ОБЗОРЫ И РЕЦЕНЗИИ

О.М. Самойлов
аспирант, Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ), (Москва)

А.Н. Татарко
д.психол.н., профессор, Национальный исследовательский университет «Высшая школа экономики» (НИУ ВШЭ), (Москва)

СОЦИАЛЬНО-ПСИХОЛОГИЧЕСКИЕ ФАКТОРЫ ДОВЕРИЯ ИСКУССТВЕННОМУ ИНТЕЛЛЕКТУ: СОСТОЯНИЕ ИССЛЕДОВАНИЙ[1]

Аннотация. Аннотация. В статье представлен теоретический обзор литературы за последние десять лет, посвящённой анализу социально-психологических факторов доверия искусственному интеллекту. Повсеместное внедрение автоматизированных ИИ-систем, связанное с ожидаемым экономическим ростом, снижением ресурсных затрат и оптимизацией ряда рабочих процессов, на практике зачастую сталкивается с недоверием пользователей к новым инструментам и отсутствием готовности трансформировать классические рабочие процессы. Совокупность факторов снижения доверия к искусственному интеллекту приводит к низкой экономической эффективности внедрения инноваций, несмотря на широкие технические возможности. Помимо важности учёта когнитивных и аффективных факторов доверия искусственному интеллекту, особую значимость приобретают социально-психологические аспекты, которые определяют этическую и ценностную приемлемость использования автоматизированных ИИ-систем. В рамках теоретического анализа было обнаружено, что индивидуализирующие моральные основания положительно связаны с доверием искусственному интеллекту в ситуациях, когда применение ИИ-систем приносит пользу обществу. Пользователи, отдающие приоритет сплачивающим моральным основаниям, проявляют большее недоверие к искусственному интеллекту и меньшую готовность делегировать часть задач автоматизированным ассистентам. Ценности «Открытости изменениям» и «Самопреодоления» в большей степени положительно связаны с доверием к ИИ и цифровым инновациям, исключая ситуации высокого риска для жизни или социальной несправедливости. Ценности «Самоутверждения» также положительно связаны с доверием ИИ-инструментам, но преимущественно в ситуациях, когда искусственный интеллект упрощает достижение целей пользователя или расширяет возможности человека для этого. Между ценностями «Сохранения» и доверием искусственному интеллекту наблюдается неоднородная структура связей, обусловленная культурными особенностями. Хотя необходимо отметить, что именно ценности «Сохранения» чаще всего рассматриваются в качестве предикторов недоверия автоматизированным ИИ-системам. Обсуждается важность рассмотрения ценностной конгруэнтности между пользователями и воспринимаемыми профилями нейросетей. Для разработчиков ИИ-систем обнаруживается необходимость уделять особое внимание возможностям адаптивной персонализированной настройки ценностных профилей генеративных моделей под пользователей, что приведёт к повышению эффективности взаимодействия системы человек-машина. Подчёркиваются перспективы исследований данной области в рамках разработки системной модели доверия искусственному интеллекту.

[1] Исследование осуществлено в рамках Программы фундаментальных исследований НИУ ВШЭ (HSE-BR-2025-52).
This article is an output of a research project HSE-BR-2025-52 implemented as part of the Basic Research Program at HSE University.

Ключевые слова: *доверие, искусственный интеллект, социально-психологические факторы, ценности, моральные основания.*
JEL: O33, D83, M15, A13, Z13
УДК: 159.9, 316.6
DOI: 10.52342/2587-7666VTE_2026_2_209_228

ДЛЯ ЦИТИРОВАНИЯ: *Самойлов О.М., Татарко А.Н.* Социально-психологические факторы доверия искусственному интеллекту: состояние исследований // Вопросы теоретической экономики. 2026. №2. С. 209–228. DOI: 10.52342/2587-7666VTE_2026_2_209_228.
FOR CITATION: *Samoilov O., Tatarko A.* Socio-Psychological Factors of Trust in Artificial Intelligence: State of Research // Voprosy teoreticheskoy ekonomiki. 2026. No. 2. Pp. 209–228. DOI: 10.52342/2587-7666VTE_2026_2_209_228.

