

Доверие к искусственному интеллекту как форма взаимодействия личности с цифровой средой

Научный руководитель – Воробьева Ксения Андреевна

Яхудина Татьяна Семеновна

Аспирант

Российский университет дружбы народов имени Патриса Лумумбы, Москва, Россия

E-mail: tany771@yandex.ru

Яхудина Татьяна Семеновна

Аспирант

Российский университет дружбы народов имени Патриса Лумумбы, Москва, Россия

E-mail: <mailto:tany771@yandex.ru>

Воробьева Ксения Андреевна

Доктор психологических наук, доцент, доцент

Российский университет дружбы народов имени Патриса Лумумбы, Москва, Россия

E-mail: kseniaprimavera@mail.ru

В условиях интенсивной цифровизации общества технологии искусственного интеллекта (ИИ) перестают быть исключительно инструментальными средствами обработки данных и становятся полноправными участниками повседневного взаимодействия личности с цифровой средой. Современный субъект все чаще вступает в коммуникативные, когнитивные и регулятивные отношения с алгоритмическими системами, что требует психологического анализа феномена доверия к ИИ как специфической формы отношения человека к искусственному агенту.

Проблема доверия в классической психологии связана с формированием высших психических функций, опосредованных культурными средствами и социальным взаимодействием [5]. В логике культурно-исторического подхода цифровые технологии могут рассматриваться как новые знаково-опосредующие системы, включенные в структуру деятельности. Следовательно, доверие к ИИ выступает не только установкой, но и элементом организации деятельности в цифровой среде, когда субъект опирается на алгоритм как на посредника в понимании, выборе и саморегуляции [7; 15].

Однако этого уровня анализа оказывается недостаточно, когда ИИ перестает восприниматься как нейтральный инструмент и начинает занимать место «почти-участника» взаимодействия. В таких ситуациях меняется не только способ выполнения задачи, но и то, как человек переживает собственную роль, возникает вопрос об авторстве решения, где граница контроля и на ком ощущается ответственность за итог [14; 15]. Именно в этом месте становится логичным переход к философско-антропологическому подходу, потому что он позволяет описывать доверие к ИИ не только как психологическую установку, но и как форму изменения человеческого опыта в условиях цифровой среды [9].

Философско-антропологический анализ противопоставления естественного и искусственного показывает, что внедрение технических систем трансформирует границы человеческого опыта, меняются критерии «понятности» мира, способы ориентировки и самоопределения, а также основание того, что человек считает «своим» действием [9]. Если в культурно-исторической рамке мы фиксируем опосредование деятельности [5], то философско-антропологическая рамка помогает увидеть то, как опосредование начинает затрагивать саму субъектность. И человек вынужден соотносить себя с системой, которая выдает рекомендации и решения, но остается частично непрозрачной [7; 12]. Это особенно важно для понимания доверия, потому что доверие здесь связано не только с оценкой

качества результата, но и с переживанием уязвимости: «я полагаюсь, но не до конца понимаю, как это устроено» [12; 15].

Отсюда возникает парадокс доверия к ИИ. С одной стороны, цифровая среда подталкивает к делегированию, экономит усилия, и алгоритм воспринимается как более последовательный источник суждения [18]. С другой стороны, как только пользователь сталкивается с ошибкой или допускает возможность скрытой предвзятости, доверие может резко снижаться, формируя избегание алгоритмических рекомендаций даже при общей эффективности системы [13]. В повседневных и профессиональных контекстах это проявляется как амбивалентность, когда человек одновременно и стремится опереться на ИИ и удерживать контроль (чаще проверяет итог) [7; 15]. В этой связи философско-антропологический подход методологически оправдан, он позволяет связать доверие к искусственному интеллекту с тем, как личность выстраивает отношения с цифровой средой, регулирует неопределенность и сохраняет субъектную позицию в условиях частично автономных технологий [9; 12; 15].

Зарубежные и отечественные эмпирические исследования показывают, что доверие к искусственному интеллекту не является постоянным. Оно формируется в конкретной ситуации, когда человек решает, можно ли опереться на рекомендацию системы или нужно перепроверить результат. На доверие влияет не только общее отношение к технологии, но и условия взаимодействия, например, насколько понятен принцип работы, насколько прозрачен вывод, как распределяются контроль и ответственность, и какой риск несет ошибка в данной задаче [7; 15].

