

Интелигентни агенти в медицината – философска перспектива

Ива Венциславова Георгиева

Институт по философия и социология, Българска академия на науката

ivavgeorgieva@gmail.com

Intelligent Agents in Medicine – A Philosophical Perspective

Iva Ventsislavova Georgieva

Institute of Philosophy and Sociology, Bulgarian Academy of Sciences

ivavgeorgieva@gmail.com

***Резюме:** Статията разглежда използването на изкуствените агенти в медицинската среда, като се вземат предвид както положителните, така и отрицателните ефекти от тяхното приложение в здравеопазването. Въвеждането на изкуствения интелект в сферата на предоставянето на грижи естествено поражда етични опасения, наблягайки на понятия като отговорен и морален изкуствен интелект. Проектирането на етични изкуствени агенти, способни да изпълняват задачи, традиционни за хората, поставя сложни дилеми, и води до заключението, че ние носим отговорност за тяхното правилно интегриране в критични области като здравеопазването и грижите. В този смисъл философската перспектива за изкуствените агенти в здравеопазването и грижите повдига важни въпроси относно тяхното приемане, регулиране и етични последици, като подчертава необходимостта от продължаващо философско изследване в сферата на приложение на изкуствените агенти в медицината.*

***Ключови думи:** изкуствени агенти, интелигентни агенти, изкуствен интелект, медицина, грижа, здравеопазване.*

***Abstract:** The article examines the use of artificial agents in the medical environment, considering both the positive and negative effects of their use in healthcare. The introduction of artificial intelligence into the realm of care delivery naturally raises ethical concerns, emphasizing concepts such as responsible and moral artificial intelligence. Designing ethical artificial agents capable of performing tasks traditional to humans poses complex dilemmas, and leads to the conclusion that we have a responsibility to correctly integrate them into critical areas such as*

health and care. In this sense, the philosophical perspective on artificial agents in health and care raises important questions about their acceptance, regulation, and ethical implications, highlighting the need for continued philosophical inquiry into the field of application of artificial agents in medicine.

Keywords: *artificial agents, intelligent agents, artificial intelligence, medicine, care, healthcare.*

Това изследване цели да обсъди едно особено и ново специално място на технологията в контекста на медицината и във връзка със скорошните развития в областта на изкуствения интелект (ИИ) като въвеждането на Large Language Models (LLMs) в широка масова употреба, която разраства сериозно потенциала им за специфична употреба като тази в контекста на медицината.

Намесата на ИИ в сферата на медицинската наука е нескорошен феномен, но последните тенденции, по-конкретно свързани с прогнозиране, диагностициране и терапия (Georgieva & Georgiev 2019) са изключително интересни с това, че показват едно ново място на технологията и нейната роля в медицината. Заедно с тази нова роля идват нови проблематични полета и съответно необходимостта от нови регулации, които трябва да отговорят бързо на също така скоростно развиващите се способности за въздействие и намеса от страна на ИИ в проблеми, които преди това не са били решавани с негова помощ.

Един основен елемент от намесата на технологиите в сферата на медицинските науки обаче се повлиява от агентите на ИИ, които могат да имат най-различна форма и роля в подпомагането на развитието на здравеопазването и поддържането на здравето и добруването, или така нареченото *wellbeing*. В този контекст, това изследване разчита на коректността на възприемане на този термин и избира да използва него вместо *wellness*, който има един по-широк смисъл и преминава отвъд терапевтичните приложения към едно по-всекидневно значение.

Агентите на ИИ, които се използват в сферата на здравеопазването в този посочен по-горе контекст, често са наричани *care agents*, тъй като те подпомагат пациенти, обекти на здравна грижа. Първото и най-важно нещо, което трябва да бъде уточнено във връзка с терминологичната яснота по въпроса за изкуствените агенти обаче е разграничението в названията на тези агенти и самите им типове, които се различават по функция, приложение

и т.н. Идеята за изкуствени агенти предполага, че те са интелигентни такива и заместват естествената помощ в медицински контекст или я допълват с разширените си възможности. И все пак понятията изкуствени агенти и интелигентни агенти не се припокриват напълно.

