Петровичева Анастасия Евгеньевна
студентка
Иванцова Серафима Александровна
студентка
Научный руководитель
Бобровникова Наталия Сергеевна
старший преподаватель

ФГБОУ ВО «Тульский государственный педагогический университет им. Л.Н. Толстого»
г. Тула, Тульская область

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ И ПСИХОЛОГИЧЕСКАЯ БЕЗОПАСНОСТЬ: ВОЗМОЖНОСТИ, РИСКИ И ГРАНИЦЫ ПРИМЕНЕНИЯ

Аннотация: *в статье анализируются возможности, риски и границы применения искусственного интеллекта (далее – ИИ) в контексте психологической безопасности личности, включая особенности обращения пользователей к чат-ботам за эмоциональной поддержкой. Показаны потенциальные возможности ИИ для психологического просвещения, модерации вредоносного контента и обучения специалистов. Сделан вывод, что ИИ может применяться только как вспомогательный инструмент при условии прозрачного информирования пользователей, защиты данных, независимой оценки безопасности и сохранения ведущей роли квалифицированных специалистов при оказании профессиональной психологической помощи и принятии клинически значимых решений в пределах их компетенции.*

Ключевые слова: *искусственный интеллект, психологическая безопасность, психическое здоровье, цифровая среда, психологическая помощь.*

Искусственный интеллект (далее – ИИ) представляет собой совокупность технологий, способных анализировать данные, распознавать закономерности и

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

формировать рекомендации или ответы. В повседневной жизни такие технологии применяются для поиска и обработки информации, отбора публикаций в социальных сетях, модерации сообщений и создания текстовых ответов. Широкий доступ к цифровой информации также влияет на когнитивные процессы: исследования показывают, что доступность внешних источников информации может изменять особенности запоминания и обработки сведений – так называемый «эффект Google» [2].

В контексте взаимодействия человека с ИИ психологическая безопасность рассматривается через защищённость личности от потенциально деструктивных психологических воздействий, сохранение автономии, устойчивости социальных связей и способности критически оценивать получаемую информацию. Особое значение эти характеристики приобретают при обращении пользователей к чат-ботам за эмоциональной поддержкой [1]. Ориентированный на поддержку и связность ответ нейросети может создавать у пользователя впечатление, что его понимает и консультирует компетентный специалист, хотя система не обладает профессиональной квалификацией и ответственностью, а её ответ не доказывает достоверность сведений. Поэтому для психологической безопасности важно информировать пользователя о возможностях и ограничениях ИИ и определить границы, отделяющие вспомогательную цифровую поддержку от профессиональной психологической помощи. Для этого необходимо рассматривать деятельность психологов и психотерапевтов, чтобы определить границы безопасного использования ИИ.

Риски применения ИИ наиболее заметны в ситуациях, когда пользователь обращается к чат-боту за эмоциональной поддержкой. В четырёхнедельном рандомизированном исследовании Кэти Мэнъин Фанг и соавторов 981 участник ежедневно взаимодействовал с ChatGPT; было проанализировано более 300 000 сообщений. Условия использования – текстовый или голосовой режим, личные либо нейтральные темы – не дали статистически значимых различий между группами. Вместе с тем более длительное самостоятельное общение с чат-ботом

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

было связано с более выраженным одиночеством, меньшим количеством контактов с людьми, эмоциональной зависимостью и проблемным использованием ИИ. Эти результаты показывают связь интенсивного взаимодействия с чат-ботом и неблагоприятных социальных переживаний, но не доказывают, что именно чат-бот является их причиной. С позиции психологической безопасности подобные результаты представляют интерес, поскольку эмоциональная зависимость от цифровой системы и сокращение социальных контактов могут выступать потенциальными факторами психологической уязвимости личности [3].

Сходный вывод позволяет сделать лонгитюдное исследование Дуниган Фолк и Элизабет Данн, в котором участвовали более 2000 взрослых из четырёх западных стран. Более частое использование чат-ботов-компаньонов предсказывало усиление эмоциональной изоляции по одному из показателей, а снижение социальной связанности, напротив, предшествовало увеличению обращений к таким программам. Следовательно, отношения между одиночеством и использованием ИИ могут быть двусторонними: люди с недостатком социальных контактов могут чаще искать поддержку у чат-ботов, а длительное взаимодействие с ними способно быть связано с дальнейшей изоляцией [4].

Одним из аспектов психологической безопасности является риск подмены профессионального психологического взаимодействия общением с ИИ. Зайнаб Ифтихар и соавторы проанализировали 137 сессий взаимодействия с языковыми моделями и выделили 15 типов этических рисков: недостаточный учёт жизненного контекста, обманчивую эмпатию, дискриминационные ответы, подкрепление ошибочных убеждений и недостатки кризисного реагирования. Авторы подчёркивают, что модели могут имитировать понимание общими фразами, но не обладают профессиональной ответственностью и не способны гарантированно оценить состояние человека [6].

Результаты исследования НИУ ВШЭ – Санкт-Петербург также указывают на необходимость осторожной оценки таких технологий. Сопоставление ответов ИИ и специалистов позволяет оценить потенциальные ограничения автоматизи-

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

рованной психологической поддержки. В контексте психологической безопасности это важно, поскольку пользователь может ошибочно воспринимать ответы нейросети как профессиональные рекомендации. Поскольку данные представлены в форме пресс-релиза, их следует рассматривать как предварительные результаты конкретной проверки, а не как характеристику всех ИИ-систем. Однако они показывают, почему ответ нейросети не должен подменять профессиональную помощь, особенно при риске самоповреждения, насилия или выраженного ухудшения психического состояния [5].

