УДК 004.89

Объяснимый искусственный интеллект как инструмент взаимодействия врача, пациента и технического специалиста

Власова Вита Владимировна(*) vlasova.v@bmstu.ru
SPIN-код: 4308-3080

Ляпунцова Елена Вячеславовна lev86@bmstu.ru
SPIN-код: 3500-4160

МГТУ им. Н.Э. Баумана, Москва, Россия

Аннотация. *Рассмотрены ключевые проблемы объяснимости искусственного интеллекта (ИИ) в медицине. Обозначена необходимость согласования интересов и информационных потребностей трех сторон — медицинских работников, пациентов и разработчиков технологий. Показано, что объяснимость становится не только технологическим, но и этическим и коммуникативным условием доверия. Особое внимание уделено адаптации объяснений к уровню подготовки пользователя и роли ИИ в клинической практике.*

Ключевые слова: *объяснимый ИИ, медицина, доверенный ИИ, взаимодействие, интерпретируемость, пациент – врач – разработчик*

Explicable artificial intelligence as a tool for interaction between a doctor, patient and a technician

Vlasova Vita Vladimirovna(*) vlasova.v@bmstu.ru
SPIN-code: 4308-3080

Lyapuntsova Elena Vyacheslavovna lev86@bmstu.ru
SPIN-code: 3500-4160

BMSTU, Moscow, Russia

Abstract. The key problems of explainability of artificial intelligence (AI) in medicine are considered. The need to coordinate the interests and information needs of three parties — medical professionals, patients and technology developers — is outlined. It is shown that explainability becomes not only a technological, but also an ethical and communicative condition of trust. Special attention is paid to adapting explanations to the user's level of training and the role of AI in clinical practice.

Keywords: explicable AI, medicine, trusted AI, interaction, interpretability, patient – doctor – developer

Введение. Медицинские ИИ-системы демонстрируют высокую эффективность в диагностике, прогнозировании и поддержке принятия решений, но остаются «черным ящиком» для пользователей. Это создает противоречие между точностью и интерпретируемостью, что особенно критично в здравоохранении. Объяснимый ИИ (XAI) становится условием доверия и гарантией

безопасного внедрения технологий [1]. Важной задачей является выстраивание междисциплинарного взаимодействия между врачами, пациентами и разработчиками.

Методы и материалы; результаты. Современные методы (LIME, SHAP, Attention Maps) позволяют локально или частично интерпретировать решения моделей, но требуют адаптации под мультимодальные медицинские данные, включающие изображения, тексты и временные ряды [2]. При этом, врачи нуждаются в клинически релевантных объяснениях: какие признаки были ключевыми, как они соотносятся с диагнозом, стадией заболевания и возможными терапевтическими стратегиями. Это облегчает интеграцию ИИ в практику и снижает риск врачебных ошибок. Также, пациенты должны понимать логику принятия решений простым языком [3]. Понятные визуализации, сценарии «что, если» и сравнение альтернативных вариантов лечения повышают уровень информированного согласия и уменьшают тревожность.

Обсуждение. Новизна подхода заключается в рассмотрении объяснимости систем ИИ как точки взаимодействия трех сторон [4, 5]. Для врача объяснимость — это инструмент профессиональной ответственности и проверки корректности алгоритмов. Для пациента — гарантия прозрачности и участия в принятии решений, укрепляющая доверие к медицинской системе. Для разработчика — условие успешного внедрения технологий и соответствия нормативным требованиям. Такой междисциплинарный взгляд формирует «замкнутый цикл взаимодействия»: ИИ → объяснение → врач → пациент → → обратная связь → доработка системы. Кроме того, объяснимость способствует выявлению системных ошибок и предвзятостей в алгоритмах, что критически важно для справедливости и безопасности медицинской помощи.

Заключение. Объяснимый ИИ в медицине выходит за рамки технической задачи и становится основой доверия, безопасности и этики. Таким образом, объяснимый ИИ становится связующим звеном в экосистеме «врач – пациент – разработчик», обеспечивая ответственное применение технологий в здравоохранении.

Список источников

[1] Карпов О.Э., Андриков Д.А., Максименко В.А., Храмов А.Е. Объяснимый искусственный интеллект для медицины. *Врач и информационные технологии*, 2022, № 2, с. 4–11.

[2] Шевская Н.В. Объяснимый искусственный интеллект и методы интерпретации результатов. *Моделирование, оптимизация и информационные технологии*, 2021, № 9(2). https://doi.org/10.26102/2310-6018/2021.33.2.024

[3] Исаков А.О., Гусарова Н.Ф., Добренко Д.А., Голубев А.А. Объяснимость поведения агентов в системах поддержки принятия врачебных решений. *Экономика. Право. Инновации*, 2024, № 4, с. 50–59.

[4] Doshi-Velez F., Kim B. Towards A Rigorous Science of Interpretable Machine Learning. Arxiv preprint:1702.08608, 2017. https://doi.org/10.48550/arXiv.1702.08608

[5] Samek W., Montavon G., Vedaldi A., Hansen L.K., Müller K.R. Explainable AI: Interpreting, Explaining and Visualizing Deep Learning. Springer, 2019. https://doi.org/10.1007/978-3-030-28954-6