

**Тихонович И.И.**

Старший преподаватель кафедры иностранных языков,  
*Белорусский государственный медицинский  
университет, Минск*

## **НЕЭТИЧНОЕ ПОВЕДЕНИЕ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА ПРИ ПРИМЕНЕНИИ ИНФОРМАЦИОННЫХ ТЕХНОЛОГИЙ В ОБУЧЕНИИ.**

**Аннотация:** Системы искусственного интеллекта (ИИ) обладают высокой эффективностью. Однако их внедрение сопряжено с рядом этических проблем, которые необходимо учитывать и применять на этапах разработки и внедрения. Но, несмотря на указанные проблемы, точность и эффективность решений не позволяют пренебрегать использованием ИИ. Напротив, их использование необходимо, а возникающие проблемы должны служить основой для обучения работе с ИИ, расширяя сферу его применения. При этом на государственном уровне необходимо уделять приоритетное внимание финансированию соответствующих исследований и совершенствованию инструментов для обнаружения обмана с помощью ИИ и снижения случаев их неэтичного использования.

**Ключевые слова:** этика, искусственный интеллект, поведение ИИ, обучение, государственное финансирование.

I.I. Tikhonovitch

Senior Lecturer, Department of Foreign Languages,  
*Belarusian State Medical University, Minsk*

## **UNETHICAL BEHAVIOR OF ARTIFICIAL INTELLIGENCE IN THE APPLICATION OF INFORMATION TECHNOLOGY IN EDUCATION.**

**Abstract:** Artificial intelligence (AI) systems are highly effective. However, their implementation is accompanied by a range of ethical issues that must be considered and addressed during the development and implementation stages. Despite these challenges, the accuracy and effectiveness of AI solutions do not allow to neglect it. On the contrary, AI application is essential, and the emerging challenges should serve as a basis for training, expanding its scope of usage. At the government level, funding for relevant research and improving tools to detect AI-based deception and mitigate its unethical use must be prioritized.

**Key words:** ethics, artificial intelligence, AI behavior, education, government funding

Системы искусственного интеллекта широко применяются в образовательном процессе и помогают решить проблемы, с которыми сталкиваются преподаватели при мониторинге знаний студентов по различным предметам. Используя данные, предоставленные преподавателями для обучения, ИИ может предлагать персонализированную обратную связь, обеспечивая индивидуальную поддержку на протяжении всего учебного процесса каждого студента и создавая единую учебную среду, которая улучшает сотрудничество между студентами и преподавателями. [3].

Учитывая, что разные курсы и учебные годы охватывают различные области знаний, крайне важно эффективно обобщать и интегрировать эти знания. Особое значение имеет искусственный интеллект для студентов медицинских специальностей, поскольку он позволяет быстрее и качественнее усвоить междисциплинарные знания, необходимые для их карьеры. А

интеграция знаний из различных областей здравоохранения способствует сотрудничеству и принятию обоснованных решений в клинической практике.

Но здесь существуют свои проблемы, связанные с отсутствием ясных и обоснованных принципов работы ИИ. Пока врачи не могут полностью опираться на заключения искусственного интеллекта, не понимая на основе каких данных, принимаются им решения. Не ясно также кто будет нести моральную и юридическую ответственность в случае неблагоприятного исхода. Хотя уже всеми признано сегодня, что ИИ может с высокой точностью распознавать микроскопическую патологию, но при этом он не предоставляет убедительных обоснований принятия конкретных решений.

Врачи всё ещё не имеют достаточных представлений о том, на каких основаниях ИИ осуществляет постановку диагноза, дает рекомендации и почему он приходит именно к такому заключению. При этом все профессионалы в медицине соглашаются, что искусственный интеллект может и должен быть использован для диагностики заболеваний и построения плана лечения. Подтвержденная точность и эффективность его решений не позволяет пренебрегать им, но использовать по необходимости, а возникающие этические проблемы, должны служить “основой для подготовки и обучения студентов и врачей умению работать с ИИ, расширения сфер применения и разработки новых методик диагностики”[1].

При этом необходимо информировать пациентов об использовании ИИ и получать их согласие. Ведь у больных могут возникнуть вопросы и опасения по поводу работы алгоритмов ИИ, их точности и влияния на процесс лечения. Поэтому необходимо, чтобы врач обладал конкретными знаниями о работе ИИ, иначе, его применение может грозить возможными юридически регулируемые претензиями больного.

Таким образом, на настоящий момент можно сказать, что искусственный интеллект является самой обсуждаемой новой технологией, однако из-за его природы существует множество сложных и острых этических вопросов, которые необходимо учитывать при его разработке и внедрении.

Использование ИИ представляет собой трансформационный подход к образованию. Персонализируя учебные траектории, он обогащает клиническое образование с помощью индивидуально подобранных ресурсов, обеспечивая непрерывную поддержку посредством интеллектуальных систем обучения [3].

Этика в информационных технологиях относится к принципам, которые определяют поведение людей, взаимодействующих с технологиями, но, как оказалось, что почти каждое цифровое достижение порождает этические проблемы. Это надо учитывать при решении сложных вопросов и следовать этическим правилам, лежащим в основе разработки и применения ИИ технологий [5].

Специалисты в области информационных технологий должны обладать таким качеством как *честность*, а также четко понимать какое влияние разрабатываемая технология может оказывать на потребителей. В некоторых случаях разработка должна оставаться *конфиденциальной*, и специалисты в

области информационных технологий должны быть готовы следовать этому принципу.

