

РЕЦЕНЗИЯ

от проф. д.н. Даниела Ананиева Орозова, Тракийски университет - Стара Загора
на дисертационен труд за придобиване на образователна и научна степен „доктор“
по професионално направление 4.6. „Информатика и компютърни науки“,
докторска програма „Информатика“ (01.01.12)
на тема: „Evaluation Framework of Retrieval-Augmented Generation“
(„Рамка за оценка на съдържание, генерирано с разширено извличане“)

от **Мирослава Дончева Димитрова**

Със заповед № 75/ 27.03.2026 г. на Директора на ИИКТ-БАН и на основание чл. 4, ал. 2 от Закона за развитие на академичния състав в Република България, във връзка с процедура за придобиване на образователна и научна степен „доктор“ по професионално направление 4.6. „Информатика и компютърни науки“, докторска програма „Информатика“ от Мирослава Дончева Димитрова с дисертационен труд на тема „Evaluation Framework of Retrieval-Augmented Generation“ („Рамка за оценка на съдържание, генерирано с разширено извличане“), съм определена за член на Научното жури и рецензент.

При оценката на дисертационния труд са приложени изискванията на Закона за развитие на академичния състав в Република България (ЗРАСРБ), Правилника за прилагане на ЗРАСРБ (ППЗРАСРБ, ПМС № 26 от 13 февруари 2019 г.) и Правилника за специфичните условия в ИИКТ-БАН за прилагането на закона:

Съгласно чл. 27, ал. 1 от ППЗРАСРБ „дисертационният труд трябва да съдържа научни или научно-приложни резултати, които представляват оригинален принос в науката. Дисертационният труд трябва да показва, че кандидатът притежава задълбочени теоретични знания по съответната специалност и способности за самостоятелни научни изследвания“.

Според чл. 27, ал. 2 от ППЗРАСРБ „дисертационният труд трябва да бъде представен във вид и обем, съответстващи на специфичните изисквания на първичното звено. Дисертационният труд трябва да съдържа: заглавна страница; увод; изложение; заключение – резюме на получените резултати с декларация за оригиналност; библиография“.

Минималните изисквания по групи показатели за придобиване на образователна и научна степен „доктор“ съгласно ППЗРАСРБ са покрити.

Научен ръководител на дисертационния труд е акад. Иван Попчев.

1. Общи данни за дисертационния труд

Дисертационният труд е представен на английски език.

Формулираната цел е: „...да се разработи рамка за оценка на RAG, която подпомага вземането на решения за конфигуриране на извличането в RAG системи при езикови модели с отворен код, с особен акцент върху настройването на прага на сходство.“ Английският еквивалент (стр. 26 от дисертационния труд) е: „The aim of the dissertation is 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.“

За постигане на целта на стр. 27 от дисертационния труд са формулирани четири задачи:

Задача 1 (Obj.1): Да се дефинират и реализират основните компоненти на рамката за оценка, интегрираща три компонента: (a) процедура за оценка, отчитаща прага на сходство чрез композитни показатели (composite scoring); (b) платформата за оценка на производителността PaSSER (Performance Assessment System for Similarity Evaluation and Retrieval) [15], осигуряваща инфраструктура за възпроизводимост, чрез запис на резултатите върху блокчейн; и (c) контролиран експериментален дизайн, за извеждане на сравнителни данни между различни модели и области.

Задача 2 (Obj.2): Установяване на критерии за избор на модели. Дефиниране на критерии за подбор, съобразени с възможностите за локално внедряване, лицензионен режим и изчислителни ресурси. Това включва профилиране на избраните модели по отношение на техния контекстен прозорец и декодинг настройки.

Задача 3 (Obj.3): Определяне на процедури за подбор и изчисляване на метрики. Да се подберат метрики, съответстващи на оценяваните характеристики - лексикално припокриване, семантично сходство, гладкост на текста, точност и езиково моделиране - и да се осигури последователното им изчисляване при всички модели и експериментални условия.

