Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Anton Razzhigaev

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Crossroads of text, vision, and graph models: advancing multimodal systems

Supervisor: Associate Professor Alexander Panchenko
Co-supervisor: Dr. Andrey Kuznetsov

Name of the Reviewer: **Sergey Zagoruyko**

I confirm the absence of any conflict of interest	**Date: 26-09-2024**

The purpose of this report is to obtain an independent review from the members of PhD defense Jury before the thesis defense. The members of PhD defense Jury are asked to submit signed copy of the report at least 30 days prior the thesis defense. The Reviewers are asked to bring a copy of the completed report to the thesis defense and to discuss the contents of each report with each other before the thesis defense.

If the reviewers have any queries about the thesis which they wish to raise in advance, please contact the Chair of the Jury.

Reviewer's Report

This review evaluates the thesis titled "Crossroads of Text, Vision, and Graph Models: Advancing Multimodal Systems" by Anton Razzhigaev.
This thesis delves into the integration of language processing, image synthesis, and knowledge-based systems using advanced representation learning techniques. It highlights key innovations such as a multilingual Knowledge Graph Question Answering system and the Kandinsky text-to-image synthesis model. The research also examines the geometric properties of transformer-based embeddings, providing insights into model optimization and multimodal performance. By developing and analyzing diverse embedding techniques, the thesis sets new benchmarks in machine learning and paves the way for future advancements in natural language processing, computer vision, and knowledge representation.
Strengths: Anton did an outstanding research work spanning across multiple fields.
The research works are unified by the Transformer architecture, which made this kind

of research possible. Actually, it would be very difficult to even imagine it even five years ago, in pre-Transformer times. The unification of the aforementioned fields has multiple advantages. From the academic point, using the same or similar approaches leads to faster research exploration. From the practical point, it speeds up adoption of AI techniques. Anton's research makes an important step towards unification. The thesis makes a substantial contribution to the field by proposing novel multimodal integration techniques. The research is methodologically sound and the results are compelling.
Weaknesses: Some areas could benefit from further exploration, such as the scalability of the proposed methods and their application to different datasets (e.g. BPE encoding for images). Additionally, the thesis lacks visionary thoughts on the impact and future implications of the overall research work.
The thesis work resulted in high quality publications in most impactful machine learning conferences and journals.

Provisional Recommendation

☑ *I recommend that the candidate should defend the thesis by means of a formal thesis defense*
☐ *I recommend that the candidate should defend the thesis by means of a formal thesis defense only after appropriate changes would be introduced in candidate's thesis according to the recommendations of the present report*
☐ *The thesis is not acceptable and I recommend that the candidate be exempt from the formal thesis defense*