В зарубежных исследованиях взаимодействия человека и автоматизированных систем доверие чаще описывается как вопрос соразмерной опоры на технологию. Если опора становится чрезмерной, то у человека снижается критичность к результату, и система начинает применяться там, где это ситуационно или методологически неоправданно, тем самым возрастает риск ошибочного и некритичного использования. Если же опора оказывается недостаточной, человек, наоборот, избегает алгоритмических рекомендаций даже в тех случаях, когда они могли бы реально повысить качество решения и возникает отказ от полезной поддержки [15; 17]. При этом исследования отмечают, что даже одна заметная ошибка алгоритма может резко снизить доверие к системе. Человек фиксирует сбой и начинает чаще избегать алгоритмических рекомендаций, даже если в целом система работает хорошо [13]. Из-за этого взаимодействие становится более осторожным и пользователь чаще перепроверяет выводы, выбирает только «безопасные» задачи для делегирования и старается удерживать контроль за итоговым решением [7; 15]. Но сохраняется и обратная тенденция, когда в ряде задач люди готовы предпочесть алгоритмическое суждение человеческому, особенно когда алгоритм воспринимается как более последовательный и менее подверженный субъективным колебаниям [18].

Отечественные данные в целом поддерживают эту картину. Так, в исследованиях экономических решений показано, что доверие алгоритму и доверие человеку различаются в зависимости от того, как субъект переживает ответственность и риск. В одних ситуациях алгоритм воспринимается как удобная опора, в других, как источник потенциальной ошибки, за последствия которой все равно внутренне отвечает человек [3]. Поэтому доверие к ИИ в реальной практике выглядит как процесс постепенной настройки, когда субъект выбирает, где допустимо делегирование, а где важно сохранять контроль и перепроверку результата. Существенную роль здесь играют прозрачность процедур, обратная связь и дизайн взаимодействия, которые позволяют «калибровать» доверие и удерживать его в пределах адекватной опоры, избегая как слепого принятия, так и полного отказа от технологии [7; 14; 15].

В повседневной жизни происходит процесс «одомашнивания» ИИ, при котором циф-

ровые устройства интегрируются в структуру семейных и бытовых практик [1; 10]. Трансформация семейных ролей и распределения функций в условиях цифровизации быта свидетельствует о постепенной нормализации присутствия алгоритмических систем в интимном пространстве [6; 11]. Искусственный интеллект становится элементом повседневной инфраструктуры, а доверие к нему приобретает характер привычного, операционального доверия, основанного на опыте эффективности и предсказуемости.

В образовательной сфере ИИ выступает как вспомогательный когнитивный ресурс, влияющий на процессы обучения и самоорганизации [4]. Включение интеллектуальных систем в учебную деятельность актуализирует вопрос о соотношении цифровой компетентности и доверия. Чем выше уровень цифровой компетентности субъекта, тем более осмысленным и критичным становится его взаимодействие с алгоритмическими агентами. Таким образом, доверие к ИИ выступает формой регуляции неопределенности в цифровой среде.

Современные исследования показывают, что искусственный интеллект постепенно включается в рутинные сценарии принятия решений и становится партнером по распределенной когнитивной деятельности [2]. Взаимодействие человека и алгоритма приобретает характер кооперации, когда человек формулирует задачу, оценивает результат и принимает окончательное решение, тогда как система обеспечивает расширение когнитивных возможностей. Доверие в этой ситуации выступает механизмом включения алгоритма в структуру деятельности, а не безусловным принятием его выводов.

Чтобы описать доверие к ИИ как динамический процесс, важно показать, из каких условий оно возникает, какие компоненты включает и как затем влияет на выбор стратегии взаимодействия с алгоритмом. На схеме ниже представлена логика формирования и проявления доверия к ИИ в цифровой среде, от контекста цифровой социализации и уровня цифровой компетентности к компонентам доверия, ожиданиям, стратегиям поведения и последующим изменениям в субъектной позиции человека.

На схеме показано, что доверие к ИИ формируется в контексте цифровой социализации и регулярных практик использования технологий. Качество доверия задается уровнем цифровой компетентности. При более высокой компетентности усиливается критичность, возрастает частота проверки результата и лучше осознаются ограничения системы. В центральном блоке отражены компоненты доверия, такие как когнитивный, аффективный, поведенческий и мотивационный, которые вместе определяют ожидания пользователя и выбор стратегии (делегирование, проверка, контроль). Опыт взаимодействия и его последствия приводят к коррекции ожиданий и решению о повторном применении ИИ. В итоге доверие к ИИ влияет на перестройку границ субъектности и ответственности и поддерживает конструктивное взаимодействие личности с цифровой средой.