Термини като изкуствени хора (artificial humans (Ishiguro & Nishio 2007)) допълнително правят сложно дефинирането на понятието като насочват повече в посока трансхуманизъм. Това изследване обаче не иска да се съсредоточава в тази посока, а по-скоро в контекста на машините, роботи и андроиди, и софтуера или ИИ, които са заложили в тях, всички в медицинска среда и със съответната функция в нея.

В този смисъл изкуствените агенти могат да изпълняват много разнообразни функции, някои от които могат да бъдат обобщени в следните категории в сферата на здравеопазването:

- наблюдение
- изследване
- диагностика
- грижа
- помощник
- заместител
- лечение, и др.

Ако все пак се опитаме да дефинираме ролята на изкуствените агенти и интелигентните агенти според различните им приложения в медицината бихме казали, че те са съответно а) компютърна програма или система, която е проектирана да възприема своята среда, да взема решения и да предприема действия за постигане на конкретна цел или набор от цели, като освен това работи автономно, което означава, че не се контролира директно от човешки оператор, и б) въз основа на степента на възприемана интелигентност и способности се наблюдават агенти, които могат да бъдат разделени на пет типа или както следва: рефлексен агент, базиран на модел агент, базиран на цел агент, полезен агент и обучаващ се агент.

От по-горните определения се вижда, че понятията за изкуствен и интелигентен агент не се припокриват, а по-скоро надграждат, в смисъла, че изкуствените агенти притежават различен вид интелигентност и в този смисъл могат да бъдат сложени под общия знаменател на интелигентни агенти. Интересно е по какъв начин се разглежда идеята

за изкуствени агенти като например нови издания на електронни списания като *Computers in Human Behavior* обръщат своя интерес към тематиката на *Artificial Humans* в ново едноименно подразделение на изданието и определят фокуса на интереса към изкуствени агенти (социални роботи, базирани на алгоритми агенти и ИИ), взаимодействията им с хората, включително опосредствано от технологични импланти и директните интерфейси между човек и компютър, между хора, киборги и социални роботи, обръщайки внимание върху обществени аспекти на тези взаимодействия. Интересното тук обаче е, че не собствено философските, а по-скоро психологическите и социалните следствия от тези промени са главен фокус на изданието. Това показва отново една сериозна необходимост от мета-анализ на промените, които настъпват в сферата на медицината от цялостното навлизане на ИИ в нея и от значимостта на философското разсъждение по проблемите, които могат да възникнат, тъй като те са фундаментално засягащи концепциите за човешко здраве, грижа, етически измерения на намесата в природата на човека и много други, които попадат в рамките на едно философско разсъждение за медицината и изкуствените агенти в нея.

Един друг интересен въпрос се повлиява от понятието за изкуствени хора и появата на изкуствени агенти в медицината и това е самата връзката лекар-пациент, която е крайъгълен камък на „добрата“ медицинска практика и все пак тя изглежда се трансформира във връзка лекар-пациент-ИИ. Предизвикателството, пред което са изправени създателите на ИИ, регулаторите и политиците, е да определят стабилни стандарти и изисквания за този нов тип „лечебна връзка“, за да се гарантира, че интересите на пациентите и моралната цялост на медицината като професия не са фундаментално увредени от въвеждането на ИИ (Mittelstadt 2021).