Введение

С каждым годом инструменты искусственного интеллекта (ИИ) всё стремительнее находят применение в реальных практических задачах, значительно трансформируя экономические системы и социальное взаимодействие. Новые ИИ-инструменты становятся неотъемлемой частью повседневной жизни и профессиональной деятельности — от умных устройств и медиа-платформ до глобальных логистических систем [*Stein et al.,* 2024]. Нейросетевые технологии позволяют изменять функционирование экономики на макроуровне, посредством автоматизации как рутинных процессов, так и сложных когнитивных задач [*Aghion, Jones, Jones*, 2017]. Несмотря на это, общество с осторожностью относится к активному внедрению ИИ. К наибольшим рискам при использовании нейросетей относят этические проблемы, связанные с конфиденциальностью данных, галлюцинациями, моральным статусом ИИ, ложным приписыванием авторства [*Stein et al.,* 2024; *Bai, Yang,* 2025]. Экономическая эффективность внедрения инструментов ИИ в классические рабочие механизмы во многом зависит не только от технологической готовности и производительности, но и от моральной релевантности использования ИИ-инструментов в привычных социальных структурах [*Panari Lorenzi, Mariani,* 2021][1]. Несоответствие между существующим технологическим потенциалом и его восприятием подчёркивает значительную важность исследования социально-психологических факторов, влияющих на доверие к современным технологиям у представителей разных демографических групп.

Существующие прогностические исследования дают основания полагать, что применение ИИ может значительно ускорить производительность труда и снизить затраты [*Макаров,* 2020]. Особенно это касается сфер, которые характеризуются активной работой с данными, в том числе финансовых услуг, здравоохранения, профессиональных консультаций, аналитики и других информационно зависимых сфер. Однако недавно реализованные экономические исследования демонстрируют, что 95% проектов по внедрению ИИ крупными компаниями не привели к увеличению прибыли или снижению затрат, несмотря на полную технологическую оснащённость [*Massachusetts Institute of Technology,* 2025]. У таких результатов две основные причины. С одной стороны, существует пробел в обучении генеративных систем конкретным рабочим задачам и недостаток опыта у сотрудников по «обучению» ИИ-инструментов [*Filippucci et al.,* 2024]. С другой стороны, обнаруживается проблема доверия к нейросетям не только у руководителей организаций, но и у работников, и обычных пользователей цифровых инструментов [*Vuori, Burkhard, Pitkäranta,* 2025; *Zhang et al.,* 2025].

[1] В настоящей работе акцент сделан в целом на анализе социально-психологических факторов доверия к ИИ. Свои особенности имеет влияние доверия к ИИ на экономические процессы, но это тема специального исследования, которое во многом должно опираться на результаты, полученные при рассмотрении общих социально-психологических проблем, связанных с доверием людей к ИИ.

В контексте ИИ доверие рассматривается с точки зрения двух различных аналитических позиций: как доверие непосредственно к инструментам ИИ, так и к социальным институтам, которые разрабатывают и внедряют различные инновационные решения во взаимодействия с ИИ [*Петрунин, Нуралиева,* 2025]. Именно доверие является одним из ключевых механизмов снижения воспринимаемых рисков и угроз от ИИ, а также облегчает делегирование контроля чат-ботам и другим инструментам [*Wen, Wang, Chen,* 2025]. Однако восприятие рисков от ИИ имеет две противоположные стороны: чрезмерное доверие и недоверие. С одной стороны, чрезмерное доверие снижает осторожность людей и может приводить к совершенно разным негативным последствиям [*Robinette et al.,* 2016]. С другой стороны, патологическое недоверие ИИ, к примеру, вызванное положительным отношением к конспирологическим теориям, искажает восприятие возможных рисков технологического поведения, что отрицательно отражается на готовности пользоваться различными ИИ-инструментами [*Казун, Поршнев,* 2021; *Stein et al.,* 2024]. С экономической точки зрения доверие к автоматизированным системам обусловливает поведенческие проявления готовности использовать технологию с последующей экономической выгодой, в то время как недоверие алгоритмам, напротив, связано с недостаточным использованием всего технического потенциала и низкой эффективностью [*Dietvorst, Simmons, Massey,* 2015].

