

## **РЕЦЕНЗИЯ**

от проф. д.н. Даниела Борисова – ИИКТ-БАН

на дисертационен труд на Мирослава Дончева Димитрова  
на тема „РАМКА ЗА ОЦЕНКА НА СЪДЪРЖАНИЕ, ГЕНЕРИРАНО С РАЗШИРЕНО  
ИЗВЛИЧАНЕ“

докторска програма „Информатика“, професионално направление 4.6.  
Информатика и компютърни науки

Съгласно заповед на № 75/27.03.2026 г. на Директора на ИИКТ-БАН  
съм определена да бъда член на научното жури, а съгласно протокол от  
проведеното първо заседание на 30.03.2026 г. съм избрана да подготвя  
рецензия. За целта съм получила всички документи, представени от  
кандидата по процедурата.

### **АКТУАЛНОСТ И ПОЗНАВАНЕ НА ИЗСЛЕДВАНИЯ ПРОБЛЕМ**

В дигиталното време в което живеем, данните играят ключова роля за  
определянето на бързи и достоверни решения. В този контекст,  
представеният дисертационен труд разглежда интересна и актуална  
тематика, свързана с генерирането на съдържание с разширено извличане.  
Подобни системите намират все по-широко приложение в области, където  
фактическата точност и проследимостта пряко влияят върху качеството на  
вземане на решения като здравеопазване, право и др. От представения обзор,  
както и описаните резултати може да се установи, че докторантът е добре  
запознат с естеството на изследваната проблематика.

### **АНАЛИТИЧНА ХАРАКТЕРИСТИКА**

Дисертационният труд е в общ обем от 214 страници. Структуриран е  
както следва: списък на таблици, списък на фигури, списък на уравнения,  
списък на алгоритми, речник на термини и съкращения, увод, 5 глави,

заключение – резюме на получените резултати, 3 приложения, библиография, списък на публикации, списък на цитирания, списък на проекти с участие, благодарности, декларация за оригиналност на резултатите.

На стр. 26 е определена целта на дисертационния труд „да се разработи рамка за оценка на генериране на данни чрез добавено търсене, която да поддържа решения за конфигурация на търсене, базирани на доказателства, за RAG системи с LLM с отворен код, с особен фокус върху конфигурацията на прага на сходство“, и са формулирани 3 изследователски въпроса: “*V1: Промяната на прага на сходство води ли до измерими промени в качеството на генериране? V2: Различават ли се ефектите на прага на сходство в различните езикови модели? V3: Сравними ли са диапазоните на праговете на сходство в различните области на знанието?*”

На стр. 27 са представени още 4 цели, които се преследват: „*Цел 1: Дефиниране и прилагане на основните компоненти на рамката за оценка чрез интегриране на три слоя. Цел 2: Установяване на критерии за избор на модел. Цел 3: Дефиниране на процедури за избор и изчисляване на показатели. Цел 4: Провеждане на контролирани тестове и анализи.*“ Дисертационният труд продължава с формализиране на 3 научно-приложни приноса. Тази структура значително се отличава от общоприетата структура на дисертационен труд, при която на базата на литературния обзор се формулира целта и задачите за постигането ѝ.

В Глава 1 се проследява развитието на Retrieval-Augmented Generation (RAG) от извличането на информация и обработката на естествен език, представя съществуващите RAG архитектури и характерните им недостатъци, разглежда подходи за оценяване и идентифицира липсите в литературата. Описани са различни направления на развитие и подходите за оценяване, като идентифицира селективността при извличането или точно избора на праг на сходство като фактор с недостатъчно изследвано влияние върху качеството на генерирането.

В Глава 2 е представен дизайн и архитектура на система за структурирана оценка на генериране, допълнено с извличане (PaSSER). Цели се интегрирано извличане, съобразено с праговете, многометрично

оценяване и регистриране на произход, подкрепено от блокчейн, в унифициран работен процес за контролирано експериментиране. Система за оценка на производителността за оценка и извличане на сходство реализира възпроизводимост за оценка на RAG, съобразена с праговете. Архитектурата интегрира три функционални слоя: интерфейс, базиран на браузър, позволяващ конфигуриране на експерименти и визуализация на резултатите, Python backend, координиращ операциите по извличане чрез векторно съхранение на ChromaDB и извеждане на езикови модели чрез Ollama API, и Antelope blockchain слой, осигуряващ регистриране на произхода без възможност за несанкционирано вмешателство за всички резултати от оценката.

В Глава 3 са описани компонентите, необходими за експерименталния анализ, а именно големи езикови модели с отворен код и метрики за оценка на качеството на генерираните отговори. Тук е дефиниран композитен показател за производителност и композитен показател, за стабилност на прага, използвани за агрегиране на набор от метрики.

