

REVIEW

by Prof. Daniela Borissova, D.Sc. at IICT-BAS
of a dissertation by Miroslava Doncheva Dimitrova
on the topic “Evaluation Framework of Retrieval-Augmented Generation”
Doctoral Program “Informatics”, Professional Direction 4.6. “Informatics and
Computer Science”

According to order No 75/27.03.2026 of the Director of IICT-BAS, I have been appointed to be a member of the scientific jury, and according to the minutes of the first meeting held on 30.03.2026, I have been selected to prepare a review. For this purpose, I have received all the documents submitted by the candidate under the procedure.

RELEVANCE AND KNOWLEDGE OF THE RESEARCHED PROBLEM

In the digital age, data plays a key role in making fast and reliable decisions. In this context, the presented dissertation examines an interesting and topical topic related to the generation of content with advanced retrieval. Such systems are increasingly used in areas where factual accuracy and traceability directly affect the quality of decision-making, such as healthcare, law, etc. From the presented review and the described results, it can be established that the doctoral student is well acquainted with the nature of the research problem.

ANALYTICAL CHARACTERISTICS

The dissertation has a total volume of 214 pages. It is structured as follows: list of tables, list of figures, list of equations, list of algorithms, glossary of terms and abbreviations, introduction, 5 chapters, conclusion - summary of the obtained results, 3 appendices, bibliography, list of publications, list of citations, list of

projects with participation, acknowledgments, declaration of originality of the results.

On page 26, the aim of the dissertation is defined as “*to develop an evaluation framework for Retrieval-Augmented Generation that supports evidence-based retrieval configuration decisions for RAG systems with open-source LLMs, with particular focus on similarity threshold configuration*”, and 3 research questions are formulated: “*Q1: Does varying the similarity threshold produce measurable changes in generation quality? Do similarity threshold effects differ across language models? RQ3: Do comparable similarity threshold ranges hold across knowledge domains?*”

On page 27, 4 more objectives are presented, which are pursued: “*Objective 1: Define and implement the core components of the evaluation framework; Objective 2: Establish model selection criteria; Objective 3: Define metric selection and computation procedures; Objective 4: Conduct controlled testing and analysis.*” The dissertation continues with the formalization of 3 scientific and applied contributions. This structure differs significantly from the generally accepted structure of a dissertation, in which the goal and tasks for achieving it are formulated based on the literature review.

Chapter 1 traces the development of Retrieval-Augmented Generation (RAG) from information retrieval and natural language processing, presents existing RAG architectures and their characteristic shortcomings, discusses evaluation approaches, and identifies gaps in the literature. Different development directions and evaluation approaches are described, identifying the selectivity of retrieval, or, more precisely, the choice of similarity threshold, as a factor with an insufficiently studied influence on generation quality.

Chapter 2 presents the design and architecture of a system for structured generation evaluation supplemented with retrieval (PaSSER). The goal is to integrate threshold-aware retrieval, multi-metric scoring, and blockchain-supported provenance logging into a unified workflow for controlled experimentation. A performance evaluation system for similarity scoring and retrieval implements reproducibility for threshold-aware RAG evaluation. The architecture integrates three functional layers: a browser-based interface that allows for experiment configuration and result visualization, a Python backend that coordinates retrieval operations via ChromaDB vector storage and language

model inference via the Ollama API, and an Antelope blockchain layer that provides tamper-proof provenance logging for all evaluation results.

Chapter 3 describes the components required for the experimental analysis, namely, large open-source language models and metrics for assessing the quality of the generated responses. Here, a composite performance metric and a composite threshold stability metric are defined, used to aggregate a set of metrics.

Chapter 4 presents a description of the conducted research to assess the influence of the similarity threshold on the quality of RAG generation across seven open-source language models and two subject areas, using PaSSER.

In essence, Chapter 5 summarizes the results obtained, described in the previous chapters, systematizes the scientific and applied contributions in narrative form, and outlines some directions for future work.

ABSTRACT AND AUTHOR REFERENCE

The presented abstract reflects the content of the dissertation work and complies with the requirements of the LAW ON THE DEVELOPMENT OF THE ACADEMIC STAFF IN THE REPUBLIC OF BULGARIA (LDASRB) and Regulations for its implementation. From the presented declaration of originality, as well as from the publications, it can be determined that the described results are the personal work of the author. In addition, it should be noted that the result of the plagiarism check performed on the website strikeplagiarism.com showed the following: Similarity coefficient 1: 2.73% and Similarity coefficient 2: 0.81%.

ASSESSMENT OF COMPLIANCE WITH THE MINIMUM NATIONAL REQUIREMENTS AND WITH THE ADDITIONAL REQUIREMENTS UNDER ART. 1A, AL. 2 OF THE REGULATIONS FOR IMPLEMENTATION OF LDASRB

A total of 5 publications are presented, all in English and co-authored. Two of the publications are indexed in IEEE Xplore, and another 2 are IF and indexed in Scopus and WoS.

According to the Regulations for implementation of LDASRB, to fulfill the minimum national requirements for obtaining the educational and scientific degree “Doctor” in the professional field 4.6 “Informatics and Computer Sciences”, 30 points are required for Group of Indicators D. The total sum of the

points for the indicators from Group D is 104 points, which significantly exceeds both the specific requirements of IICT-BAS and the minimum national requirements for professional field 4.6 “Informatics and Computer Sciences” for acquiring the educational and scientific degree “Doctor”.

CONTRIBUTIONS

I assess the contributions of the doctoral student as scientifically applied, leading to the enrichment of existing knowledge, reformulated as follows:

1. A formulation for a composite performance score (CPS) is presented.
2. A formulation for a threshold-dependent composite performance score (T-CPS) is presented.
3. An architecture of a system for structured generation evaluation, supplemented by extraction, is proposed.

CRITICAL COMMENTS AND RECOMMENDATIONS

The dissertation is relatively well-balanced. The presentation is described in a way that is too descriptive for a scientific study, instead of focusing on the specific results and their analysis. This does not diminish its scientific value, but it would be good in the future for the results to be presented in a more rigorous form. There are incorrectly written mathematical formulas, such as (3.1) on page 89 and (3.32) on page 107.

I have the following questions for the doctoral student:

1. On page 20 in the list of abbreviations, “PaSSER (Platform for Systematic and Structured Evaluation of RAG)” is written, and on page 59, “Performance Assessment System for Similarity Evaluation and Retrieval (PaSSER)”. What requires two different interpretations of the same abbreviation? What is the difference between a platform, a framework, a system and a browser-based environment?
2. What are the differences in the optimal values of the thresholds in different thematic areas due to?
3. Why were the tools of explainable artificial intelligence not used, which allow interpretation of the results?

CONCLUSION

The presented results on the topic of the dissertation convincingly show that the doctoral student Miroslava Doncheva Dimitrova possesses the necessary theoretical knowledge and practical skills in the field of informatics and computer sciences, as well as proven abilities for independent scientific research. The presented dissertation meets the requirements of the Act on the Development of Academic Staff in the Republic of Bulgaria, the Regulations for its implementation and the Regulations on the Specific Conditions for Acquiring Scientific Degrees and for Holding Academic Positions at the IICT-BAS. **The obtained results on the topic of the dissertation research are sufficient grounds for me to give a categorically positive assessment of the presented dissertation and I propose to the esteemed Scientific Jury to award Miroslava Doncheva Dimitrova the educational and scientific degree "Doctor" in the doctoral program "Informatics", professional direction 4.6. "Informatics and Computer Sciences".**

30 April 2026

Scientific Jury 1

