

РЕЦЕНЗИЯ

по процедура за защита на дисертационен труд на тема:
*„Изкуствен интелект – дефиниция, реализация и последствия
Какво е това, как да го направим и какво ще правим, след като го направим?“*
за придобиване на образователната и научна степен „доктор“

от математик **Димитър Димитров Добрев,**
докторант на самостоятелна подготовка

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

Професионално направление: **4.5. Математика**

Докторантска програма: **„Математическа логика“**

Научен ръководител **доц. д-р Любомир Иванов**

Секция **„Алгебра и логика“**

Институт по математика и информатика към БАН

Рецензията е изготвена от: **акад. проф. д-рн Веселин Стоянов Дренски**, асоцииран член на ИМИ към БАН, член на научното жури съгласно Заповед № 108/29.04.2024 г. на Директора на ИМИ – БАН.

1. Обща характеристика на дисертационния труд и представените материали

Дисертационният труд е представен в два варианта – на български и английски език. Българският вариант съдържа 115 страници, от които 6 уводни страници, 103 страници основен текст, 6 страници заключение, завършващо с 92 цитирани заглавия. Основната част се състои от увод, завършващ с кратка справка за приносите, и три глави. Останалите представени материали съдържат стандартните за една процедура документи (автобиография със списък на публикациите, данни за докторантурата, като се започне от зачисляването и се свърши с решението на Научния съвет и заповедта на директора за състава на Научното жури, информационна карта на НАЦИД, списък от 5 публикации, на които се базира дисертационният труд и 20 свързани с труда други публикации, копия от тези 25 публикации, списък на цитирания на трудовете на кандидата, авторска справка за приносите на дисертационния труд и на петте статии, на които той се базира, автореферат на български и английски език).

Всички материали по дисертацията са качени със свободен достъп на страницата на ИМИ. Имам някои критични бележки към представянето на публикациите, имащи отношение към дисертацията. От петте статии, включени в дисертацията, само 2 са представени в оригинал. Една от другите 2 статии е на български, а кандидатът е представил превод на английски, а останалите две са представени в ръкописен вариант, въпреки, че оригиналите са достъпни в интернет. От останалите 20 публикации само 5 са представени в оригинал. При това, някои от останалите (например част от тези, представени в трудове на конференции) са представени във вариант, който е много по-дълъг от оригиналния. Би било разумно, ако някои от статиите не са със свободен достъп, техният оригинален

текст да бъде достъпен само за членовете на Научното жури, а ръкописните варианти да бъдат със свободен достъп.

