

DOI: <https://doi.org/10.17816/DD430348>

Этические проблемы внедрения искусственного интеллекта в медицине

М.И. Коньков

Российский национальный исследовательский медицинский университет имени Н.И. Пирогова,
Москва, Российская Федерация

АННОТАЦИЯ

Системы искусственного интеллекта (ИИ) высокоэффективны, что делает невозможным игнорирование их имплементации в медицинскую практику, но создаёт ряд этических проблем. Проблема «чёрного ящика» является базовой для философии искусственного интеллекта, но имеет свою специфику применительно к медицине. С целью изучения проблем внедрения ИИ в медицину проведён отбор релевантных статей за последние три года через Pubmed и Google Scholar по цитированиям и их анализ. Одной из центральных проблем внедрения ИИ в медицину является проблема «чёрного ящика» — отсутствие ясных и обоснованных принципов работы ИИ. Как врачи могут опираться на заключения ИИ, не имея достаточного количества данных, объясняющих принятие того или иного решения? А в случае неблагоприятного исхода (летального случая, тяжёлой травмы) на ком будет лежать ответственность за окончательно принятое решение? В рутинной практике медицинские решения основываются на комплексном подходе: понимании патофизиологии и биохимии, интерпретации полученных в прошлом результатов, клинических исследованиях и когортных наблюдениях. Искусственный интеллект может быть использован для построения плана диагностики заболеваний и лечения, но не предоставляет убедительных обоснований для принятия конкретных решений. Это и создаёт «чёрный ящик»: не всегда понятно, какую информацию ИИ считает важной для вынесения заключения, каким образом и почему ИИ приходит именно к такому выводу. Хуан Мануэль Дуран пишет: «Даже если мы заявляем, что понимаем принципы, лежащие в основе маркировки и обучения ИИ, всё равно сложно, а часто даже невозможно претендовать на понимание внутренней работы таких систем. Врач может интерпретировать или проверять результаты этих алгоритмов, но врач не может объяснить, как алгоритм пришёл к своим рекомендациям или диагнозу». Сегодня модели ИИ обучены распознавать микроскопические аденомы и полипы в толстой кишке. Но, несмотря на высокую точность, врачи всё ещё имеют недостаточно представлений о том, как ИИ дифференцирует различные виды полипов, а признаки, являющиеся ключевыми для постановки диагноза ИИ, остаются неясными для опытных врачей-эндоскопистов. Другой пример — биомаркеры колоректального рака, распознаваемые ИИ. При этом врач не знает, как алгоритмы определяют количественные и качественные критерии выявляемых биомаркеров для формулировки окончательного диагноза в каждом индивидуальном случае, т.е. возникает «чёрный ящик» в патологии процесса. Для того чтобы завоевать доверие врачей и пациентов, необходимо обеспечить расшифровку и пояснение процессов, лежащих в основе работы ИИ, описать, как последовательно, шаг за шагом происходит работа и формулируется конкретный результат. Хотя алгоритмы «чёрного ящика» нельзя назвать прозрачными, стоит рассмотреть возможность применения данных технологий в практической медицине. Несмотря на перечисленные проблемы, точность и эффективность решений не позволяет пренебрегать использованием ИИ, даже, наоборот, делает это использование необходимым. Возникающие проблемы должны служить основой для подготовки и обучения врачей умению работать с ИИ, расширения сфер применения и разработки новых методик диагностики.

Ключевые слова: искусственный интеллект; врачи; пациенты; «чёрный ящик».