Задача 4 (Obj.4): Провеждане на контролирано тестване и анализ. Подготовка на тематични масиви от данни (корпуси) и набори от въпроси и отговори, включваща специфична предварителна обработка и конфигурация на извличането. Изпълнение и оценка на серия от тестове и изследване на целия диапазон от прагове на сходство (threshold sweeps). Анализ и обобщаване на резултатите, за да се установи как прагът на сходство влияе върху качеството на генериране, възпроизводимостта и избора на подходяща конфигурация.

На стр. 26 от дисертационния труд са формулирани три изследователски въпроса:

- Въпрос 1 (RQ1): Води ли промяната на прага на сходство до измерими промени в качеството на генерираното съдържание?
- Въпрос 2 (RQ2): Различава ли се ефектът от прага на сходство при различните езикови модели?
- Въпрос 3 (RQ3): Валидни ли са сходни диапазони на прага при различна тематична област (домейн)?

Дисертационният труд е в обем от 215 страници и включва: списък на таблици, списък на фигури, списък на уравнения, списък на алгоритми, речник на термините и съкращенията (стр. 16–23), увод (стр. 24–29), пет основни глави (стр. 30–175), заключение – резюме на получените резултати (стр. 176–178), три приложения (A, B, C, стр. 179–191), библиография (стр. 192–204), списък на публикациите по дисертационния труд (стр. 205), забелязани цитирания (стр. 207–212), справка за участие в научни проекти (стр. 213), благодарности (стр. 214) и декларация за оригиналност (стр. 215). Представени са 25 фигури и 50 таблици. Библиографията съдържа 145 заглавия и обхваща основополагащи и съвременни източници от областите на информационното извличане, обработката на естествен език, генеративните езикови модели, оценъчните рамки и възпроизводимостта на изследванията. Съотношението между основното научно изложение (Глави 1–5) и спомагателните раздели е балансирано.

Структурата следва логиката на научно-приложното изследване: преглед на състоянието на проблема, проектиране на инструментален апарат, дефиниране на метричен апарат, систематичен експеримент и обсъждане. Преходите между частите са плавни, а обратната проследимост от дефицит през изследователски въпрос, задача, глава и принос е обобщена в Таблица I.1 (стр. 28).

2. Оценка на съдържанието по глави

Глава 1 „Retrieval-Augmented Generation“ (стр. 30–58) е литературно-аналитичен преглед, който проследява развитието на информационното извличане от основните индекси и класификационни схеми до съвременните модулни архитектури Self-RAG, CRAG, FLARE, GraphRAG и LightRAG. Прегледът обхваща архитектурните компоненти на RAG, иновациите в областите на ефективността, данните, итеративното извличане, мултимодалното разширение, адаптацията към тематичната област (domain adaptation) и верификацията. В раздел 1.3.8 е отделено специално внимание на праговете на подобие и селективността на извличането, с което се подготвя теоретичната обосновка на експерименталната част. В раздел 1.4 са анализирани утвърдените средите за оценка (evaluation frameworks) RAGAS, RGB, TREC 2024 RAG Track и TruLens, а в 1.4.6 са очертани три изследователски дефицита: (D1) отсъствие на прагово-чувствителна оценка, (D2) недостатъчна инфраструктура за възпроизводимост, (D3) ограничени сравнителни данни за отворени модели с 7–8 млрд. параметри. Прегледът е систематичен, с ясна проследимост между открояваните дефицити и последващите изследователски въпроси.

Глава 2 „Design and Architecture of PaSSER“ (стр. 59–78) представя разработената платформа PaSSER (Performance Assessment System for Similarity Evaluation and Retrieval) – браузър-базирана, модулна система за конфигурация и оценка на RAG с отворени LLMs. Трислойната архитектура (уеб интерфейс, бекенд-услуги и блокчейн-подсистема) е обоснована от гледна точка на повторяемост и проследимост. Интегрирана е блокчейн мрежа Antelope, разгърнатата като частна мрежа с разрешение (permissioned private blockchain network), в която се записват експерименталните конфигурации, параметрите на извличане, декодинг настройките (decoding settings) и идентификаторите на изпълнения. Това представлява оригинално инфраструктурно решение, което свързва оценъчната процедура с контролирано документирание на условията. Функционалностите за конфигуриране на системата, управление на данни, настройване на извличането, взаимодействие с модела и оценка са описани с необходимата техническа прецизност и са приведени към публикувани източници [17], [16], които представят платформата.