При соблюдении ограничений ИИ может выполнять полезные вспомогательные функции. При этом психологическое просвещение не является эквивалентом психологического консультирования, поскольку не предполагает полноценной индивидуальной оценки ситуации пользователя и профессионального сопровождения. В систематическом обзоре Лиин Ван и соавторов отмечаются перспективы генеративного ИИ для распространения такой информации и развития эмоциональной осведомлённости. Одновременно авторы подчёркивают недостаток данных о диагностической точности, культурной чувствительности и качестве длительного взаимодействия с реальными пользователями. Поэтому материалы, выдаваемые нейросетью, должны проверяться специалистами, а пользователь должен быть заранее информирован о пределах возможностей программы [7].

Другой потенциальный ресурс – автоматизированная модерация цифрового контента. Системы ИИ способны выявлять оскорбительные высказывания, угрозы и отдельные виды потенциально вредных публикаций, чтобы направлять их на дополнительную проверку. Такая технология может уменьшить вероятность контакта с некоторыми видами агрессивного контента, но не гарантирует улучшения психологического состояния пользователя. Безопасность модерации требует ясных правил, возможности обжаловать решение и участия человека в рассмотрении неоднозначных случаев.

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

ИИ может быть полезен и в профессиональном обучении психологов: например, при создании тренировочных диалогов, в которых студент отрабатывает первичную беседу, уточняющие вопросы и навыки реагирования. Такой формат позволяет многократно моделировать сложные ситуации без риска для реального клиента. Однако автоматическая обратная связь может быть неточной, поэтому она не заменяет преподавателя, супервизора и анализ реальных профессиональных стандартов.

В профессиональной практике психолог может участвовать в оценке психологических рисков использования ИИ, разработке рекомендаций по безопасному взаимодействию с цифровыми системами и психологическом просвещении пользователей. Таким образом, роль специалиста заключается не только в оказании помощи, но и в предупреждении потенциально деструктивных последствий цифрового взаимодействия.

Обеспечение психологической безопасности при использовании ИИ предполагает прозрачное информирование о взаимодействии с программой, защиту чувствительных данных, проверку систем на разных группах пользователей и заранее определённый порядок действий при признаках кризиса. В рекомендациях Всемирной организации здравоохранения выделяются защита автономии человека, благополучие и безопасность, прозрачность, ответственность, справедливость и устойчивость ИИ [8]. В отдельном руководстве ВОЗ о больших мультимодальных моделях подчёркиваются риски ошибок, предвзятости, нарушения конфиденциальности и манипуляции [9].

Следовательно, ИИ способен расширить доступ к информации и отдельным инструментам самоподдержки, однако его влияние на психологическую безопасность нельзя считать однозначно положительным или отрицательным. Результат зависит от цели использования, качества системы, характера запроса и степени человеческого контроля. Применение ИИ в психологии требует соблюдения границ, обеспечивающих психологическую безопасность личности. ИИ может использоваться для психологического просвещения, информационной поддержки и обучения специалистов, однако не должен рассматриваться как полноценная

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)

замена профессионального психологического консультирования и психотерапии. Особое значение имеет сохранение автономии пользователя, предупреждение эмоциональной зависимости и обеспечение профессионального участия в ситуациях повышенного риска.

Список литературы

1. Баева, И. А. Психологическая безопасность образовательной среды: становление направления и перспективы развития / И. А. Баева // Экстремальная психология и безопасность личности. – 2024. – Т. 1, № 3. – С. 5–19.

2. Sparrow, B. Google effects on memory: cognitive consequences of having information at our fingertips / B. Sparrow, J. Liu, D. M. Wegner // Science. – 2011. – Vol. 333, No. 6043. – P. 776–778.

3. How AI and Human Behaviors Shape Psychosocial Effects of Extended Chatbot Use: A Longitudinal Randomized Controlled Study / C. M. Fang, A. R. Liu, V. Danry [et al.] // arXiv. – 2025. – arXiv:2503.17473.

4. Folk, D. How does turning to AI for companionship predict loneliness and vice versa? / D. Folk, E. Dunn // Psychological Science. – 2026. – Vol. 37, No. 4. – P. 276–286.

5. Ученые Питерской Вышки раскрыли риски использования ИИ в роли психологов // НИУ ВШЭ – Санкт-Петербург. – 5 февраля 2026 г. – URL: https://spb.hse.ru/news/1125629146.html (дата обращения: 12.10.2026). – Текст : электронный.

6. How LLM Counselors Violate Ethical Standards in Mental Health Practice: A Practitioner-Informed Framework / Z. Iftikhar, A. Xiao, S. Ransom [et al.] // Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society. – 2025. – Vol. 8, No. 2. – P. 1311–1323.

7. Evaluating Generative AI in Mental Health: Systematic Review of Capabilities and Limitations / L. Wang, T. Bhanushali, Z. Huang [et al.] // JMIR Mental Health. – 2025. – Vol. 12. – e70014.

8. World Health Organization. Ethics and governance of artificial intelligence for health. – Geneva : WHO, 2021.

Содержимое доступно по лицензии Creative Commons Attribution 4.0 license (CC-BY 4.0)

9. World Health Organization. Ethics and governance of artificial intelligence for health: guidance on large multi-modal models. – Geneva : WHO, 2024.

Content is licensed under the Creative Commons Attribution 4.0 license (CC-BY 4.0)