2. Данни и лични впечатления за кандидата

Димитър Димитров Добрев е завършил Факултета по математика и информатика на СУ „Св. Климент Охридски“, специализация „Математическа логика и приложенията ѝ“ и е защитил дипломна работа на тема „Периодични зацикляния в Пролог“ под ръководството на проф. дмн Димитър Скордев. По време на следването си е бил на петмесечна специализация в Университета в Твенте, Холандия по програмата Tempus. От 1996 г. досега работи като математик в Института по математика и информатика към БАН, най-напред в секция „Математическа логика“, а след обединяването на секцията със секция „Алгебра“, в секция „Алгебра и логика“. Води занятия (лекции и упражнения) към катедра „Математическа логика и приложенията ѝ“ към ФМИ на СУ и в Нов български университет. Научните интереси на Димитър Добрев са в областта на изкуствения интелект и логическото програмиране, системи от рода на „Бонус-Малус“. Освен това, той е един от авторите на проект за движение на влаковете в метрото без машинисти и без да спират на всички станции. Димитър Добрев е бил част от колектива за транслитерацията и клавиатурните подредби, съобразени с българския език. Създател е на програмни и информационни продукти и има няколко регистрирани патента, както и заявки за патенти. Димитър Добрев има добре поддържана страница в интернет, от която човек може да получи подробна информация за цялостната му дейност. Освен това той е автор на книгата „Пророкът Гочо“, за която авторът обещава да допълни с нови глави, която е написана на колоритен език и за която има множество противоречиви оценки. Хвалебствените стигат до там, че това е съвременен вариант на „Бай Ганьо“. Димитър Добрев има и редица публични изяви по телевизията и други медии. Обикновено там го представят като писател и/или специалист по изкуствен интелект и много рядко като математик. В автобиографията си Димитър Добрев е представил списък от 21 научни публикации (за периода 2005 – 2022 г.) и 11 научно-популярни публикации (за 2000 – 2018 г.). От научните статии 2 са в *Mathematica Balkanica*, 2 са в *Serdica Journal of Computing*, 3 са в *International Journal “Information Theories and Applications”*, 1 е в *International Journal „Information Technologies and Knowledge“*, 5 са в трудове на конференции в България и Гърция, 8 са в препринтните бази arXiv и viXra. (В автобиографията статия [19] е дадена като препринт, но в списъка от 20 статии, имащи отношение към дисертацията, статията е излязла през 2021 г. в *Serdica Journal of Computing*.) Популярните статии са: 8 в *PC Magazine – Bulgaria* и по 1 в *eWeek Bulgaria*, *Култура* и *viXra*. Бих препоръчал на кандидата да подбира по-добре списанията, където публикува своите трудове, за да получи повече признание не само у нас, но и в чужбина.

В заключение ще отбележа, че понякога поведението на Димитър Добрев не е типично за математик, но въпреки това имам положително мнение за него и считам, че той е един полезен член на българската математическа колегия.

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

Тематика на дисертацията. Бих характеризирал изследванията, представени в дисертацията като изследвания в областта на изкуствения интелект, инспирирани от математическата логика. Кандидатът има 4 статии, реферирани в Zentralblatt für Mathematik, които съгласно Mathematics Subject Classification – MSC2020 са в раздела 68-XX Computer science, подраздели 68Txx Artificial intelligence (68T01 General topics in artificial intelligence и 68T05 Learning and adaptive systems in artificial intelligence) и 68Wxx Algorithms in computer science (68W01 General topics in the theory of algorithms). Аз бих причислил публикациите му също към 68T27 Logic in artificial intelligence.

Актуалност на разглежданите проблеми. Считаю, че в дисертацията се изучават и разглеждат важни и актуални задачи. В потвърждение на това ще отбележа, че в Zentralblatt für Mathematik са реферирани над 115000 публикации в 68Txx Artificial intelligence, над 7000 от тях са книги, а само от 2020 г. досега има почти 13000 публикации, от които над 640 са книги.

Научни приноси. По мнението на дисертанта, целта на неговата дисертация е да разсее две заблуди, свързани с изкуствения интелект (че ИИ е функция без памет и че ИИ е псевдонаука) и да зададе трите най-важни въпроса, свързани с тази тема: „Какво е ИИ?“, „Как да го създадем?“ и „Какви ще са последствията от създаването му?“

Според мене, така формулирано, твърдението за двете заблуди е пресилено. Що се отнася до това, че за повечето изследователи ИИ е функция без памет, би трябвало да констатираме, че просто теорията на ИИ има две подобласти – на ИИ без памет (тесен, но съвсем не фалшив ИИ, както се твърди в дисертацията) и ИИ с памет или глобален ИИ. Не съм убеден, че преобладаващото мнение е, че ИИ е псевдонаука, както се твърди в дисертацията. Мисля, че ИИ и в частност глобалният ИИ съвсем не са пренебрегнати както от обществото, така и от изследователите. В потвърждение ще отбележа две публикации, които се появиха тези дни в популярна българска новинарска страница в интернет. Там се отбелязват 70 години от смъртта на Алан Тюринг (7 юни 1954 г.), където се казва, че една от главните негови заслуги е, че е баща на ИИ. Другата публикация е интервю с доц. Мариана Тодорова и нейното участие в международна конференция в Университета по информационни науки и технологии в Нанджин, Китай, където учени от различни страни ще обсъждат именно бъдещето на човечеството в ерата на изкуствения интелект във всички сфери, включително и възможностите на генералния изкуствен интелект. Не съм съгласен и с твърдението: „Дори и сред специалистите по математическа логика има известен скептицизъм относно ИИ, а логиката е областта на математиката, която е най-близка и най-подходяща за създаването на ИИ. Може да се каже, че математиката е в основата на всички точни науки, а логиката е математиката на математиката. Всяка математическа дисциплина е изградена върху фундамента на логиката.“ (За мене твърдението за важността на математическата логика е инспирирано от известното изказване на Гаус:

„Математиката е кралица на науките, а теорията на числата е кралица на математиката.“) Аз бих изказал твърдението за скептицизма на логиците в по-мека форма: „Има специалисти по математическа логика, които не са уверени, че ИИ е свързан с математическата логика.“

Първата глава на дисертацията е озаглавена „Какво е изкуствен интелект?“. Първият риторичен въпрос, който задава дисертантът е: *Трябва ли ни да знаем какво е ИИ?* На този въпрос той отговоря просто: *Да, ако искаме да го открием, то определено ще е по-лесно да го намерим, ако знаем какво търсим.* Разбира се, тази постановка е добре известна на изследователите във всички области на познанието. Първият отговор на въпроса какво е ИИ е даден в статията на Тюринг от 1950 г. Тестът на Тюринг е на основата на игра с името „Imitation game“ (Игра на имитация). Мъж и жена се намират в една стая и си разменят бележки с водещия, който е в друга стая. Задачата на водещия е да определи кой от двамата е жената. Задачата на мъжа е да обърка водещия. Задачата на жената е да помогне на водещия да направи правилен избор. Тюринг задава въпроса какво ще стане, ако мъжът се замени с машина. Ще може ли водещият да направи правилен избор и ще се поддаде ли на машината да го излъже? Отговорът на този въпрос показва може ли машината да се явява като мислеща. В дисертацията това е формулирано така: *„Ако компютърът имитира човека достатъчно добре, така че да не можем да го отличим от човека, тогава това е ИИ.“* Тази дефиниция е добра за ИИ без памет, но дисертантът иска да даде дефиниция на ИИ с памет. В дисертацията се приема тезисът на Чърч, че всяко изчислително устройство може да се моделира с програма и се дава следната неформална дефиниция: **ИИ е такава програма, която в произволен свят би се справила не по-зле от човек.** Условието програмата да работи успешно в произволен свят е съществено. Обикновено конкретните разработки на ИИ работят в конкретен свят. Например, преди няколко години беше създаден тест на Тюринг за финанисти.

Jasmina Hasanhodzic, Andrew W. Lo, Emanuele Viola, Is It Real, or Is It Randomized?: A Financial Turing Test, arXiv:1002.4592v1 [q-fin.GN].

Експериментите с този тест показват, че човек умее да открива закономерности в случайни на пръв поглед събития. В случая, това е проверено с данни от финансовите пазари, т.е. разработката се отнася до конкретен свят.

Следващата стъпка в дисертацията е формализирането на дефиницията за ИИ. За мене е очевидно, че формализацията използва идеята за машината на Тюринг, но в конструкцията има няколко принципно нови моменти, а в разглежданията са преодоляни и технически трудности. В частност, уточнява се дадената по-рано дефиниция на Ернандес-Орайо и Миная-Койадо и нейното подобрене от Маркус Хутер, защото дефиницията на ИИ, дадена в дисертацията, не зависи от дължината на живота и от езика за описание на света. Разглеждат се различни варианти на ИИ – когато светът е детерминиран и е достатъчно просто устроен, когато е детерминиран, но е сложен и когато има елементи на случайност. Описват се алгоритми на стратегия, която е близко до най-добрата стратегия и които са най-подходящи за всеки един от случаите. Формулират се няколко твърдения, които не се доказват, но се дават аргументи, че би трябвало да са верни.