По какъв начин се повлиява медицината от изкуствените агенти, или по-скоро кои сфери на медицинската практика търпят въздействие от тяхната поява? Malik et al. (2019) говорят за няколко сфери на въвеждане на изкуствени агенти – разработване на лекарства, диагностика на заболяването, мониторинг на здравето, дигитална консултация, анализ на здравни планове, хирургично лечение, управление на медицински данни, персонализирано лечение и др. В рамките на тези се вижда как агентите в дадени сфери подпомагат медицинския персонал и в други значително заместват персонала в работата му. Въпросът, който би вълнувал едно философско разсъждение в контекста на философия на медицината

и технологиите в нея е дали е възможно тези агенти напълно да заместят медицинския персонал в дадени сфери, кои са те, как ще стане регулирането на работата им и разрешаването на нововъзникналите етически проблеми и какви нови проблеми биха възникнали от такава стъпка ако тя изобщо е възможна на този етап на развитие на ИИ.

Собствено философските проблеми не спират до тук. Важно е да помислим върху фундаментални въпроси като изобщо възможно ли е медицинските експертиза и опит на практикуващите от дълго време лекари да бъдат пресъздадени в уменията на изкуствените агенти; възможно ли е тези агенти да бъдат натоварени с каквато и да е (*морална*) преценка при вземането на решения, когато има възможност за намеса единствено от тяхна страна; и най-вече биха ли били те в състояние да придобият някаква форма на емпатия, каквато е неизменна за лекарската професия, за да не превърнат тези усилия в опасно дехуманизиране на медицинското изкуство и т.нар. медицински нихилизъм (Васева-Дикова 2020).

Очевидно е, че ИИ, както се случва в момента чрез въздействието на LLMs в обществото, има силата да доведе до сериозна промяна в традицията на медицинските практики и съответно и в ежедневието ни. Без съмнение технологиите (и особено ИИ) водят до нови етични и социални проблеми поради това, че те подтикват (*nudge*) хората да взимат решения по нов начин и да променят поведението си (Capasso & Umbrello 2022). Новото при подобни технологии е, че те започват да доказват потенциала си като цифрови здравни инструменти за наблюдение, консултиране и предоставяне на съвети и насоки, по този начин като посредници между здравната система като цяло и обществеността. Това създава нов тип отговорност – не само очевидната за създателите им, но и за потребителите им като самоотговорност. В тази връзка дизайн на асистенти с ИИ се разработва под идеята за *AI for Social Good* и така нареченото отговорно подтикване на потребителите (Umbrello & van de Poel 2021).

В тази връзка други изследователи като Coeckelbergh (2015) питат доколко изобщо е (*етически*) приемливо използването на изкуствени агенти в здравеопазването. Това питане бива разширено в разсъждението за необходимостта от критерии за добра грижа и следващото от това обсъждане дали машините като изкуствени агенти, които поемат задачите свързани с грижа за пациенти могат да отговарят на тези критерии. Тук авторът набляга на термините за грижа, посредничество и поемане на отговорност (*care, agency, take over responsibility*) като водещи в разсъждението за мястото на изкуствените агенти в

здравеопазването. Макар самият той да няма възражения срещу използването на машини в медицината и здравеопазването, идеята те да функционират и да изглеждат като „изкуствени агенти“ е проблематична в много смисли (например бихме казали трябва ли те да изглеждат колкото е възможно повече като медицински персонал, т.е. хора или да бъде ясно разграничението им с човешките същества и те да бъдат отявлено разпознаваеми като машини, т.е. изкуствени) и ни насочва към фундаменталните проблемите в грижата за хората, които вече са налице, преди да навлязат визиите за грижи от страна на машини на сцената, които той определя като *машини (технологии) за грижи*.

Loveys et al. (2019) виждат изкуствените агенти в медицината, които могат да варират например от роботи до чатботове, като иновация в предоставянето на грижи, която може да намали самотата на пациентите чрез предоставяне на социална подкрепа и заместване на персонала, където той не би смогнал да предостави цялостна грижа. Подобно на взаимоотношенията между лекар и пациент, качеството на връзката на пациента с изкуствения агент също може да повлияе на ефективността на подкрепата, както и на ангажираността в грижите, т.е. да има двустранно въздействие.