Глава 3 „Model Selection and Evaluation Metrics“ (стр. 79–111) дефинира аналитичния апарат. Изборът на седем отворени LLMs в диапазона 7–8 млрд. параметри (Mistral 7B v0.1 и v0.3, Llama 2 7B, Orca 2 7B, Granite 3.2 8B, DeepSeek R1 8B, Llama 3.1 8B) е обоснован чрез критериите за възможност за локално внедряване, лицензна съвместимост и изчислителни изисквания. Дефинирани са 24 метрики, групирани в пет категории – лексикално припокриване (lexical overlap), семантично подобие (semantic similarity), плавност и качество на отговора (fluency, predictive, and answer quality), статистическа корелация (statistical correlation) и приближение към човешка четимост (human-readability inspired metrics, B-RT). Формулирани са оригиналните композитни показатели Composite Performance Score (CPS), Threshold-aware Composite Performance Score (T-CPS) и Balance Score, които обединяват разнородните метрики в единен показател за систематично сравнение. Процедурата за статистическа проверка чрез своени t-тестове с отчитане на размера на ефекта (Cohen’s d) е формализирана в Алгоритъм А.3 от Приложение А.

Глава 4 „Experimental Evaluation and Results“ (стр. 112–154) е най-обемната и представя резултатите от четири последователни експериментални фази с нарастваща сложност. Фаза I осигурява системна валидация и измерване на времевата ефективност под фиксиран top-k режим. Фаза II въвежда вариация на прага на подобие в диапазона 0.50–0.80 и установява чувствителността на моделите. Фаза III разширява диапазона до 0.50–0.95 върху четири обновени модела в домейна земеделие, като въвежда статистическа проверка и регистрира CPS-подобрения до +4.58% (Mistral 7B v0.3 при праг 0.95, Таблица B.1). Фаза IV валидира преносимостта на резултатите към нова тематична област — биоразнообразието, където регистрираните CPS-подобрения достигат +13.32% (Таблица 4.14). Изместванията на оптималните прагове при прехода от земеделие към биоразнообразие са в диапазона от –0.05 (Llama 3.1 8B) до –0.35 (DeepSeek R1 8B). Общият брой индивидуални оценки надхвърля 38 000, което осигурява солидна емпирична база.

Глава 5 „Discussion and Future Work“ (стр. 155–175) систематизира отговорите на трите изследователски въпроса, формулира трите научно-приложни приноса, документира ограниченията на изследването в четири групи (обхват, експериментален план, измерване и причинно-следствена интерпретация) и очертава посоки за бъдеща работа. Заключение – резюме (стр. 176–178) обединява основните резултати и трите приноса.

Формулираните цел и задачи са постигнати: платформата PaSSER (Задачи 1–2), метричната рамка CPS/T-CPS/Balance Score (Задача 3) и четирите експериментални фази (Задача 4) са изпълнени с конкретни резултати. Изследователските въпроси RQ1–RQ3 получават емпирични отговори: RQ1 е потвърден със статистически значими ефекти на праговата конфигурация; RQ2 е потвърден с модел-специфични профили на чувствителност; RQ3 е получил частичен отговор – наблюдава се чувствителност към тематичната област, изискваща калибрация, съобразена с конкретния случай.