Рассмотрение доверия к искусственному интеллекту в ситуациях экономических решений обнаруживает ряд важных результатов. Советы при принятии экономических решений от ИИ вызывают меньше доверия, чем рекомендации от человека [*Винокуров, Садовская,* 2023]. Хотя в принятии быстрых решений люди в большей степени склонны доверять ИИ, ориентируясь в первую очередь на критерий оперативности. Важно отметить, что согласно исследованиям, если человек и ИИ работают совместно, то такое партнёрство повышает их индивидуальную точность и улучшает совместный результат [*Ulfert, Antoni, Ellwart,* 2022]. В работе экономических систем совместная деятельность человека и ИИ позволяет снижать временные затраты на анализ данных, но при этом позволяет сохранять точность посредством адаптации человеком материала к контексту решения задачи, а также контроль соблюдения этики со стороны эксперта [*Лукичев,* 2024; *Манахова, Маковская,* 2025]. Но если партнёрство человек-ИИ подразумевает постоянную перепроверку друг друга, то это приводит к снижению общей результативности [*Zhang, Lee, Carter,* 2022].

Наиболее известные концепции доверия демонстрируют, что основания доверия технологиям и ИИ шире, чем ориентация только на их техническую точность и производительность (см., например, обзор эмпирических работ [*Glikson, Woolley,* 2020]). Зачастую техническая оценка дополняется социально-психологическими факторами, включающими морально-нормативную оценку ИИ и соответствие этой системы ценностным профилям пользователя. Это соотносится с классической моделью Р.С. Майера и его коллег, где доверие рассматривается не только через компонент способностей, но и через факторы благожелательности и честности [*Mayer, Davis, Schoorman,* 1995]. Следовательно, системное рассмотрение доверия к ИИ невозможно свести только к изучению его когнитивных аспектов, включающих восприятие полезности, точности, надёжности, ожидание производительности и другие факторы, широко представленные в ряде исследований [*Xiong et al.,* 2023; *Cheung, Ho*, 2025; *Kai, Ping, Xiaomin,* 2026]. Если пользователи воспринимают автоматизированные системы как несправедливые или морально неправильные, то они не будут им доверять даже при широких возможностях их применения и их высокой технической точности [*Bigman, Gray,* 2018].

В структуре социально-психологических факторов доверия искусственному интеллекту выстраивается логика двух ключевых направлений: анализ ценностей и анализ моральных оснований. Исследования показывают устойчивую взаимосвязь между ценностями и доверием к цифровым инструментам, онлайн-сервисам, ИИ-технологиям [*Morselli, Spini, Devos,* 2012; *Glikson, Woolley,* 2020; *You et al.*, 2022; *Stanciu, Partsch, Lechner*, 2024], в то время как экспериментальные данные демонстрируют высокую значимость конгруэнтности

ценностей пользователя и воспринимаемых ценностей ИИ-системы в повышении доверия к таким инструментам [*Mehrotra, Jonker, Tielman,* 2021]. Необходимо отметить, что ИИ не может обладать собственной субъективной системой ценностей, поэтому в ряде статей исследователи самостоятельно задают моделям ценностные приоритеты (посредством коэффициентов для каждой ценности). В свою очередь, конгруэнтность выражается в восприятии пользователями поведения ИИ-системы, сгенерированной на основании заданных ценностных приоритетов, и соотношении этого поведения с собственным ценностным профилем. Анализ моральных оснований определяет этическую допустимость использования нейросетей, а также доверие инструментам ИИ в социально чувствительных сферах, таких как здравоохранение, политика, управление, правосудие, образование [*Awad et al,* 2018; *Jobin, Ienca, Vayena,* 2019; *Gerke, Minssen, Cohen,* 2020]. Постоянное возрастание включённости ИИ в принятие социально значимых решений [*Glikson, Woolley,* 2020] повышает значимость ценностных и моральных оснований доверия пользователей к ИИ, которые модифицируют когнитивные оценки технических преимуществ автоматизированных систем.

