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

The thesis presents novel research on multimodal machine learning systems, focusing on the integration of language processing, image synthesis, and knowledge-based systems. The text consists of 10 chapters, providing a comprehensive overview of the field and the author's contributions.

The topic of the dissertation is fully relevant to its content. The author explores various aspects of multimodal AI systems, including multilingual knowledge graph question answering, text-to-image synthesis, and analysis of transformer-based models, which align closely with the thesis title and objectives.

The methods used in the dissertation, including embedding techniques, diffusion models, and transformer analysis, are highly relevant to the current state of the art in natural language processing, computer vision, and knowledge representation. The approaches demonstrate competitive or superior performance compared to existing methods.
The scientific significance of the results obtained and their compliance with the international level and current state of the art

The thesis presents several significant contributions:
The M3M system for multilingual KGQA, showing improvements over existing baselines.
The Kandinsky model for text-to-image synthesis, achieving competitive results among open-source solutions.
Novel insights into the linear properties and anisotropy of transformer-based models.
The OmniDialog benchmark and OmniFusion model for multimodal dialogue systems.
These contributions align with and in some cases advance the current state of the art in their respective areas.

The results have clear applications in improving efficiency and performance of NLP and computer vision models. The M3M system has potential applications in multilingual information retrieval, while the Kandinsky model advances the field of AI-generated imagery. The insights into transformer models could lead to more efficient model architectures.

The candidate's research findings have been published in high-quality venues, including CORE A* and A ranked conferences, demonstrating their compliance with international standards and current research trends.

There are some issues or suggestions that could be discussed during the defense:
A more detailed analysis of the M3M system's performance on ambiguous queries would strengthen the work, particularly given the known limitations of datasets like SimpleQuestions.
Further discussion on the discrepancy between FID scores and human evaluation for different priors in the Kandinsky model would be beneficial.
It would be interesting to hear the candidate's thoughts on potential real-world applications of the linear relation discovered in transformer layers.

More discussion on the challenges from the perspective of methods, rather than focusing primarily on real-world application challenges, would enhance the methodological contribution of the thesis.

I recommend Anton Razzhigaev for a doctorate.
Reviewer: Prof. Dr. Evgeny Burnaev