

РЕЦЕНЗИЯ

по процедура за защита на дисертационен труд на тема:

„Интелигентен агент за диалог на естествен език с отчитане на контекста“
(“Intelligent Context-Aware Natural Language Dialogue Agent”)

за придобиване на образователна и научна степен „доктор“

от

кандидат: **Момчил Емилов Хардалов,**

Област на висше образование: **4. Природни науки, математика и информатика**

Професионално направление: **4.6. Информатика и компютърни науки**

Докторска програма: **„Софтуерни Технологии – Откриване на знания“,**

в катедра: „Софтуерни технологии“

Факултет по математика и информатика (ФМИ)

Софийски университет „Св. Климент Охридски“ (СУ)

Рецензията е изготвена от: проф. д-р Олга Илиева Георгиева

катедра „Софтуерни технологии“, ФМИ – СУ в качеството ми на член на научното жури, съгласно Заповед № РД 38-600/28.10.2022 г. на Ректора на Софийския университет „Св. Климент Охридски“.

1. Обща характеристика на дисертационния труд и представените материали

Представеният за рецензиране дисертационен труд е оформен в общо 163 страници, които включват въведение, четири глави по същество и заключение и съдържа 18 фигури и 39 таблици. Част от основния труд са и две приложения в обем на 8 страници. Прави впечатление големият брой заглавия в библиографската справка - общо 350, всички издадени след 2015 година и представени в престижни научни издания – списания и сборници на конференции в тематиката на дисертационния труд. Научният труд е написан на английски език, а авторефератът е представен както на английски, така и на български език.

Първа глава е въвеждаща, в която са дефинирани целта и задачите на дисертационния труд.

Втора глава представя съдържателен аналитичен обзор върху съществуващите системи с разговорни агенти, класифицирани според характера и целите на диалога, техниките за разработка. Обзорът обръща специално внимание на съществуващите набори от данни в помощ на създаването на системи с поддържане на разговор с потребителите.

В трета глава е представен езиков модел със съвместен слот-стойност елемент за разбиране на естествен език, който надгражда известните модели BERT и неговата версия RoBERTa. Валидността и подобренията на предложениния модел е показана експериментално и е сравнена със съществуващи решения върху различни бази данни.

Четвърта глава е фокусирана върху методи, които използват допълнителна информация, за да се отговори на потребителски въпрос. Предложени са подходи за създаване на разговорни агенти чрез алгоритми за генериране на отговори с отчитане на контекста на разговора.

Пета глава анализира и предлага решения, които реализират автоматизиран разговор с управление на диалога чрез генериране на нови отговори, различни от използваните. Реализирана е задачата на многоезични модели за междуезичен трансфер без допълнително обучение.

Последната – шеста глава е заключителна. В нея са посочени научните приноси на дисертационния труд и са представени насоки за бъдещи изследвания.

Кандидатът демонстрира задълбочени познания и възможности да ползва широк набор от съвременните технологии в областта на дисертационния труд. Трудът е написан с голяма осведоменост по тематиката на изследването. Научните тези и анализи са дефинирани и доказани с необходимата задълбоченост и точност. Материалът е оформен старателно, добре структуриран и илюстриран.

Тези факти ме уверяват, че всички изисквания относно обема, структурата, съдържанието и оформлението на дисертационния труд са спазени.

В документите по конкурса са предоставени и други в подкрепа на работата на Момчил Хардалов – автореферат на български и английски език, копие на всички научни статии по процедурата, доклад, протокол и становище за сходство, административни документи, уверяващи в изпълнение на ЗРАСРБ.

2. Данни и лични впечатления за кандидата

Момчил Хардалов е завършил бакалавърска специалност „Компютърни системи и технологии“ в Техническия университет – София. Магистърската си степен е придобил

със специалност „Извличане на информация и откриване на знания“ във ФМИ СУ с отлична оценка.

Познавам кандидата от времето на зачисляването му през февруари 2017 г. като задочен докторант към катедра „Софтуерни технологии“. Преките ми впечатления са свързани с отличното му представяне на ежегодните атестации и презентации като докторант. Като член на комисията на докторантския минимум имах възможност да оценя неговата задълбоченост и високото ниво на знания в материята на докторантурата. Работата на Момчил Хардалов като хоноруван асистент към катедрата се отличаваше с отговорност, както и с умение да представя теми от областта на машинното самообучение по привлекателен за студентите начин.

3. Съдържателен анализ на научните и научно-приложните постижения на кандидата, съдържащи се в представения дисертационен труд и публикациите към него, включени по процедурата

Научните изследвания в дисертационния труд са изключително актуални. Получените научни и научно-приложни резултати са с теоретична и практическа стойност за изграждане на софтуерни системи с комуникационен канал с потребителите. Съществуването на такъв канал е вече наложила се съвременна тенденция във внедряваните системи. Това е съществен елемент, който отговаря на една съществуваща нужда - автоматизиране на работата по обслужване на клиентите. Дисертационният труд предоставя оригинални научни резултати и иновативни практически решения в отговор на тази нужда като се справя с предизвикателството за създаване на разговорен агент с разбиране на естествен език.

