

СТАНОВИЩЕ

от

проф. дн Александра Андреева Соскова

Факултет по математика и информатика, СУ

(член на научно жури, заповед № 109/29.04.2024 г. на Директора на ИМИ-БАН)

за дисертацията на *Димитър Димитров Добрев*

„Искусствен интелект – дефиниция, реализация и последствия“

представена за придобиване на научна степен „доктор“ в научна област 4. „Природни науки, математика и информатика“, професионално направление 4.5 „Математика“, научна специалност „Математическа логика“.

1. Обща характеристика на дисертационния труд и представените материали

Представеният дисертационен труд е на български език, в обем 115 страници, включващи списък от литература, съдържаща 92 заглавия. Дисертацията, с научен ръководител: доц. д-р Любомир Иванов, е в областта на Изкуствения Интелект, едно интензивно развиващо се научно направление. Дисертацията започва с увод, последван от три глави, в които авторът отговаря на въпросите: „Какво е ИИ?“, „Как да го създадем?“ и „Какви ще са последствията от създаването му?“ и завършва със заключение, публикации, свързани с дисертацията, декларация за оригиналност на резултатите и литература. Авторефератът е в обем от 37 страници и правилно отразява приносите на дисертанта.

Представените материали включват всички изискуеми от ЗРАСРБ и Правилника за приложението му документи, както и тези изискуеми от съответния Правилник на БАН. Приложената съгласно изискванията на НАЦИД справка (документ 8) е коректно направена и показва, че кандидатът удовлетворява минималните национални изисквания.

1. Данни и лични впечатления за кандидата

Димитър Добрев завършва през 1995, специалност Математика, специализация „Математическа логика и приложенията ѝ“ във Факултета по математика и информатика на Софийския университет. Дипломната му работа „Периодични зацикляния в Пролог“, е с научен ръководител проф. д-р Димитър Скордев. От 1996 г. е математик в секция Математическа логика (сега Алгебра и Логика) в Института по математика и информатика, БАН. Води часове във ФМИ, СУ по „Strawberry Prolog и Изкуствен Интелект“, „Математическа логика“, „Логическо програмиране“, „Дискретна математика“, „Дискретни структури“ и в Нов Български Университет по „Strawberry Prolog и Изкуствен Интелект“. Участва в 5 научно-изследователски проекти, има признати 6 патенти и заявка за 3 нови патента. Автор е на 21 научни публикации и на 11 научно-популярни.

Познавам Димитър Добрев от студентските му години. Винаги ме е впечатлявал с новите си идеи, винаги нестандартни. Отначало беше увлечението му по Пролог, направи програмна система „Strawberry Prolog“, която и досега той преподава във ФМИ, на студентите. Нашият екип по Логическо програмиране, в който и аз участвах,

използвахме неговата система успешно в упражненията. Следваше интересът му в създаване на транслитерация и клавиатурни подредби, която се внедри, после се беше запалил по схеми за движението на влаковете в метрополитена и системи от рода на „Бонус-Малус“. И последните години интересът му е изключително към дефиницията и основната структура на Общия Изкуствен Интелект. Като колега винаги е бил добронамерен и отзивчив. Участвал е активно с доклади на нашите семинари по математическа логика, и на различните конференции, организирани от катедрата „Математическа логика и приложенията й“. Участва като хоноруван преподавател в няколко четения на катедрата от много години.

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

Теорията на „общия“ изкуствен интелект се развива напоследък доста бурно. Интелигентността е твърде сложна, за да бъде описана с една теория; вместо това изследователите изграждат йерархия от теории, които я характеризират на множество нива на абстракция. На най-ниските нива на тази йерархия, невронните мрежи, генетичните алгоритми и други форми на изчисления, ни позволяват да разберем процесите на адаптация, възприятие и взаимодействие с физическия свят, които трябва да са в основата на всяка форма на интелигентна дейност. На по-високо ниво са логическият формален извод, дедукцията, индукцията, алгоритмите, моделите и много други начини на разум. Дизайнерите на експертни системи, интелигентни агенти и програми за разбиране на естествен език разпознават ролята на социалните процеси в създаването, предаването и поддържането на знания.

Изучаването на механичните или „формалните“ разсъждения започва с философи и математици в древността. Логическите изследвания водят до теорията на изчислимостта, създадена от Алън Тюринг, Алонзо Чърч, Стивън Клини, Емил Пост и др. Това, заедно с едновременните открития в кибернетиката, теорията на информацията и невробиологията, кара изследователите да обмислят възможността за изграждане на „електронен мозък“. Съвременните изследвания на изкуствен интелект (ИИ) започват в средата на 1950 г. Самият термин „изкуствен интелект“ е официално въведен от Джон Маккарти на семинара в Дортмунд през 1956 г.

Първото поколение изследователи на ИИ са убедени, че общият изкуствен интелект е възможен и че ще съществува само след няколко десетилетия. Въпреки това, в началото на 70-те години на миналия век става очевидно, че изследователите силно са подценили трудността на проекта. В началото на 1980 г. японският компютърен проект от пето поколение съживява интереса към ИИ. Но целите на компютърния проект от пето поколение никога не се изпълняват.

През първите десетилетия на 21-ви век достъпът до големи количества данни, по-евтини и по-бързи компютри са успешно приложени към много проблеми във всички сфери. Бумът на ИИ започва с първоначалното разработване на ключови архитектури и алгоритми, което довежда до развитие на големи езикови модели, показващи човешки черти на разсъждение, познание, внимание и креативност.