В Глава 4 са представени проведените изследвания за установяване влиянието на прага на сходство върху качеството на генериране в RAG при седем езикови модела с отворен код и две тематични области, използвайки PaSSER.

По същество, Глава 5 прави обобщение на получените резултати, описани в предходните глави, систематизира научно-приложните приноси в разказна форма и очертава някои насоки за бъдеща работа.

#### **АВТОРЕФЕРАТ И АВТОРСКА СПРАВКА**

Представеният автореферат отразява съдържанието на дисертационния труд и съответства на изискванията на ЗРАСРБ и ППЗРАСРБ. От представената декларация за оригиналност, както и от публикациите, може да се определи, че описаните резултати са лично дело на автора. В допълнение трябва да се спомене, че резултатът от проверката за плагиатство, направена в сайта [strikeplagiarism.com](https://strikeplagiarism.com), показва следните резултати: Коефициент за сходство 1: 2.73 % и Коефициент за сходство 2: 0.81%.

## **ОЦЕНКА ЗА СЪОТВЕТСТВИЕ С МИНИМАЛНИТЕ НАЦИОНАЛНИ ИЗИСКВАНИЯ И С ДОПЪЛНИТЕЛНИТЕ ИЗИСКВАНИЯ ПО ЧЛ. 1А, АЛ. 2 ОТ ППЗРАСРБ**

Представени са общо 5 публикации, като всички те са на английски език и са в съавторство. Две от публикациите са индексирани в IEEE Xplore, а други 2 са с IF и индексирани в Scopus и WoS.

Съгласно ППЗРАСРБ, за изпълнение на минималните национални изисквания за получаване на ОНС „Доктор“ по професионално направление 4.6 „Информатика и компютърни науки“ се изисква наличие на 30 точки по Група показатели Г. Общата сума от точките за показателите от Група Г е 104 точки, което съществено превишава както специфичните изисквания на ИИКТ-БАН, така и на минималните национални изисквания за проф. напр. 4.6 „Информатика и компютърни науки“ за придобиване на образователната и научна степен „Доктор“.

### **ПРИНОСИ НА ДОКТОРАНТА**

Приносите на докторанта оценявам като научно-приложни, водещи до обогатяване на съществуващи знания, преформулирани както следва:

1. Представена е формулировка за композитен резултат от ефективност (CPS).
2. Представена е формулировка за прагово-зависим композитен резултат от ефективност (T-CPS).
3. Предложена е архитектура на система за структурирана оценка на генериране, допълнено с извличане.

### **КРИТИЧНИ ЗАБЕЛЕЖКИ И ПРЕПОРЪКИ**

Дисертационният труд е сравнително добре балансиран. Изложението е представено прекалено описателното за научно изследване вместо да се фокусира върху конкретните резултати и техния анализ. Това не намалява неговата научна стойност, но би било добре в бъдеще резултатите за бъдат представяни в по-строга форма. Има неправилно изписани математически формули като например (3.1) на стр. 89, (3.32) на стр. 107.

Към докторанта имам следните въпроси:

1. На стр. 20 в списъка със съкращения е записано „PaSSER (Platform for Systematic and Structured Evaluation of RAG)“, а на стр. 59 е записано „Performance Assessment System for Similarity Evaluation and Retrieval (PaSSER)“. Кое налага две различни интерпретации на една и съща абривиатура? Каква е разликата между платформа, рамка, система и браузър-базирана среда?
2. На какво се дължат различията в оптималните стойности на праговете при различните тематични области?
3. Защо не са използвани средствата на обяснимия изкуствен интелект, които позволяват интерпретация резултатите?

#### ЗАКЛЮЧЕНИЕ

Представените резултати по темата на дисертационния труд убедително показват, че докторантът **Мирослава Дончева Димитрова** притежава необходимите теоретични знания и практически умения в областта на информатиката и компютърните науки, както и доказани способности за самостоятелни научни изследвания. Представеният дисертационен труд отговаря на изискванията на Закона за развитието на академичния състав в РБългария, Правилника за неговото прилагане и на Правилника за специфичните условия за придобиване на научни степени и за заемане на академични длъжности в ИИКТ-БАН. Получените резултати по темата на дисертационното изследване са достатъчно основание да дам категорично положителна оценка на представения дисертационния труд и **предлагам на уважаемото Научно жури да присъди на Мирослава Дончева Димитрова образователната и научна степен „доктор“ по докторска програма „Информатика“, професионално направление 4.6. „Информатика и компютърни науки“.**

Дата 30.04.2026 г.

Под  
/проц

НА ОСНОВАНИЕ

ЗЗЛД