КАК ЦИТИРОВАТЬ

Коньков М.И. Этические проблемы внедрения искусственного интеллекта в медицине // *Digital Diagnostics*. 2023. Т. 4, № 1 Supplement. С. 70–72.
DOI: <https://doi.org/10.17816/DD430348>

СПИСОК ЛИТЕРАТУРЫ

1. Holm E.A. In defense of the black box // *Science*. 2019. Vol. 364, N 6435. P. 26–27. doi: 10.1126/science.aax0162
2. Durán J.M., Jongsma K.R. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI // *J Med Ethics*. 2021 Mar 18;medethics-2020-106820. doi: 10.1136/medethics-2020-106820
3. Poon A.I.F., Sung J.J.Y. Opening the black box of AI-Medicine // *J Gastroenterol Hepatol*. 2021. Vol. 36, N 3. P. 581–584. doi: 10.1111/jgh.15384
4. Wang F., Kaushal R., Khullar D. Should Health Care Demand Interpretable Artificial Intelligence or Accept “Black Box” Medicine? // *Ann Intern Med*. 2020. Vol. 172, N 1. P. 59–60. doi: 10.7326/M19-2548
5. London A.J. Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability // *Hastings Cent Rep*. 2019. Vol. 49, N 1. P. 15–21. doi: 10.1002/hast.973
6. Yang G., Ye Q., Xia J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond // *Inf Fusion*. 2022. Vol. 77. P. 29–52. doi: 10.1016/j.inffus.2021.07.016
7. Quinn T.P., Jacobs S., Senadeera M., Le V., Coghlan S. The three ghosts of medical AI: Can the black-box present deliver? // *Artif Intell Med*. 2022. Vol. 124. P. 102158. doi: 10.1016/j.artmed.2021.102158
8. Handelman G.S., Kok H.K., Chandra R.V., et al. Peering Into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods // *AJR Am J Roentgenol*. 2019/ Vol. 212, N 1. P. 38–43. doi: 10.2214/AJR.18.20224
9. Мелдо А.А., Уткин Л.В., Трофимова Т.Н. Искусственный интеллект в медицине: современное состояние и основные направления развития интеллектуальной диагностики // *Лучевая диагностика и терапия*. 2020. Т. 11, № 1. С. 9–17. doi: 10.22328/2079-5343-2020-11-1-9-17
10. Малых В.Л. Системы поддержки принятия решений в медицине // *Программные системы: теория и приложения*. 2019. Т. 10, № 2 (41). С. 155–184. doi: 10.25209/2079-3316-2019-10-2-155-184

DOI: <https://doi.org/10.17816/DD430348>

Ethical issues of implementing artificial intelligence in medicine

Maxim I. Konkov

N.I. Pirogov Russian National Research Medical University, Moscow, Russian Federation

ABSTRACT

Artificial intelligence (AI) systems are highly efficient. However, their implementation in medical practice is accompanied by a range of ethical issues. The “black box” problem is basic to the AI philosophy, although having its own specificity in relation to medicine. A selection of relevant papers for the last three years by citations and their analysis through PubMed and Google Scholar search engines was conducted to study the problems of the AI implementation in medicine. One of the central problems is that the algorithms to justify decisions are still unclear to doctors and patients. The lack of clear and reasonable principles of AI operation is called the “black box” problem. How can doctors rely on AI findings without enough data to explain a particular decision? Who will be responsible for the final decision in case of an adverse outcome (death or serious injury)? In routine practice, medical decisions are based on an integrative approach (understanding of pathophysiology and biochemistry and interpretation of past findings), clinical trials and cohort studies. AI may be used to build a plan for disease diagnosis and treatment, while not providing a convincing justification for specific decisions. This creates a “black box”, since the information that the AI considers important for making a conclusion is not always clear, nor is it clear how or why the AI reaches that conclusion. Thus, Juan M. Durán writes, “Even if we claim to understand the principles underlying AI annotation and training, it is still difficult and often even impossible to understand the inner workings of such systems. The doctor can interpret or verify the results of these algorithms, but cannot explain how the algorithm arrived at its recommendations or diagnosis”. Currently, AI models are trained to recognize microscopic adenomas and polyps in the colon. However, doctors still have insufficient understanding of how AI differentiates between different types of polyps despite the high accuracy, and the signs that are key to making an AI diagnosis remain unclear to experienced endoscopists. Another example is the biomarkers of colorectal cancer recognized by AI. The doctor does not know how algorithms determine the quantitative and qualitative criteria of detectable biomarkers to formulate a final diagnosis in each individual case, i.e., a “black box” of process pathology emerges. For the trust of doctors and patients to be earned, the processes underlying the work of AI must be deciphered and explained, describing how it is done sequentially, step by step, and a specific result is to be formulated. Although the “black box” algorithms cannot