Тези изследователи смятат, че включването на поведение за изграждане на привързаност като това, което се наблюдава при бозайниците, в обработката на невронни мрежи като част от разработките по възможностите на изкуствените агенти, може да подобри качеството на тази връзка между пациентите и изкуствените агенти и да увеличи ангажираността на последните. По този начин насърчават разработчиците на изкуствени агенти да включат дизайнерски идеи от еволюционната невропсихология в разработките на агенти, които са главно предназначени да облекчат самотата на пациентите.

Изглежда, че новите приложения на изкуствените агенти в медицината включват използването на мултиагентни системи в сферата на здравеопазването, което от своя страна създава нови приложения като персонализирани и социализирани системи за здравеопазване (Iqbal et al. 2016).

Алгоритми за машинно обучение (ML) за диагностични и прогностични цели създават сътрудничество между човешкия интелект, въплътен от медицинския персонал, и машинния интелект, осигурен от ИИ алгоритми и методологии. Много институции са се включили в това усилие: професионални организации (например Association for Computing Machinery (ACM)) и Institute of Electrical and Electronics Engineers (IEEE)), занимаващи се с

популяризирането на стандарти, правителствата в ролята им на определяне и приемане на разпоредби и законодателство, както и изследователи и експерти в областта на компютърните науки, правото и философията, които допринасят за дискусиата. Така предоставените насоки описват общи принципи, които трябва да се спазват, но не предоставят решения на реални проблеми. След формулирането на регламенти е необходимо да се посочи как те могат да бъдат приложени към конкретни случаи на прилагане на ИИ технология в областта на здравеопазването (Grasso & Pareschi 2022).

Важно е да се вземе предвид, че криза на индустрията за предоставяне на грижи, подсилена от демографски срив и фактори като пандемия, международни конфликти и икономическа стагнация и т.н., оправдават още повече въвеждането на изкуствените агенти в медицината и обуславят тяхното бързо интегриране, което може да доведе до не достатъчно регулирани принципи на употреба и намеса в грижата за хората.

Persson et al. (2022) насочват вниманието към тези въпроси и свързаните с тях проблеми, които могат да бъдат обобщени до проблема, че все още се знае малко за това как болногледачите са повлияни от въвеждането на роботи и изкуствени агенти като цяло в тяхната работна среда и най-вече от два показателя: а) до колко използването на такива заместители продължава във времето, и б) до каква степен те не създават допълнителна работа и необходимост от квалификация.

Анализът показва, че използването на роботи може да повлияе както на физическата, така и на психосоциалната работна среда, както по положителен, така и по отрицателен начин. Те се използват в грижите, за да намалят физическите и умствените изисквания към болногледачите, но всъщност могат да увеличат тяхното натоварване, понеже изискват специално внимание и научване на нови условия на работни среди. По този начин анализът показва, че роботите могат да подобрят качеството на работа, но те рядко работят като пряк път към повишена ефективност или времева ефикасност, а дори понякога забавят и утежняват вече изградената мрежа от участници в процеса на предоставяне на грижи.

Други автори (Wykowska et al. 2016) разглеждат експерименталните протоколи, които включващи изкуствени агенти, и по-специално хуманоидите роботи, като среда, която е изключително богата за мета-анализ, тъй като предоставя задълбочена и проникателна информация относно социалните когнитивни механизми в човешкия мозък, докато хората се опитват да интегрират изкуствените агенти в живота и работата си,

особено като помощници в процеса на грижа за други човешки същества. Използването на изкуствени агенти позволява манипулиране и контрол на различни параметри на поведение, външен вид и изразителност в едната от страните за взаимодействие (изкуствения агент) и за изследване на ефекта на тези параметри върху другата страна във взаимодействието (човека). В същото време използването на изкуствени агенти означава въвеждане на присъствието на изкуствени, но все пак човешки (създадени от човека за човека) системи в човешката социална сфера. Това позволява да се тестват по контролиран, но екологично валиден начин основните човешки механизми на социалното познание както на поведенческо, така и на невронно ниво и да се достигнат до важни изводи за природата на взаимодействие в такава среда, каквито изводи не биха били възможни без използването на такъв феномен като ИИ.