В дисертацията се отбелязва като признание към разработката, че „тази дефиниция получи голямо разпространение, защото в момента в Google на въпроса *Definition of Artificial Intelligence* първият резултат, който се появява, е тази статия.“ Направих проверка. Наистина, като се напише въпросът в Google първите два реда са: „Scholarly articles for Definition of Artificial Intelligence

A definition of artificial intelligence - Dobrev - Cited by 131“

Втората глава „Как да го направим?“ свежда задачата за създаването на ИИ към задачата за намирането на подходящия език за описание на света. За разлика от езиците за програмиране, които описват само изчислими функции, в дисертацията се разглежда език, който описва малко по-широк клас от функции. Идеята е да се създаде нов език, който да описва един конкретен свят и след това да се покаже, че този език е подходящ за описанието на произволен свят. Описанието на света се разглежда като структура, състояща се от много пластове и която в дисертацията е представена като пирамида, за която първият пласт е основата. В частност, дава се обобщение на понятието алгоритъм. Разглежданията се илюстрират с различни варианти на игра на шах – когато играчът играе срещу себе си и когато има двама играчи. Както в първата глава и тук разглежданията се формализират на език, който според мене е инспириран от разглеждания, типични за математическата логика и теорията на алгоритмите.

В третата глава „Какво ще правим, след като го направим?“ се разглеждат философски въпроси, свързани с последствията от създаването на ИИ. Един от въпросите, които се задава в края на дисертацията, е „Трябва ли технологията ИИ да бъде общодостъпна?“ Според дисертанта това е опасна технология и сериозните статии в областта на ИИ трябва да бъдат засекретени. Дискутират се и други въпроси: *Можем ли да не създаваме ИИ? Как се отнасят към тематиката сериозните списания? Кой трябва да има достъп до компютрите? Как ще изглежда животът ни след появата на ИИ?* и други от този род. Не съм съгласен с много от твърденията в тази глава на дисертацията, които според мене звучат екстремистки. Например, там се твърди, че за хартиените списания рецензията е неизбежна, защото хартията е ограничен ресурс и не може всеки да публикува каквото му скимне. Освен това, хартиените списания са една отживелица от миналото. В електронните списания рецензентите са нужни, но не за да казват „да“ или „не“, а за да оценяват статиите и да насочват читателите струва ли си тази статия да бъде прочетена. Тоест рецензентите са нужни, но не в качеството им на цензори, а в качеството им на критици. Друго спорно твърдение в дисертацията е, че след появата на ИИ животът ни ще бъде измамно лесен. Няма да има нужда да мислим за прехраната си, няма да ни се налага да работим, дори няма да е нужно да се забавляваме един друг, защото ИИ ще ни забавлява много по-добре, отколкото който и да е човек би ни забавлявал. *Аз мисля, че един средно интелигентен читател лесно ще отдели полезните неща от екстремистките.* Лично аз съм привърженик на мнението на Тюринг. Създавайки своя тест за проверка дали настреща имате машина или човек, на базата на който и днес стъпват експериментите с роботи, Тюринг създава правила и пръв казва на глас, че

компютрите не могат и не трябва да направят хората безполезни, а само да бъдат в тяхна помощ. Това е един хуманен поглед върху бъдещето, който вдъхва надеждата, че въпреки напредъка, машината никога няма да бъде човек.

Искам специално да отбележа нетрадиционния стил, използван в дисертацията. Обясненията са написани в достъпен за широката (включително нематематическа) публика стил, но който е далеч от стандартния стил за писане на труд в областта на математиката. Този начин на изложение е полезен за популяризирането на постиженията, но според мене не е най-добрият начин за писане на дисертация в областта на математиката, информатиката и техните приложения.