3. Научно-приложни приноси

В раздел 5.2 от дисертационния труд (стр. 159–162) авторът е формулирал три научно-приложни приноса, които изграждат интегрираната рамка за оценка (evaluation framework). Приемам формулировките в представения вид:

Принос 1 (C1): Компонент „Процедура за оценка“ (5.2.1 Threshold-aware Evaluation Procedure, стр. 159). „Въвежда процедура за оценка, отчитаща прага на сходство, която включва Композитен показател за производителност (Composite Performance Score, CPS), Композитен показател за стабилност на прага (Threshold-aware Composite Performance Score, T-CPS) и Показател за баланс (Balance Score, BS) за характеризиране на селективността на извличането при различни настройки на прага на сходство.“ [15], [16]

Принос 2 (C2): Инфраструктурен компонент (5.2.2 Reproducibility Infrastructure, стр. 161). „Реализира се инфраструктура за тестване и възпроизводимост на резултатите, съчетаваща запис на експерименталните данни в блокчейн и пълно документиране на конфигурациите чрез платформата PaSSER.“ [15], [17]

Принос 3 (C3): Компонент „Емпирични данни“ (5.2.3 Practical Guidance for Open-Source Deployments, стр. 162). „Извежда практически насоки за внедряване на RAG системи с модели с отворен код, основани на сравнителни емпирични данни, които свързват чувствителността към прага на сходство, качеството на генериране и приложимостта при внедряване при седем модела в диапазона 7–8 милиарда параметри, изследвани при контролирани експериментални условия.“ [16], [18]

Приемам посочените три научно-приложни приноса като коректно формулирани и съответстващи на извършената изследователска работа. Мащабът от над 38 000 индивидуални оценки в две приложни области представлява една от по-обемните систематични прагово-чувствителни оценки на отворени LLMs в рецензираната литература.

4. Публикации по дисертационния труд

По дисертационния труд са представени пет рецензирани публикации, които отразяват съществените елементи на изследването:

- [15] I. Radeva, I. Popchev, M. Dimitrova, „Similarity thresholds in retrieval-augmented generation,“ в Proc. 2024 IEEE 12th Int. Conf. on Intelligent Systems (IS), 2024, pp. 1–7, doi: 10.1109/IS61756.2024.10705214. Публикацията обосновава CPS-формулировката и анализа на прагова чувствителност от Глава 4, Фаза II.
- [16] M. Dimitrova, I. Popchev, I. Radeva, „PaSSER: A platform for evaluating LLMs in RAG,“ в Proc. 2025 IEEE BdKCSE, 2025, p. 7, doi: 10.1109/BdKCSE67969.2025.11300500. Публикацията описва архитектурата и функционалностите на платформата PaSSER, представени в Глава 2.
- [17] I. Radeva, I. Popchev, L. Doukovska, M. Dimitrova, „Web application for retrieval-augmented generation: Implementation and testing,“ Electronics, vol. 13, no. 7, p. 1361, 2024, doi: 10.3390/electronics13071361. Публикацията представя платформата PaSSER и началния метричен апарат, обсъдени в Глава 2.
- [18] I. Radeva, I. Popchev, L. Doukovska, M. Dimitrova, „Multi-agent coordination strategies vs. retrieval-augmented generation in LLMs: A comparative evaluation,“ Electronics, vol. 14, no. 24, p. 4883, 2025, doi: 10.3390/electronics14244883. Публикацията документира T-CPS и Balance Score, представени в Глава 4, Фаза IV.
- [20] M. Dimitrova, „Retrieval-augmented generation (RAG): Advances and challenges,“ Problems of Engineering Cybernetics and Robotics, vol. 83, 2025, doi: 10.7546/PECR.83.25.03. Публикацията представя литературния преглед и анализ на RAG-рамките, които изграждат основата на Глава 1

Запазена е номерацията на публикациите от библиографията на дисертационния труд (стр. 192–204).

Анализът на публикационната активност показва: две публикации в списание Electronics (MDPI), индексирани в Web of Science с импакт-фактор (JCR-IF) и в Scopus (SJR); две публикации в реферирани сборници на IEEE конференции, индексирани в IEEE Xplore; една самостоятелна публикация в списанието на Института по информационни и комуникационни технологии – БАН „Problems of Engineering Cybernetics and Robotics“. Публикация [20] е самостоятелна. В раздел „CITATION RECORD“ на дисертационния труд (стр. 207–212) са представени 64 забелязани цитирания върху четири от петте публикации. Публикациите покриват всички основни аспекти на изследването и достатъчно пълно апробират резултатите пред научната общност.

