Skoltech
Skolkovo Institute of Science and Technology

Jury Member Report – Doctor of Philosophy thesis.

Name of Candidate: Sergey Sosnin

PhD Program: Computational and Data Science and Engineering

Title of Thesis: Exploration of chemical space by machine learning

Supervisor: Professor Maxim Fedorov

Name of the Reviewer:

I confirm the absence of any conflict of interest (Alternatively, Reviewer can formulate a possible conflict)	**Signature:**  **Date: 20-11-2020**

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

The subject of this thesis is the study of the chemical space, which is estimated to comprise of 10^{60} chemical structures, using artificial intelligence (AI) programs. While the pharmaceutical and chemical industries are eager to find organic compounds that can be used as new drugs and specialized materials, using currently-available computational methods to analyze the entire chemical space is not feasible because of its enormous size. Thus, newly developed algorithms and approaches are needed to address this challenge which makes this thesis very relevant.

The candidate has developed several approaches, programs, and computational tools to examine the chemical space "both for predictions of the properties of organic compounds, for the visualization of regions of chemical space, and for the sampling of compounds from chemical space." This includes AI-based programs which have been developed to predict the bioconcentration and toxicity of compounds, and a library of molecular "rules" (Legogram) to create organic compounds. The candidate has shown that these newly developed tools and approaches outscore the current state-of-the-art methods, thus demonstrating the significance of this work.

The thesis is well written and the goals and results are fairly understandable -- even for readers like this reviewer who has no expertise in AI. Therefore, the reviewer cannot judge the methods applied and developed in this thesis, but the number of publications in highly respected journals based on the candidate's work testifies to the quantity and quality of the work presented in this thesis.

In summary the study is of high significance, and the results of applying the newly developed programs and approaches clearly demonstrate the superiority of these new methods to the current methods. The conclusions are appropriate, and the thesis is well written and of high quality. I therefore recommend that this thesis be accepted in partial fulfillment of the requirement for a degree of Doctor of Philosophy.

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*