В ряде исследований акцентируется внимание на возрастных различиях пользователей по доверию к нейросетям. Согласно некоторым эмпирическим данным, молодые люди, а также пользователи с более высоким уровнем образования демонстрируют более высокое доверие к ИИ-инструментам и готовность к их использованию. В то время как пожилые люди склонны к большему недоверию и скептицизму в отношении приватности и прозрачности алгоритмов работы ИИ [*Horowitz, Kahn,* 2021].

При этом в ряде работ, выполненных в другой исследовательской логике, нами были обнаружены позитивные установки пожилых людей по отношению к ИИ, к технологичным роботам [*Gursoy et al.,* 2025], а также к цифровым технологиям в сфере здравоохранения [*Anisha et al.,* 2025]. Необходимо подчеркнуть, что отмеченное нами, равно как и многими другими исследователями, дифференцированное отношение пожилых людей к современным технологиям обусловливается уровнем развития цифровой грамотности респондентов, принадлежащих к этой возрастной группе, и контекстом проводимых исследований [*Li, Wei,* 2025]. Поэтому пожилые люди демонстрируют высокие показатели доверия по отношению к нейросетевым инструментам, если, с их точки зрения, условия применения ИИ оптимальны: например, соблюдается условия простоты использования технологий, поддержки пользователей и безопасности.

Таким образом, доверие инструментам ИИ связано со скоростью внедрения технологий и готовностью делегировать ИИ часть решаемых ныне человеком задач. Это напрямую обусловливает экономическую целесообразность изучения данного феномена. Кроме того, уровень доверия связан с повышением эффективности выполнения поставленных ИИ задач и оптимизацией связанных с его применением затрат, а также точностью оценок рисков внедрения ИИ. Несмотря на значительный рост публикаций, связанных с развитием нейросетей, в том числе систематических обзоров, ныне большинство исследований сосредоточены на изучении когнитивных факторов доверия к ИИ. Основной акцент в таких работах смещён на восприятие и оценку технических характеристик ИИ-инструментов (их точности, простоте, компетентности, профессионализме, доступности, прозрачности, понятности). Однако социально-психологические характеристики, такие как ценности и моральные основания, анализируются фрагментарно. Рассмотренные теоретические предпосылки указывают на то, что моральная приемлемость использования нейросетей и ценностная конгруэнтность могут модерировать эффекты когнитивной оценки на доверие к ИИ-инструментам. Этот факт подчёркивает значимость анализа социально-психологических факторов (ценностей и моральных оснований) в структуре доверия к искусственному интеллекту и их прогностическую силу, в том числе в экономическом контексте. Поэтому целью данного теоретического исследования стало проанализировать роль ценностей и моральных оснований в структуре доверия искусственному интеллекту.

Материалы и методы исследования

Основным исследовательским методом был выбран сплошной теоретический обзор литературы, наточенный на целенаправленный поиск, отбор и анализ эмпирических работ, посвящённых изучению социально-психологических предикторов доверия ИИ. Данный обзор не включал в себя мета-анализ вследствие гетерогенности методологии исследований доверия к ИИ, а также невозможности сопоставлять количественные результаты опросных методов с экспериментальными данными и качественными интервью.