5. Наукометричен образ

Наукометричният образ на докторанта е документиран въз основа на отчети от Web of Science Core Collection (Clarivate Analytics, реален период 2024–2026 г.) и Scopus/SciVal (Elsevier, реален период 2023–2025 г., данни към 15.04.2026 г., справка към 23.04.2026 г.).

В Web of Science Core Collection са индексирани две от публикациите по дисертационния труд: [17] Radeva, Popchev, Doukovska, Dimitrova, „Web Application for Retrieval-Augmented Generation: Implementation and Testing“, Electronics, vol. 13, no. 7, 2024, с 13 цитирания и среден годишен поток на цитиранията 4.33; и [18] Radeva, Popchev, Doukovska, Dimitrova, „Multi-Agent Coordination Strategies vs. Retrieval-Augmented Generation in LLMs“, Electronics, vol. 14, no. 24, 2025, без регистрирани цитирания, което е очаквано за публикация от декември 2025 г.

Обобщените показатели от Scopus/SciVal за реалния период 2023–2025 г. са представени в Таблица 1.

Таблица 1. Наукометрични показатели от Scopus/SciVal за периода 2023–2025 г.

Показател	Стойност	Референтна стойност
Scholarly Output (общ брой публикации)	6	–
Citations (общ брой цитирания)	50	–
Citations per Publication	8.3	зависи от научната област
h-индекс	3	зависи от научната област и етапа на кариерата
h5-индекс	2	зависи от научната област
Field-Weighted Citation Impact (FWCI), общо	1.64 (медиана 1.22)	1.00 (световна норма)
Field-Weighted Citation Impact за 2024 г.	3.70	1.00 (световна норма)
Публикации в горните 10% най-цитирани (полево-претеглено)	33.3% (за 2024 г. – 100%)	10.0% (по дефиниция)
Публикации с отворен достъп (Open Access)	33.33%	ок. 35% (Scopus, 2024 г., ориентиrowъчно)
Patent-Citations Count	1	0 (типична стойност за докторант)

Сред представените показатели четири са с определяща тежест за наукометричния образ на докторанта: FWCI = 1.64 (медиана 1.22) за целия период и FWCI = 3.70 за 2024 г. документират „citation impact“ над световната норма (референтна стойност 1.00). Делът от 33.3% публикации в горните 10% най-цитирани (полево-претеглено) потвърждава, че „citation impact“ не е следствие от една отделна публикация. Индексът на Хирш $h = 3$ при шест публикации показва разпределение на цитиранията върху повече от една работа, което поддържа устойчивостта на профила. Регистрираното патентно цитиране (Patent-Citations Count = 1) е необичаен за докторантски етап индикатор за приложно въздействие извън академичната сфера.

Концентрацията на публикационната активност е в издания от втори квартал (Electronics, MDPI, Q2) и в индексирани сборници на IEEE конференции. Наукометричният образ подсказва посоки за бъдещо развитие: увеличаване на дела на самостоятелните (single-author) публикации, насочване към списания от Q1 по CiteScore и преминаване към потвърдителен (а не проучвателен) статистически режим при следващи експериментални разширения.

6. Участие в научни проекти

Дисертационният труд е разработен в рамките на две изследователски програми на Министерството на образованието и науката: Националната научна програма

„Интелигентно растениевъдство“ (Решение на МС № 866/26.11.2020 г.) и проект „BG PLANTNET: Създаване на националната информационна мрежа за генбанка – растителни генетични ресурси“ (КП-06-Н36) към Фонд „Научни изследвания“. Участието в тези програми очертава приложната среда, в която са формирани доменните корпуси за земеделие и биоразнообразие.

7. Автореферат

Авторефератите са изготвени на български и английски език в обем от 48 стр. и представят основните резултати, приносите и заключенията на дисертационния труд. Съдържанието на авторефератите съответства на съдържанието на дисертационния труд.