Received: 15.05.2023

Accepted: 05.06.2023

Published Online: 10.07.2023

be called transparent, the possibility of applying these technologies in practical medicine is worth considering. Despite the above problems, the accuracy and efficiency of solutions does not allow to neglect the use of AI. On the contrary, this use is necessary. Emerging problems should serve as a basis for training and educating doctors to work with AI, expanding the scope of application and developing new diagnostic techniques.

Keywords: artificial intelligence; doctors; patients; “black box”.

FOR CITATION

Konkov MI. Ethical issues of implementing artificial intelligence in medicine. *Digital Diagnostics*. 2023;4(1S):70–72. DOI: <https://doi.org/10.17816/DD430348>

REFERENCES

1. Holm EA. In defense of the black box. *Science*. 2019;364(6435):26–27. doi: 10.1126/science.aax0162
2. Durán JM, Jongsma KR. Who is afraid of black box algorithms? On the epistemological and ethical basis of trust in medical AI. *J Med Ethics*. 2021 Mar 18:medethics-2020-106820. doi: 10.1136/medethics-2020-106820
3. Poon AIF, Sung JJY. Opening the black box of AI-Medicine. *J Gastroenterol Hepatol*. 2021;36(3):581–584. doi: 10.1111/jgh.15384
4. Wang F, Kaushal R, Khullar D. Should Health Care Demand Interpretable Artificial Intelligence or Accept “Black Box” Medicine? *Ann Intern Med*. 2020;172(1):59–60. doi: 10.7326/M19-2548
5. London AJ. Artificial Intelligence and Black-Box Medical Decisions: Accuracy versus Explainability. *Hastings Cent Rep*. 2019;49(1):15–21. doi: 10.1002/hast.973
6. Yang G, Ye Q, Xia J. Unbox the black-box for the medical explainable AI via multi-modal and multi-centre data fusion: A mini-review, two showcases and beyond. *Inf Fusion*. 2022;77:29–52. doi: 10.1016/j.inffus.2021.07.016
7. Quinn TP, Jacobs S, Senadeera M, Le V, Coghlan S. The three ghosts of medical AI: Can the black-box present deliver? *Artif Intell Med*. 2022;124:102158. doi: 10.1016/j.artmed.2021.102158
8. Handelman GS, Kok HK, Chandra RV, et al. Peering Into the Black Box of Artificial Intelligence: Evaluation Metrics of Machine Learning Methods. *AJR Am J Roentgenol*. 2019;212(1):38–43. doi: 10.2214/AJR.18.20224
9. Meldo AA, Utkin LV, Trofimova TN. Artificial intelligence in medicine: current state and main directions of development of the intellectual diagnostics. *Diagnostic radiology and radiotherapy*. 2020;11(1):9–17. (In Russ). doi: 10.22328/2079-5343-2020-11-1-9-17
10. Malykh V. Decision support systems in medicine. *Program Systems: Theory and Applications*, 2019;10(2(41)):155–184. (In Russ). doi: 10.25209/2079-3316-2019-10-2-155-184

ИНФОРМАЦИЯ ОБ АВТОРЕ

Коньков Максим Игоревич;

ORCID: <https://orcid.org/0009-0002-2803-1020>;

e-mail: konkovmaksim18@gmail.com

AUTHOR'S INFO

Maxim I. Konkov;

ORCID: <https://orcid.org/0009-0002-2803-1020>;

e-mail: konkovmaksim18@gmail.com