



REVIEW

For the Public Defense of Doctoral Dissertation

Prepared by: Prof. Anton Gerunov, Ph.D., D.Sc., Department of Industrial Economics and Management, Faculty of Economics and Business Administration

Dissertation Topic: Application of Machine Learning Methods for Solving Business Problems

Doctoral Student: Sergey Sergeev Vichev

1. Professional Profile

Sergey Sergeev Vichev graduated from the University of National and World Economy (UNWE) with excellent results - initially as a bachelor (2012), and then as a master (2018) in business administration. He was enrolled in the Faculty of Economics in 2021 as a full-time doctoral student in the PF 3.8 Economics and was discharged with the right to defend in 2024 by order RD-20-1364/29.07.2024. In addition to his academic pursuits, Sergey Vichev also has extensive practical experience - as a data analyst, founder and managing director of companies, and currently as the chief scientific director of a consulting company with a focus on artificial intelligence. I consider the extensive practical experience to be an excellent foundation, but also a valuable addition to the scientific profile of the doctoral student.

2. Overview of the Dissertation

The presented dissertation consists of 155 pages of main text and a list of information sources (bibliography). The structure of the dissertation is logical, with the results divided into three chapters, which consistently fulfill the set goals and objectives. From a structural point of view, it is appropriate to decide that the individual chapters have their own tasks, as well as to end with a review and summary of the main results. The introduction contains the main necessary requisites - relevance of the topic, goal and objectives, subject and object of the study. We note the lack of an explicitly derived scientific hypothesis in the introductory part, which would be useful, although the hypotheses are clearly implied in the research chapters themselves.

The first chapter "Theoretical Foundations" is essentially an overview. It outlines the theoretical foundation on which the subsequent research is based. The main concepts are examined and detailed information about the major language models (LLMs) is provided. An overview is also given of the applications of this class of models in the service of business, with the dissertation focusing on multimodality, generation with enriched information extraction and fine-tuning. The chapter creates confidence in the reader that the doctoral student has a thorough knowledge of the basic theoretical and practical propositions in the field of the dissertation. This is confirmed by the large volume of current scientific research in the bibliography. In a previous version of the dissertation, Chapter One was significantly longer, but in the final version it has been significantly shortened (from 82 pp. to 54 pp.), which makes it more readable and more useful for researchers and practitioners.

Chapter Two "Methodology for Building an Automated System for Solving Deep Business Problems on Database Information" presents a general description of the architecture of the developed system. The methods for forecasting with LLM, approaches for improving embedding models, as well as a detailed block diagram of the proposed automated system are considered. The architecture in question is significant in volume, with the doctoral student developing each of its components in detail and conscientiously. In some, the doctoral student includes relatively large blocks of code in the main text, but many of them are essentially not of a contributing nature. I would recommend that the blocks of code (or at least the majority of them) be exported into a dedicated application. This is especially true for less essential ones such as used versions of Python libraries (p. 101), common code available in the documentation of the relevant libraries and packages (e.g. p. 103, but also others), code for importing used libraries (p. 106), etc. In many cases, the description of the procedures and parameters is available in the documentation of the library itself (often this is Langchain) and I recommend that the doctoral student refer to the original source instead of including the same text in the dissertation. This is especially true for the description of the arguments of the functions used.

The third chapter "Simulations and experiments with automated database decision systems" is essentially the main contribution chapter of the dissertation. Here, two different applications of the proposed methodology (called "experiments") are tested. In the first experiment, the proposed architecture is tested on the serious game SQL Murder Mystery (Canale & Farinetti, 2022), considering three of its variants - a memoryless agent, an agent with contextual memory and a multi-agent system. In essence, we emphasize that this game was developed more for educational purposes and in this sense does not meet the narrow definition of a business application. I recommend that the dissertation candidate justify why the demonstrated functionalities and results within this educational serious game are close to those we would see in real business cases. The current argumentation ("We use this particular case because it is already well thought out and repeatedly tested", p. 123) is insufficient. The second

“experiment” essentially aims to test whether the specialization training of the embedding model within the architecture proposed by the dissertation candidate leads to improved results (measured using the NDCG metric). This is probably where the most interesting result of the dissertation is achieved: it turns out that the pre-training of a (very small) embedding model leads to significantly better results compared to significantly larger models that have not undergone pre-training (fine-tuning). The conclusion is of undoubted interest for both theory and practice.

The dissertation ends with a conclusion and a set of scientific and applied scientific contributions.

3. Evaluation of Scientific and Applied Contributions

Undoubtedly, the presented dissertation is an original and interesting scientific work with all the necessary requisites for a doctoral dissertation. The contributions in the dissertation are sufficient in quantity and quality and fully meet the requirements for acquiring the ONS "doctor". I accept the contributions formulated in this way, but they should be edited and further refined.

In essence, contribution 1 should be further corrected. Since there are already sets of similar methodological frameworks for knowledge extraction (see, for example, Xie et al., 2024¹, but also others), and in this sense, the contribution should be refined to emphasize what new knowledge the dissertation provides. Contribution 2 should focus on the acquired result regarding the representation of low-dimensional language models. Contribution 3 consists mainly of comparative results with the ReAct framework.

I fully accept contributions 4 and 5. Here, the undoubted benefit of developing additional artifacts within the dissertation effort, which would allow not only the reproduction of the results obtained, but also the upgrading of the presented research, should be emphasized.