Важно учитывать и мотивационный компонент доверия. Он связан с уверенностью пользователя в результате и с тем, чего он заранее ждет от системы возможной предвзятости и ошибок [3; 7; 13]. Эти ожидания влияют на то, будет ли человек пользоваться ИИ снова, в каких задачах он готов делегировать решение и насколько тщательно будет проверять итог [7; 15; 18]. Доверие к ИИ формируется в пространстве цифровой социализации и встраивается в повседневные и образовательные практики, где ИИ становится привычной частью цифровой среды [1; 4; 10]. При этом уровень цифровой компетентности задает качество этого доверия, оно проявляется в том, чем выше компетентность, тем больше критичности, тем понятнее ограничения, и тем чаще возникает проверка результата вместо слепого принятия [5; 7; 15]. Анализ доверия к ИИ позволяет увидеть, как меняются границы субъектности и ответственности в цифровую эпоху и какие условия поддерживают конструктивное взаимодействие человека с интеллектуальными технологиями [9; 14;

15]. В этом смысле доверие к искусственному интеллекту можно рассматривать как форму взаимодействия личности с цифровой средой, через которую человек регулирует неопределенность, распределяет контроль и по-новому выстраивает свою роль в технологически насыщенной реальности [9; 10; 15].

Источники и литература

- 1) 1. Азина Д. Б., Пьянкова П. А., Козицына М. О., Потапов Д. А., Иванова В. В., Кулагина Е. В., Кузнецова А. С. Речевое взаимодействие между человеком и искусственным интеллектом посредством чат-ботов и умных устройств // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Сер. 6. Языкознание: Реферативный журнал. 2022. № 1. С. 54–65. 2. Аксентов В. А. Искусственный интеллект в повседневной жизни: развитие и перспективы // Вестник науки. 2023. № 8 (65). С. 80–82. 3. Винокуров Ф. Н., Садовская Е. Д. Экспериментальное сравнение доверия искусственному интеллекту и человеку в экономических решениях // Экспериментальная психология. 2023. Т. 16. № 2. С. 87–100. 4. Волков Н. И., Лебедева А. К. Роль искусственного интеллекта в образовательном процессе детей в домашних условиях // Вопросы педагогики. 2023. № 9-1. С. 45–59. 5. Выготский Л. С. История развития высших психических функций // Собр. соч.: В 6 т. Т. 3. М., 1983. 6. Григорьев С. В. Цифровизация быта и трансформация семейных ролей // Человек и общество. 2020. № 4 (55). С. 78–90. 7. Дашков А. А., Нестерова Ю. О. Построение доверия при использовании искусственного интеллекта // E-Management. 2021. № 2. 8. Еникеев И. М., Солнцев В. И. Доверие к искусственному интеллекту в здравоохранении и пути его укрепления // Вестник Московского университета. Серия 21. Управление (государство и общество). 2025. № 4. 9. Кутырев В. А. Естественное и искусственное: борьба миров. Н. Новгород: Изд-во «Нижний Новгород», 1994. 10. Корбут А. М. Одомашнивание искусственного интеллекта: умные колонки и трансформация повседневной жизни // Мониторинг общественного мнения: экономические и социальные перемены. 2021. № 1. С. 193–216. 11. Кривоустова А. Г. Использование искусственного интеллекта в повседневных практиках семей // Вестник науки. 2025. № 12 (93). 12. Лешкевич Т. Г. Парадокс доверия к искусственному интеллекту и его обоснование. 2023. 13. Dietvorst BJ, Simmons JP, Massey C. Algorithm aversion: people erroneously avoid algorithms after seeing them err. *J Exp Psychol Gen.* 2015 Feb;144(1):114-26. doi: 10.1037/xge0000033. Epub 2014 Nov 17. PMID: 25401381. 14. Hoff, K. A., & Bashir, M. (2015). Trust in Automation: Integrating Empirical Evidence on Factors That Influence Trust: Integrating Empirical Evidence on Factors That Influence Trust. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 57(3), 407-434. 15. Lee, J. D., & See, K. A. (2004). Trust in Automation: Designing for Appropriate Reliance. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 46(1), 50-80. 16. Logg J. M., Minson J. A., Moore D. A. Algorithm appreciation: People prefer algorithmic to human judgment // *Organizational Behavior and Human Decision Processes*. 2019. Vol. 151. P. 90-103. 17. Mayer, Roger C., et al. "An Integrative Model of Organizational Trust." *The Academy of Management Review*, vol. 20, no. 3, 1995, pp. 709–34. JSTOR, <https://doi.org/10.2307/258792>. Accessed 5 Mar. 2026. 18. Parasuraman, R., & Riley, V. (1997). Humans and Automation: Use, Misuse, Disuse, Abuse. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 39(2), 230-253.