Приносите на дисертацията са в анализа, разработването и експерименталното доказване на методи, необходими в цялостната последователност от елементи за изграждане на комуникация с потребител - разбиране на естествен език и генериране на автоматизиран диалог с потребителя. Те могат да бъдат обобщени в следните научни и научно-приложни приноси:

- **Научни и научно-приложни приноси**

- Разработен е **нов езиков модел *Transformer-NLU*** за разбиране на естествен език, надграждащ известните BERT и RoBERTa, като за целта е вложен допълнителен слой за оценка на намерения, отчитане на допълнителни характеристики на думите и предсказано разпределение от намеренията. Подобрието на предложения модел

спрямо съществуващи решения е показано чрез експерименти върху два стандартни корпуса от данни. Приложението на модела е в изграждане на чатботовете за обслужване на клиенти за отговор на потребителски заявки.

- Предложена е **оригинална методология за извличане на подходящи пасаж**и от документи, свързани със специфичен контекст, **за избор на отговор** при четене с разбиране на въпрос и **последващо приложение към целеви език без допълнително обучение**. Предложеното решение се справя с проблемите, резултат на малък ресурс на данни за обучение и липса на контексти на въпроси в новия целеви корпус.

- Предложен е **нов метод за обучение от данни с шум** за целите на изграждането на разговорни агенти. Прилага се итеративно обновяване на етикетите на база на предсказани стойности на функцията на загубите.

- Предложена е **нова методология за подреждане на кандидати за отговори** за следваща реплика на разговорни агенти чрез използване на модел за оценка на качеството въпрос-отговор с въведена вероятностна оценка на отговорите на агента. Предложената методология реализира автоматизиран разговор с управление на диалога чрез генериране на нови отговори.

- **Приложни приноси:**

Трябва да бъдат подчертани и конкретни приложни приноси на дисертанта, които са от практическа стойност както за експерименталната работа, описана в дисертационния труд, така и в полза на бъдещи изследвания и технологични разработки в областта на разговорните агенти:

- Събран е корпус от въпроси на български език за четене с разбиране на въпрос с избираем отговор.

- Създаден е нов, междуетичен и многоезичен корпус за отговаряне на въпроси от гимназиални изпити (матури) на шестнайсет езика. Публикувани са множество данни и код в помощ на работата върху многоезични модели.

- Създаден е корпус за откриване на вече проверени твърдения – двойки кратки публикации и съответни статии за проверка на факти.

В обобщение, оценката за приносната стойност на дисертационния труд е изключително висока. Постигнатите резултати са значими и имат както добре обоснована теоретична, така и ясно показана практическа стойност. Трябва да се подчертае големият обем и качество на свършената работа.

4. Аprobация на резултатите

Списъкът на публикациите, които пряко отразяват резултатите на дисертационния труд, съдържа шест заглавия. Четири от публикациите са реферирани в SCOPUS, а три от тях са с SJR. Една от публикациите е в списание, реферирано в Web of Science, а останалите пет са представени в сборниците на престижни конференции. Във всички публикации кандидатът е на първо място. За всички заглавия са представени декларации за равно участие на съавторите. Считам, че представените публикации в достатъчна степен и по безспорен начин доказват представителността на изследванията и резултатите на дисертационния труд.

Справката за изпълнение на минималните национални изисквания показва, че публикациите защитават 108 точки от групата показатели Г при изискуеми 30 точки. Така кандидатът в много голяма степен надхвърля изискуемия минимум, определен от чл. 2б, ал. 2 и 3 на ЗРАСРБ и съответно на допълнителните изисквания на СУ за придобиване на образователна и научна степен „доктор“ в научната област и професионално направление на процедурата. Трябва да се подчертае, че публикациите, представени в подкрепа на дисертационния труд са оценени от специалистите в областта. Реферираните в SCOPUS от списъка с публикации по дисертацията са цитирани в 28 публикации, отчетени до деня на изготвяне на рецензията.

Няма доказано по законоустановения ред плагиатство в представения дисертационен труд и научни трудове по тази процедура.

5. Качества на автореферата

Авторефератът на дисертационния труд е представен на български език с 47 страници, а на английски език с 45. Той обобщава точно и в достатъчна степен на детайлност теоретичните изследвания, анализ и практическата им реализация, описани в дисертационния труд. Неговото съдържание следва представеното в дисертационния труд като са описани изследванията и приносните части на всяка от главите. Авторефератът отговаря на всички изисквания за изготвянето му.

6. Критични бележки и препоръки

Нямам критични забележки по отношение на постановката на задачите, анализ, точност и пълнота на получените резултати.

7. Заключение

След като се запознах с представените в процедурата дисертационен труд и придружаващите го научни трудове и въз основа на направения анализ на тяхната значимост и съдържащи се в тях научни и научно-приложни приноси, **потвърждавам**, че представеният дисертационен труд и научните публикации към него, както и качеството и оригиналността на представените в тях резултати и постижения, отговарят на изискванията на ЗРАСРБ, Правилника за приложението му и съответния Правилник на СУ „Св. Климент Охридски“ за придобиване от кандидата на образователната и научна степен „доктор“ в научната област 4. Природни науки, математика и информатика в професионално направление 4.6. Информатика и компютърни науки. В частност кандидатът удовлетворява минималните национални изисквания в професионалното направление и не е установено плагиатство в представените по конкурса научни трудове.

Въз основа на гореизложеното, **препоръчвам на научното жури да присъди на Момчил Емилов Хардалов образователна и научна степен „доктор“ в научна област 4. Природни науки, математика и информатика, професионално направление 4.6. Информатика и компютърни науки.**

18.01.2023 г.

Изготвил рецензията:

/проф. д-р Олга Георгиева/