Поиск публикаций осуществлялся в ноябре–феврале 2025–2026 гг. в электронных библиографических базах Google Scholar, PubMed, а также РИНЦ. Поисковая стратегия заключалась в использовании комбинаций следующих ключевых слов: «социально-психологические предикторы OR socio-psychological predictors», «ценности OR personal values», «моральные основания OR moral foundations», «доверие к искусственному интеллекту OR AI trust», «искусственный интеллект OR ИИ OR artificial intelligence OR AI»[2]. В итоговый аналитический корпус обзора вошли рецензируемые статьи, опубликованные в период 2018–2026 гг. В рамках критериев включения публикаций были выбраны следующие параметры: эмпирический характер работы, соответствие ключевым словам, английский или русский язык публикации, а также доступ к полному тексту или расширенной аннотации. Исключение составили источники, описывающие Теорию базовых человеческих ценностей Ш. Шварца и Теорию моральных оснований. После изучения аннотаций, удаления дубликатов и полнотекстового анализа работ, согласно обозначенным критериям, в итоговый обзор вошло 32 публикации.

Результаты

Ценности и доверие к искусственному интеллекту

Теория базовых человеческих ценностей была разработана Ш. Шварцем и впервые описана в 1992 г., представляя на данном этапе одну из наиболее популярных социально-психологических концепций ценностей. В рамках данной теории ценности рассматриваются как надситуативные цели, которые варьируются по важности и определяют поведение индивида или группы [*Schwartz,* 1992]. При изучении ценностей автор фокусировался на двух уровнях анализа: уровень культурных ценностей и уровень индивидуальных ценностей. Первоначально было идентифицировано 10 базовых ценностей, общепризнанных во всех изучаемых культурах — Универсализм, Благожелательность, Власть, Гедонизм, Достижения, Самостоятельность, Стимуляция, Безопасность, Конформность и Традиция. Позднее теория была дополнена до 19 ценностей, что обеспечило повышение прогностических возможностей, и получила широкое распространение в эмпирических исследованиях [*Schwartz et al.,* 2012]. Все ценности организованы в циркулярную структуру, в которой противоположные полюса конфликтуют, а смежные ценности мотивационно совместимы. Ценности с высокими положительными связями между собой образуют ценности более высокого порядка, которые получили название «мета-ценности». На первом векторе мета-ценности «Открытости изменениям» (Самостоятельность, Стимуляция) противопоставлены ценностям «Сохранения» (Традиции, Конформность, Безопасность). Такая структура отражает мотивационный конфликт между стремлением к автономии, познанию нового, независимости и, напротив, ориентации на

[2] Так как материалом для анализа был значительный объём интернет-источников, то, естественно, при отборе данных в качестве вспомогательных использовались инструменты ИИ для ускорения процессов первичного поиска. Однако после проведения необходимых процедур верификации полученной информации инструменты ИИ не использовались для конечного обобщения научных результатов и создания текста статьи.

стабильность, следованию нормам, сохранении традиций. Вторую оппозицию представляют мета-ценности «Самоутверждения» (Власть, Достижение) и «Самопреодоления» (Благожелательность, Универсализм). Это измерение размещает ценности достижения личного успеха, доминирования, в противовес альтруистическим ориентирам, заботе о других людях, идеям общего равенства [*Татарко,* 2017]. При изучении поведения или влияния какого-либо внешнего фактора необходимо учитывать связь этих явлений не только с ценностями высшего порядка, но и с конкретными ценностями внутри каждого сектора. Теория базовых ценностей Ш. Шварца была многократно эмпирически валидирована в разных культурных контекстах и является надёжным конструктом с высокой прогностической силой в объяснении социально-психологических феноменов [*Schwartz et al.,* 2012; *Witte, Stanciu, Boehnke*, 2020].