4. Critical Evaluation and Recommendations

Specific critical remarks and recommendations are made in section 1 regarding the relevant chapters. It is also worth noting some considerations of a more general nature:

- The dissertation is essentially a great example of design science, in which the author creates a specific applicable artifact (Dresch et al., 2015)². Although the

¹ Xie, W., Wu, G., & Zhou, B. (2024). *Mag-sql: Multi-agent generative approach with soft schema linking and iterative sub-sql refinement for text-to-sql*. arXiv preprint arXiv:2408.07930.

² Dresch, A., Lacerda, D. P., Antunes Jr, J. A. V., Dresch, A., Lacerda, D. P., & Antunes, J. A. V. (2015). *Design science research* (pp. 67-102). Springer International Publishing.

approach and methodology of the work clearly follow the principles and practices of design science, this is not explicitly mentioned in the methodological section. I would recommend adding this as an additional nuance to the work. The application of this approach can essentially also represent a methodological contribution to the field.

- The topic of the dissertation focuses on solving business problems, while a significant part of the contributions are not explicitly business-oriented. The subtitle already points to “complex tasks in databases”, but this is also not a very clearly defined term. I would recommend that in his future research, the PhD student consider how the presented results can be applied to specific business tasks.
- A key issue in the evaluation of alternative architectures is the approach to evaluating the achieved results. The thesis focuses on NDCG, but in the literature and practice there are different approaches, especially in the decomposed evaluation of the different components of the generation with enriched information retrieval (RAG), see e.g. Es et al., 2024³. It is useful to have a more extensive justification of the methodological considerations regarding the choice of a specific evaluation framework.

I emphasize that the comments made do not diminish the value of the obtained results and should be considered primarily as potential guidelines for improving the dissertation.

5. Evaluation of Publications and Compliance with National Minimum Requirements

Doctoral student Sergey Vichev **presents evidence of significant publication activity** on the topic of the dissertation:

- Sergey Vichev, Angel Marchev, Reasoning capabilities of large language models on information extracted from databases using Text-to-SQL, PROCEEDINGS OF THE 49TH INTERNATIONAL CONFERENCE “APPLICATIONS OF MATHEMATICS IN ENGINEERING AND ECONOMICS”, 2025, doi:<https://doi.org/10.1063/5.0247481>, Ref, IF (- 2025), SCOPUS, SJR (- 2025)
- Sergey Vichev, A Multi-Agent Framework for Large Language Model-Driven SQL Analysis, KSI Transactions on KNOWLEDGE SOCIETY, vol:XV, issue:Nr. 1, 2024, ISSN (print):1313-4787

³ Es, S., James, J., Anke, L. E., & Schockaert, S. (2024, March). Ragas: Automated evaluation of retrieval augmented generation. In *Proceedings of the 18th Conference of the European Chapter of the Association for Computational Linguistics: System Demonstrations* (pp. 150-158).

- Vichev, Sergey, Marchev, Angel, RAGSQL: Context Retrieval Evaluation on Augmenting Text-to-SQL Prompts, International IEEE Conference proceedings, IS, 2024, doi:10.1109/IS61756.2024.10705186, Ref, SCOPUS, SJR (- 2024)
- Vichev, Sergey, Application of K-means Clustering Algorithm with RFM for Customer Segmentation in Logistics Sector., Vanguard Scientific Instruments in Management, issue:19, 2023, ISSN (print):ISSN: 1314-0582, Ref, etc.(CEEOL)

It should be noted here that of the four publications cited, two are indexed in the globally renowned scientific information database Scopus, which indicates the high level of the research conducted.

Additionally, a Certificate of Compliance with the Criteria and Requirements for Doctoral Theses for Obtaining the Educational and Scientific Degree "Doctor" of the Faculty of Economics, Sofia University "St. Kliment Ohridski" is presented. It shows that in addition to publications, the doctoral student has 4 reports at national scientific schools and seminars, as well as 3 more reports at international scientific forums in the country. In addition, the doctoral student also teaches the course Databases in Finance, which also testifies to his teaching experience.

Based on the documents presented, it can be concluded that doctoral student Sergey Vichev **not only meets, but significantly exceeds the national minimum requirements** for acquiring the degree "Doctor" (Art. 12 of the ZRASRB, Art. 35 of the Regulations for the Implementation of the ZRASRB and Art. 1a, para. 1 of the PPZRASRB).

6. Evaluation of the Dissertation Summary

The dissertation summary presents the main results of the dissertation in a detailed and comprehensive manner.

7. Conclusion

The dissertation work represents a thorough and conscientious scientific research, which clearly demonstrates the deep theoretical knowledge of the doctoral student, as well as the skills to conduct independent scientific research at a high level. The text contains scientific results that represent an original contribution and meet the minimum requirements of the Act on the Development of the Academic Staff in the Republic of Bulgaria (ADSRB) and the Regulations for the Implementation of the ADSRB. I believe that the presented dissertation fully meets the requirements of Art. 6, para. 3 of the ADSRB and give my positive assessment of the presented work, **as I strongly recommend to the esteemed scientific jury to award the degree "doctor"** to doctoral student Sergey Sergeev Vichev in Professional Field 3.8 Economics.

Reviewer:

Prof. Anton Gerunov, Ph.D., D.Sc.

Sofia, 16.06.2025