Иллюстрации

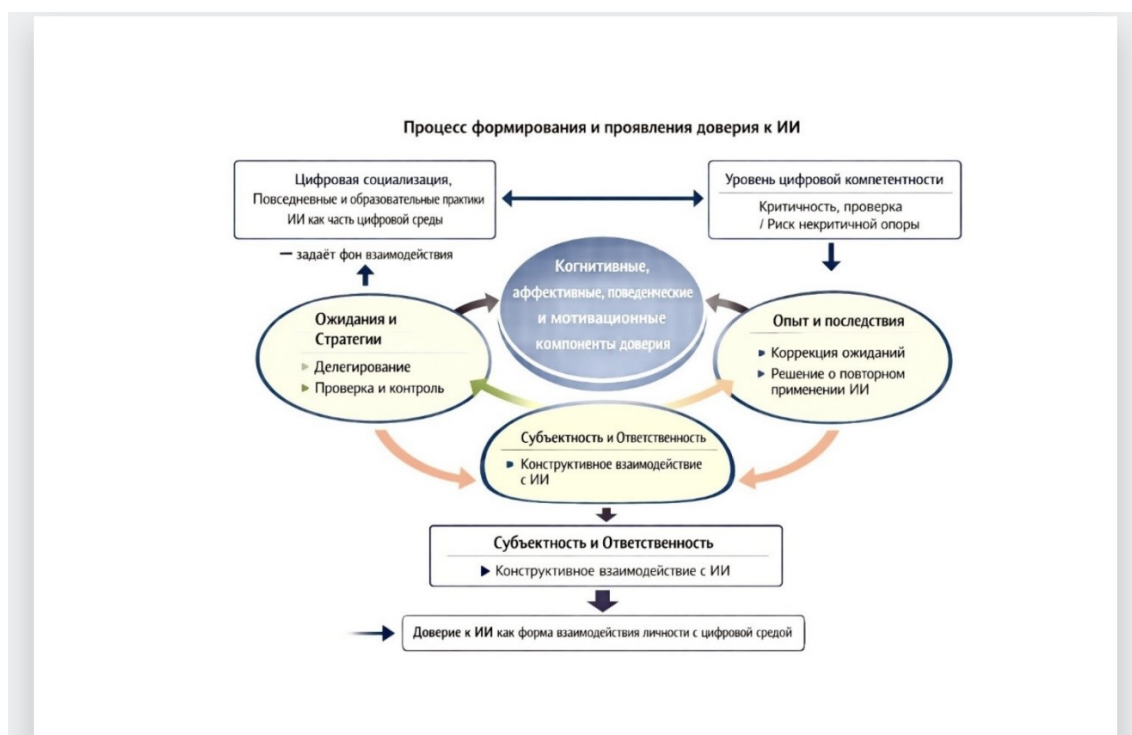


Рис. : На схеме показано, что доверие к ИИ формируется в контексте цифровой социализации и регулярных практик использования технологий. Качество доверия задается уровнем цифровой компетентности. При более высокой компетентности усиливается критичность, возрастает частота проверки результата и лучше осознаются ограничения системы. В центральном блоке отражены компоненты доверия, такие как когнитивный, аффективный, поведенческий и мотивационный, которые вместе определяют ожидания пользователя и выбор стратегии (делегирование, проверка, контроль). Опыт взаимодействия и его последствия приводят к коррекции ожиданий и решению о повторном применении ИИ. В итоге доверие к ИИ влияет на перестройку границ субъектности и ответственности и поддерживает конструктивное взаимодействие личности с цифровой средой.