Друга характерна особеност на дисертацията е, че в нея няма нито едно съществено математическо твърдение, което да е доказано. За много от другите науки провеждането на дискусия върху даден проблем може да окаже важно значение за развитието на науката. В по-малка степен това е вярно и за математиката. Например, според широката математическа аудитория едно от най-важните постижения на съветския математик Андрей Тихонов е дадената от него през 1926 г. удачна дефиниция на произведение на топологични пространства. Разбира се, аз не искам да сравнявам дисертанта с Тихонов, но намирам, че въпросите, разглеждани в дисертацията, са важни и дискусията в нея е в полезна посока.

Приложение на получените резултати. Публикациите по дисертацията и свързаните с тематиката публикации на дисертанта са намерили отзвук в публикациите на други изследователи. В документацията е приложен списък от 119 цитати в монографии, статии, дисертации и дипломни работи на английски, немски, италиански, испански, португалски, словенски, хърватски, словашки, фински и турски.

За мене е очевидно, че дисертантът познава много добре резултатите в областта на своите изследвания. В потвърждение на нетрадиционния стил на изложение ще отбележа, че в дисертацията се цитират и материали, присъствието на които е малко странно в една дисертация по математика. Цитират се Айзък Азимов, Терминаторът, Матрицата, президентът Макрон, кучето на Павлов и други подобни.

4. Аprobация на резултатите

Както вече отбелязах, дисертацията се базира на 5 статии, публикувани през 2000 – 2023 г., а още 20 статии имат пряко или косвено отношение към дисертацията. Две от петте статии са публикувани в авторитетно българско списание (с впечатляващия обем от 50 и 38 страници), една е публикувана в International Journal "Information Content and Processing", една е популярна, една е в препринтна база. Имам само косвени данни за представянето на резултатите на семинари и конференции (от списъка на публикациите в трудове на конференции и от заседанията на семинара на секция „Алгебра и логика“).

а) научните трудове отговарят на минималните национални изисквания (по чл. 2б, ал. 2 и 3 на ЗРАСРБ) и съответно на допълнителните изисквания на ИМИ – БАН за придобиване на образователната и научна степен „доктор“ в научната област и професионално направление на процедурата. При минимални изисквания от 30 точки за група от показатели Г дисертантът е представил данни за 36 точки;

б) представените от кандидата резултати в дисертационния труд и научни трудове към него не повтарят такива от предишни процедури за придобиване на научно звание и академична длъжност;

в) няма доказано по законоустановения ред плагиатство в представения дисертационен труд и в научните трудове по тази процедура.

5. Качества на автореферата

Авторефератът е доста подробен (на 37 страници) и е в стила на дисертацията. Считам, че той дава ясна и адекватна представа за съдържанието и основните резултати на дисертацията. Справката за приносите отразява вярно приносите на дисертанта. Считам, че авторефератът и справката за приносите отговарят на всички изисквания за изготвянето им.

6. Критични бележки и препоръки

Нямам особени критични бележки, освен тези, отбелязани в предишните точки от рецензията ми. Искам специално да отбележа, че отбелязаните от мене неща не развалят общото ми положително отношение от резултатите и тяхното представяне в дисертацията.

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 научноприложни 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 съответните Правилници на БАН и ИМИ – БАН за придобиване от кандидата на образователната и научна степен „доктор“ в научната област 4. Природни науки, математика и информатика, професионално направление 4.5. Математика, докторантска програма: „Математическа логика“. В частност кандидатът удовлетворява минималните национални изисквания в професионалното направление и на допълнителните изисквания на ИМИ – БАН и не е установено плагиатство в представените по конкурса научни трудове.

Въз основа на гореизложеното, **убедено препоръчвам** на научното жури да присъди на Димитър Добрев Добрев образователната и научна степен „доктор“ в научна област 4. Природни науки, математика и информатика, професионално направление 4.5. Математика (докторантска програма „Математическа логика“).

11 юни 2024 г.

Изготвил рецензията:

(акад. проф. дмн Веселин Дренски)