По-специално т.нар. социален ИИ (включващ например чатботове, виртуални герои, изкуствени асистенти или роботи) има голям потенциал за подобряване на здравеопазването и позволяване на по-персонализиран подход – напр. създаване на емпатия, сериозни игри за здравеопазване и изкуствени спътници за подпомагане на грижите за пациенти в болнична или домашна среда. Такъв нов подход към здравеопазването, а именно здравни грижи, базирани на ценности (*value-based health care*), вместо да възнаграждава обема на постигнатите резултати, има за цел да възнаграждава по-добри резултати по отношение на разходите, качеството и измерванията на резултатите от страна и във връзка с пациентите. Този метод също е допълнен от измервания на резултатите, докладвани от пациентите, които позволяват базиран на данни подход към здравеопазването за оценка на качеството на живот (Hindriks & Meyer 2019).

Бихме казали, е този подход е в основите на идеите, залегнали в комбинацията на понятията *wellbeing* и *technology* (AI, artificial intelligent agents) *in medicine*, който ще бъде обект на интерес и в бъдещи изследвания по темата. Естествено, тези теми неизменно водят и до въпроси, свързани с етическите и по-скоро морални измерения на намесата, присъствието и значението на изкуствените интелект и агенти в медицината и предоставянето на грижи.

Авторите отнасят този проблем до познат като проблем от сферата на *responsible AI* и *artificial morality*, анализът по отношение на работата на хората, които се занимават с етика на машините, показва, че тяхната задача не е в това да определят дали няма морално

значими разлики между изкуствените агенти и по-традиционните софтуерно базирани системи. Изкуствените агенти са способни да изпълняват задачи, които преди са можели да изпълняват само човешки същества и които традиционните софтуерно базирани системи не са способни да изпълняват. По-скоро определянето на това как те трябва да бъдат проектирани да изпълняват тези задачи ще изисква от нас да се справим с трудни нови етични въпроси. Но докато се занимаваме с тези въпроси, трябва да устоим на изкушението да антропоморфизираме изкуствени агенти. В края на краищата изкуствените агенти не са морални агенти – те не носят отговорност за това, което правят; ние обаче сме и носим такава отговорност, особено като работим за въвеждането им в такива сфери от живота ни като здравеопазване и грижа за хората, в които изконно човешкото е най-силно изявено.

В заключение можем да отправим няколко много важни въпроси от гледна точка на философия на медицината и технологията във връзка с темата за изкуствените агенти в здравеопазването и грижата за хората. Един от тях дали не въвеждаме твърде рано такива агенти в употреба, преди ефектите от тяхното присъствие в такива важни области да са достатъчно изследвани, особено в дългосрочен план и на едно по-дълбоко ниво. Не се ли приемат те твърде бързо като част от медицинските екипи и технологии в медицината като цяло без да бъдат добре регулирани като елементи от морални решения и човешки взаимоотношения? В този смисъл, философия на науката трябва да бъде в крак с тенденциите и да работи по проблемите, които възникват в хода на случването на нещата (Grozdanoff 2020).

Една начална точка е да се въвеждат обосновани и подкрепени от такъв анализ правила в контекста на медицината или както Lupton (2018) се опитва да изведе някакъв вариант на нов тип принципи на роботиката, позовавайки се на доклад на Engineering and Physical Sciences Research Council (EPSRC) (2011) 2011 и the Arts and Humanities Research Council (AHRC), както следва:

1. Robots should not be designed solely or primarily to kill or harm humans.
2. Humans, not robots, are responsible agents. Robots are tools designed to achieve human goals.
3. Robots should be designed in ways that assure their safety and security.