8. Критични бележки и препоръки

По отношение на представянето и оформлението на резултатите могат да се направят следните бележки:

1. Изследователските въпроси RQ1–RQ3 са емпирични по характер, а Задачи 1–3 изграждат изследователската инфраструктура. Логическата връзка между двете групи е отразена в Таблица I.1 (стр. 28) и при защитата може да бъде разяснена устно.
2. На отделни места в текста се появяват формулировки, които могат да бъдат прочетени като „загатващи“ валидация спрямо човешки оценки (напр. *validated framework*), каквато в обхвата на настоящия труд не е предвидена. При устното представяне авторът може да уточни, че работата е в режим „разработване и демонстрация“ (*developed and demonstrated*).
3. Тежестите в композитния показател CPS (Таблица 4.3: лексикално припокриване 30%, семантично сходство 25%, плавност и качество 25%, езиково моделиране 20%) са определени експертно, а устойчивостта им е подкрепена с анализ на чувствителността в публикация [18]. При устното представяне авторът може накратко да обобщи логиката на избора на теглата — това би подпомогнало читателя, който не познава [18].
4. Във Фаза IV са въведени няколко едновременни изменения спрямо Фаза III — хардуерна конфигурация, метод за генериране на въпросите и тематична област — което затруднява изолирането на чистия доменен ефект. Това е признато в Раздел 5.3 (Ограничения на експерименталния дизайн). Кратко устно пояснение при защитата би укрепило прочитата на резултатите за преносимост между области.
5. Структурата на множествените сравнения (четири модела × десет прагови нива на фаза) е причина р-стойностите да се интерпретират като проучвателни, а не като потвърдителни. Това разграничение е направено в Раздел 5.3.3 (Ограничения на измерването и анализа). Кратко устно повторение на този акцент при представянето на резултатите от Глава 4 би било полезно.
6. Фаза I е определена в труда като системна валидация и измерване на времевата ефективност (стр. 26), без пряк принос към отговорите на изследователските въпроси RQ1–RQ3. Това разграничение присъства в текста. Кратко повторение при устното представяне би помогнало на читателя да позиционира коректно четирите експериментални фази.
7. Трите представени приноса са формулирани като научно-приложни. Съгласно чл. 27, ал. 1 от ППЗРАСРБ дисертационният труд трябва да съдържа резултати, които представляват „оригинален принос в науката“. При устното представяне авторът

може да уточни кои от трите приноса според него попадат в тази категория и на какви основания.

9. Въпроси към дисертанта

1. Какви са основните критерии за избор на седемте отворени езикови модела в диапазона 7–8 млрд. параметри?
2. На какво основание е избран блокчейн Antelope за записване на произхода на експерименталните данни?
3. Какъв е броят на въпросите, използвани при оценката в домейните земеделие и биоразнообразие, и как са генерирани?
4. Каква е причината за разширяване на изследвания прагов диапазон от 0.50–0.80 във Фаза II до 0.50–0.95 във Фази III и IV?
5. Кои от представените научно-приложни приноси могат да се определят като оригинален принос в науката?
6. Кои от посоките за бъдеща работа, формулирани в раздел 5.4 (стр. 170–175), се разглеждат като най-приоритетни за непосредствено продължаване на изследването?

ЗАКЛЮЧЕНИЕ

Дисертационният труд на Мирослава Дончева Димитрова съдържа научно-приложни резултати, които представляват оригинален принос в науката, и отговаря на изискванията на Закона за развитие на академичния състав в Република България, Правилника за приложението му и Правилника за специфичните условия в ИИКТ–БАН.

Мирослава Дончева Димитрова притежава задълбочени теоретични знания в областта на обработката на естествен език, информационното извличане и оценъчните методологии за езикови модели, както и способности за самостоятелно провеждане на мащабно научно изследване.

Давам категорична положителна оценка на представения дисертационен труд, автореферата и постигнатите резултати.

Предлагам на почитаемото Научно жури да присъди образователната и научна степен „доктор“ на Мирослава Дончева Димитрова по професионално направление 4.6. „Информатика и компютърни науки“, докторска програма „Информатика“ (01.01.12).

Дата: 27.04. 2026 г.

Рецензе:
(проф. д

НА ОСНОВАНИЕ

ЗЗЛД