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: Prof. Vladislav Myasnikov

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Date: 17-09-2024**

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

Reviewers report should contain the following items:

- Brief evaluation of the thesis quality and overall structure of the dissertation.
- The relevance of the topic of dissertation work to its actual content
- The relevance of the methods used in the dissertation
- The scientific significance of the results obtained and their compliance with the international level and current state of the art
- The relevance of the obtained results to applications (if applicable)
- The quality of publications
- The summary of issues to be addressed before/during the thesis defense

Anton Razzhigaev's thesis presents an exploration of multimodal systems, specifically focusing on the convergence of natural language processing, visual synthesis, and knowledge-based frameworks. The work stands out for its emphasis on creating practical, scalable models that bridge these domains, contributing both theoretical insights and applied innovations. The thesis introduces the *M3M* system for multilingual Knowledge Graph Question Answering, which addresses the challenge of cross-lingual information retrieval, a crucial topic for today's increasingly globalized and multilingual data environments. The system's ability to efficiently handle diverse languages through integrated text and graph embeddings is a technical highlight. In terms of generative AI, the development of the *Kandinsky* model showcases a cutting-edge approach to text-to-image generation, leveraging latent diffusion techniques and image priors. This methodological choice significantly advances the quality and fidelity of generated images, positioning the work within the upper echelon of current research in image synthesis. A particularly compelling aspect of the thesis is its analysis of transformer architectures, which reveals the linearity of embedding transformations across layers. This finding not only deepens our understanding of transformers but also suggests new avenues for improving model efficiency, such as pruning or layer simplification strategies. Moreover, the examination of anisotropy and intrinsic dimensions within transformers provides foundational insights into how these models structure internal representations. While each chapter of the thesis contributes substantial findings, the connection between the sections could be articulated more clearly.

As shortcomings and issues regarding the work, I would like to highlight the following:

I) Grammatical and technical errors:

1) page 26: "Introduced by , Transformers Vaswani et al. [256]..." - punctuation is incorrect;
2) page 62: On-line KGQA-demo (https://kgqa-nlp-zh.skoltech.ru) does not work;
3) page 62: "Appendix" referred to by the author is missing;
4) page 102: Eq.(8.1) - the value of "t" is not defined;
5) Missing or incomplete positions in the list of references: [77, 92-93];
6) Formal mathematical inaccuracies:
   a) page 29: Eq.(3.1) - it is not specified how the value of "k" is related to the values of "n_samples" and emb_dim;
   b) page 29 "set of pairs {(r_1, r_2)}" and page 30 "{mu_i)}" - the specified sets are singletons, the indices and their membership in the set are lost;
   c) page 29: Eq.(3.4) - the index "i" on the right-hand side of the expression is lost.

II) Issues:

1) The author's personal contribution is not indicated anywhere in the dissertation. The pronoun "We" is used throughout the text (e.g., "We introduce Kandinsky..." - page 78, "This paper presents our...investigation..." - page 79, etc.).
2) Chapter 3: A certain contradiction in the results of the first section needs to be explained.
On the one hand, the "dimensionality" of the space, determined by the concentration of the vector(s) with the maximum singular value, decreases, approaching "1"; and on the other hand, the "internal dimension", determined by expression (3.6), increases.
Ibid.: Expression (3.6) turns out to be inapplicable in some fairly simple cases. For example, in the case of the location of points at the nodes of a uniform lattice for an arbitrary dimension.
In this case: mu = 1 (= const), F (mu) = 1 for mu > = 1 and d = log (0) / log (1) = log (0) / 0! We have the same paradoxical result for the 1D case of points of a geometric progression.
How legitimate is the use of this approach here?

3) Chapter 4, page 43: The author introduced the "metric" linearity_score. Does this metric satisfy the basic properties/axioms of metric? Since the introduced measure is not symmetric, how exactly was it used on the transformer layers (what is X and what is Y)?
4) Chapter 4: How do the results of the Chapter "On Linearity of Contextual Representations Transformations in Transformers-decoders" agree with the results of the paper
(Andreas Veit, Michael Wilber, Serge Belongie "Residual Networks Behave Like Ensembles of Relatively Shallow Networks" // arXiv:1605.06431v2 [cs.CV] 27 Oct 2016)
and how do they differ?

Razzhigaev's research is both ambitious and well-executed, as evidenced by the thoroughness of his experiments and the breadth of evaluations conducted. The practical relevance of his work is enhanced by the public availability of models and code, ensuring that these contributions can have a real-world impact. His numerous publications in top-tier venues, such as CORE A*, further validate the quality and innovation of his contributions. Given the originality, technical depth, and broad applicability of the work, I confidently recommend Anton Razzhigaev for the award of a doctoral degree.

Vladislav Myasnikov,

Professor at Samara National Research University

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*