4. Robots are artefacts they should not be designed to exploit vulnerable users by working on emotional responses or dependency. It should always be possible to tell a robot from a human.

5. It should always be possible to find out who is legally responsible for a robot [...]. (Lupton 2018: 3).

В този смисъл Lupton набляга на последните точки от тези принципи като връщащи въпросите за отговорност и регулация към хората и така отново ни насочва към собствената работа в анализирането и предлагането на адекватни принципи на действие и взаимодействие на изкуствените агенти в човешката среда, които са повече от належащи в бързо развиващото се технологично съвремие, в което грижата за добруването на хората в съвместно съществуване с изкуствени агенти е повече от реална необходимост и предизвикателство.

Библиография:

Васева-Дикова, Ю. 2020. Медицински нихилизъм от Джейкъб Стегенга. *Философски Алтернативи*, 29 (6), 86-90. (Vaseva-Dikova, YU. 2020. Meditsinski nikhilizŭm ot Džeĭkŭb Stegenga. *Filosofski Alternativi*, 29 (6), 86-90.) [In Bulgarian]

Capasso, M., & Umbrello, S. 2022. Responsible nudging for social good: new healthcare skills for AI-driven digital personal assistants // *Medicine, Health Care and Philosophy*, 25(1), 11-22.

Coeckelbergh, M. 2015. Artificial agents, good care, and modernity // *Theoretical Medicine and Bioethics*, 36, 265-277.

Georgieva, I., & Georgiev, G. V. 2019. Redesign me: Virtual reality experience of the line of life and its connection to a healthier self // *Behavioral Sciences*, 9(11), 111.

Grasso, M. A., & Pareschi, R. 2022. Human and artificial collaboration for medical best practices // *Frontiers in Human Dynamics*, 4, 1056997.

Grozdanoff, B. D. 2020. REINFORCEMENT LEARNING SYSTEM FOR A MORAL AI AGENT // *Ethical Studies*, 5 (3).

Hindriks, K. V., & Meyer, J. J. C. 2019. Artificial Intelligence in Health Care and Medicine: A Personalized Approach // *Acta scientific medical sciences*, 3(10), 71-78.

Iqbal, S., Altaf, W., Aslam, M., Mahmood, W., & Khan, M. U. G. 2016. Application of intelligent agents in health-care // *Artificial Intelligence Review*, 46, 83-112.

Ishiguro, H., & Nishio, S. 2007. Building artificial humans to understand humans // *Journal of Artificial Organs*, 10, 133-142.

Loveys, K., Fricchione, G., Kolappa, K., Sagar, M., & Broadbent, E. 2019. Reducing patient loneliness with artificial agents: design insights from evolutionary neuropsychiatry // *Journal of medical Internet research*, 21(7), e13664.

Lupton, M. 2018. Some ethical and legal consequences of the application of artificial intelligence in the field of medicine // *Trends Med*, 18(4), 100147.

Malik, P., Pathania, M., & Rathaur, V. K. 2019. Overview of artificial intelligence in medicine // *Journal of family medicine and primary care*, 8(7), 2328.

Mittelstadt, B. 2021. The Impact of Artificial Intelligence on the Doctor-Patient Relationship. Commissioned by the Steering Committee for Human Rights in the fields of Biomedicine and Health (CDBIO) for the Council of Europe.

Persson, M., Redmalm, D., & Iversen, C. 2022. Caregivers' use of robots and their effect on work environment—a scoping review // *Journal of technology in human services*, 40(3), 251-277.

Umbrello, S., and I. van de Poel. 2021. Mapping value sensitive design onto AI for social good principles // *AI and Ethics*. 1 (3): 283–296.

Wykowska, A., Chaminade, T., & Cheng, G. 2016. Embodied artificial agents for understanding human social cognition // *Philosophical Transactions of the Royal Society B: Biological Sciences*, 371(1693), 20150375.