В дисертацията се разглежда един сложен модел на работата и действието на общия изкуствен интелект. Това е вид изкуствен интелект, който съответства или надминава човешките способности в широк спектър от задачи. За разлика от това, слабият ИИ е в състояние да реши един специфичен проблем, но му липсват общи познавателни способности.

В първата глава на дисертацията се разглежда неформална дефиниция на ИИ. Авторът я сравнява с известните в литературата и показва какви подобрения е направил. Дава се и строга математическа дефиниция на ИИ. Авторът подобрява дефиницията на ИИ, първоначално дадена от Hernández-Orallo, и съществено подобрена от Marcus Hutter, като дефиницията на ИИ на Добрев не зависи от дължината на живота и не зависи от езика за описание на света.

Във втората глава авторът представя един нов подход в изследването на ИИ. Това е *Event-Driven (ED)* подходът. Идеята, която стои зад *ED* подхода, е че моделът не трябва да поема цялата входно-изходна информация, а трябва да използва само важните събития. За описанието на светове най-разпространеният инструмент е *Markov decision process (MDP)*. Показва се, че *ED* моделът е естественото обобщение на *Markov decision process*. Авторът получава по-прост модел, който описва повече светове. Разглежда също разширен модел, при който състоянието знае всичко, с помощта на който въвежда интерпретация на събитията и на *Event-Driven* моделите. Езикът за описание на светове има предимството, че описанието може да се търси автоматично. Показва се, че езикът за описание на светове може чрез простите модули, от които се състои, да описва доста сложни светове с много агенти и сложни взаимоотношения помежду им. В дисертацията се описва един свят (играта шах). В статията Dobrev (2020a) е направена компютърна програма, която на базата на това описание емулира света.

В третата глава авторът дискутира въпросът за последствията от създаване на общ изкуствен интелект. С редица примери авторът предупреждава за опасности, които могат да се предотвратят. Разглежда въпроса: как ще изглежда нашият живот след появата на ИИ.

3. Аprobация на резултатите

По същество съдържанието на първа глава са статиите:

- Dobrev, D. (2000). AI - What is this. *PC Magazine - Bulgaria*, 11/2000, pp.12-13.
- Dobrev, D. (2022a). The AI Definition and a Program Which Satisfies this Definition. arXiv:2212.03184 [cs.AI].

Съдържанието на втора глава основно е представено в статиите:

- Dobrev D. (2022b). Language for Description of Worlds. Part 1: Theoretical Foundation. *Serdica Journal of Computing* 16(2), 2022, pp. 101-150.
- Dobrev D. (2023). Language for Description of Worlds. Part 2: The Sample World. *Serdica Journal of Computing* 17(1), 2023, pp. 17-54.

Съдържанието на третата глава е представено в статията:

- Dobrev, D. (2019c). AI Should Not Be an Open Source Project. *International Journal "Information Content and Processing"*, Volume 6, Number 1, 2019, pp. 34-48.

Авторът има още с 19 публикации и един патент, които са свързани с дисертацията, но не са част от текста на дисертацията. Например, понятието *Event-Driven* модел е въведено в Dobrev (2018), разширеният модел е въведен в Dobrev (2019a). В Dobrev (2017a) са разгледани тестовите събития. Как агентът може да влияе и на събитията, които не са негови действия е описано в Dobrev (2021b). Представил е две публикации, в списанията *Mathematica Balkanica, New Series*, индексирани в *Zentralblatt* и *Serdica Journal of Computing*, индексирани в *MathSciNet*.

Представени са 8 цитата по дисертацията, от които 3 в Scopus и 119 по темата, от които в 34 Scopus. в Google Scholar само статията Dobrev (2005a) има 127 цитата (116 без самоцитирание). Авторът е докладвал резултатите си на много конференции и семинари.

Научните трудове отговарят на минималните национални изисквания (по чл. 2б, ал. 2 и 3 на ЗРАСРБ) за придобиване на образователна и научна степен „доктор“ в научната област 4. „Природни науки, математика и информатика“ и професионално направление 4.5 „Математика“ („Математическа логика“). 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. Авторът декларира оригиналността на резултатите си.

4. Качества на автореферата

Авторефератът във вариантите на български и английски отразява коректно изложените в дисертационния труд резултати и е изготвен съгласно изискванията.

5. Критични бележки и препоръки

Дисертацията е написана в нетрадиционен стил. Според мен би могло да се оформи с повече математическа строгост. Това, което ми направи впечатление е, че се коментират различни нови понятия, преди те да бъдат дефинирани формално. За мен по-логично е обратното. От друга страна положителното е, че се чете леко и е достъпна до по-широка аудитория. Дават се доста конкретни примери, пак нетрадиционно, които обаче помагат да се разберат идеите. Но все пак остава впечатлението за научно-популярно четиво. В последната глава това е доста застъпено..

6. 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 БАН за придобиване от кандидата на образователната и научна степен „доктор“ в научна област 4. „Природни науки, математика и информатика“, професионално направление 4.5 „Математика“, научна специалност „Математическа логика“.

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

10.06.2024. г.

Изготвил становището:

проф., дн, Александра Соскова

(академична длъжност, научна степен, име, фамилия)