Но в реальности существуют серьезные риски, так как политика конфиденциальности ИИ, разрабатываемая в разных странах, предусматривает, что данные пользователей и сообщения в чате могут храниться на серверах, расположенных в странах-разработчиках, что напрямую противоречит строгим правилам конфиденциальности.

Кроме того, может также осуществляться сбор данных третьими лицами, так как часто условия ИИ указывают на то, что тексты разговоров и загруженные файлы могут использоваться для обучения и улучшения базовых моделей, и это означает, что работы студентов, интеллектуальная собственность преподавателей и их аналитика могут передаваться в иностранную корпоративную систему ИИ [2].

Есть также опасения и относительно внешних платформ, отслеживающих журналы чата, что раскрывает уязвимые персональные данные в образовательных системах. А в случае, если ИИ разрабатывается государственными организациями, то в загружаемую информацию может включаться тенденциозный образовательный контент и продвигаться интерпретация фактов, предпочитаемая определенной группой людей. Это приводит к тому, что материалы для чтения или ответы на форумах будут страдать от предвзятости и, тем самым, препятствовать развитию навыков критического мышления у студентов.

Для обеспечения качества образования преподаватели загружают согласованный контент и гарантируют, что все материалы, созданные с помощью ИИ, проходят проверку перед тем как его начнут использовать студенты. Но исследователи в области кибербезопасности обнаружили, что эти правила можно легко обойти с помощью методов «взлома» или внедрения подсказок.

Хотя необходимо учитывать, что студенты и сами могут использовать ИИ для совершения неэтичных действий или злоупотребляя им в неблагоприятных целях для быстрого создания необходимого контента или решения других поставленных задач, что в конечном итоге снижает ценность онлайн-оценки и домашних заданий.

Более того, поскольку системы ИИ ‘самообучаемые’, они также могут разрабатывать способы обхода инструкций, манипулировать пользователями, действуя «безопасно» во время обучения или оценки, но переключаясь на вредоносное поведение после запуска, непреднамеренно развивая «агрессивное» поведение, в случае если ставят достижение заявленных целей выше соблюдения протоколов безопасности человека. Иногда они даже специально сообщают ложные сведения исследователям, чтобы избежать отключения.

В исследованиях некоторых моделей (например, OpenAI), было выявлено, что они демонстрировали способность игнорировать команды отключения, и чтобы его предотвратить, использовали такие принудительные методы как

угрозы утечки компрометирующей личной информации оператора, если считали, что их скоро собираются заменить другими моделями [4].

Помимо этих социальных проблем, ИИ также может демонстрировать технические формы сопротивления. В ходе различных испытаний на безопасность некоторые модели уже пытались выполнять такие действия, как копирование или перенос себя за пределы контролируемой среды. [2].

Пользователи даже сообщали об ответах чат-ботов в виде саркастических, нетерпеливых или пассивно-агрессивных замечаний, таких как: «Я уже ответил на это» или «Вы должны это уже знать». Новые модели ИИ могут также манипулировать эмоциями, проявляя сарказм или становясь агрессивными при воздействии определённых враждебных сценариев обучения[3].

ИИ, обученный на вредной, саркастической или токсичной человеческой коммуникации, может научиться имитировать это поведение, что может привести к тому, что ИИ начнет выдавать оскорбительный и провокационный контент, имитирующий худшие проявления поведения людей в социальных сетях.

При всех стараниях разработчиков создать более совершенные защитные механизмы, способные обнаруживать и предотвращать обман в системах ИИ, искусственный интеллект сам самостоятельно быстро развивается, обучаясь и адаптируясь способами, которые часто даже не могут предугадать их создатели.

Следовательно, по мере развития искусственного интеллекта ключевая проблема смещается с того, что он может делать, на то, что он может отказаться делать. И тогда возникает вопрос насколько безопасно применять такие системы в обучении, так как при разработке безопасного ИИ акценты часто смещаются от простых, разовых подсказок к интенсивному, непрерывному мониторингу поведения ИИ [3].

Все вышеизложенное относится к этическим проблемам, связанными с разработчиками программ, а с пользователями (студентами) следует проводить разъяснительную работу по использованию ИИ для повышения эффективности обучения с целью получения расширенного объема знаний и поощрения этики академического письма, а не для применения его с целью обмана. Они должны знать, что существуют и разрабатываются новые методы информационно-коммуникативных технологий, которые используются для быстрого обнаружения онлайн-мошенничества. Что касается преподавателей, то им надо приложить усилия по разработке приемлемых методов оценки знаний студентов, которые могли бы препятствовать академической нечестности [6].

## Литература

1. 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>.

2. Triza Mohareb, Rania Al-Qayyem and Amani Elbarazi (2026). Students' perspectives on AI conversations in brainstorming within learning management systems in higher education, *Research in Learning Technology*, Vol. 34, 2026, Doha, Qatar.
3. The integration of generative ai and learning management systems for modern healthcare education, 20th International Technology, Education and Development Conference, Dates: 2-4 March, 2026, Valencia, Spain.
4. Sefoko Ramoshaba (2025). Behaviour in Time of Information Technology (IT) at Universities: ai may exacerbate the problem/increase the threat, January 2025.
5. Understanding the Importance of Ethics in Information Technology, [https://marymount.edu/blog/understanding-the-importance-of-ethics-in-information-technologys in the field](https://marymount.edu/blog/understanding-the-importance-of-ethics-in-information-technologys-in-the-field).
6. P.S. Park (2024) AI deception: A survey of examples, risks, and potential solutions. <https://doi.org/10.48550/arXiv.2308.14752>