Кросс-культурное изучение связей личных ценностей исследователей из Китая и Германии и ИИ [*Lammert et al.,* 2026] демонстрирует следующие взаимосвязи. Европейские респонденты проявляли большую осторожность по отношению к ИИ, а также отдавали приоритет ценностям Самопреодоления. Китайские исследователи продемонстрировали большее доверие к ИИ и ставили в приоритет ценности Самоутверждения и Сохранения. Стоит отметить, что наиболее значимыми оказались различия между немецкими и китайскими респондентами по вопросу «Я использую искусственный интеллект без каких-либо опасений» (у группы из Китая M=3,2, а у группы из Германии M=2,3). Помимо учёта культуральных особенностей, которые играют важнейшую роль в гетерогенности результатов данного исследования, можно предположить следующую интерпретацию полученных данных. Ценности Самоутверждения повышают доверие к нейросетевым инструментам как к средству достижения успеха, в то время как ценности Самопреодоления определяют больший скептицизм и недоверие по отношению к ИИ, так как ориентированы на анализ потенциальных социальных угроз и рисков.

В другом кросс-культурном исследовании подростков из 10 стран Юго-Восточной Европы (N=10902) изучались связи между ценностями, институциональным доверием и использованием цифровых услуг [*Lep, Trunk, Babnik*, 2022]. Хотя в представленном исследовании не рассматривается влияние ценностей на доверие к ИИ, однако обнаруживается опосредующая роль институционального доверия между ценностями у подростков (в частности, ценности Самопреодоления; ценности Самоутверждения) и использованием цифровых услуг. На основании эмпирических данных выстраивается следующая логика: ценности не имеют прямого влияния на использование цифровых услуг, однако они являются фактором, определяющим институциональное доверие. При этом доверие предсказывает готовность использовать технологии. Авторы отмечают нестабильность результатов и региональную вариативность регрессионных моделей, в частности, в тех странах, где интернет использовался реже, прогностическая сила доверия была менее значимой, чем в странах с более развитым использованием цифровых технологий.

Одна из наиболее перспективных областей повышения доверия к ИИ — адаптивная корректировка заданных ценностей ИИ ценностным профилям пользователей. В эмпирическом исследовании было показано, что доверительная расслабленность пользователей при взаимодействии с ИИ в большей степени связана с их личными ценностями [*Tang, Ferronato, Bashir,* 2023]. Стоит отметить, что наибольшую предсказательную силу в этом случае имели ценности Открытости изменениям, а также то, что ценности данного блока поддерживают сохранение доверия между человеком и автоматизированной системой в процессе дальнейшего взаимодействия. Интересно, что в другом исследовании ценности Открытости изменениям поддерживали доверие ИИ в работе со статьями о социальном взаимодействии и государственном секторе, в то время как этого эффекта не наблюдалось при работе ИИ со статьями по здравоохранению или образованию, где требовалось больше ответственных решений [*Shen et al.,* 2025].

Сходство ценностей пользователей с воспринимаемыми ценностями моделей ИИ в литературе рассматривается как ключевой фактор формирования доверия к этим ИИ-инструментам. Такие выводы были получены в экспериментальном исследовании, где участники взаимодействовали с пятью нейросетями, обладающими различными ценностными профилями в ситуациях оценки рисков [*Mehrotra, Jonker, Tielman,* 2021]. ИИ-агенты, чьи ценности в большей степени совпадали с ценностными профилями пользователей, получали более высокие оценки по шкале доверия от респондентов. В другом исследовании авторы предложили модель адаптивной согласованности робота и человеческих ценностей в контексте доверия пользователя [*Bhat et al.,* 2024]. Последующие эмпирические данные показали, что персонализированная настройка ценностного профиля робота в режиме реального времени связана с высоким доверием. Кроме того, адаптивное поведение робота на основании согласования ценностей приводит не только к повышению субъективного доверия, но и к возрастанию воспринимаемой полезности таких технологических решений.

Сходство ценностных профилей нейросетей и человека дополняются исследованиями, анализирующими ценностную согласованность ответов различных крупных языковых моделей (LLM) [*Hadar-Shoval et al.,* 2024; *Segerer,* 2025; *Shen et al.,* 2025]. В ответах всех изучаемых моделей наиболее высокую приоритезацию имели ценности Самопреодоления (такие, как Благожелательность, Универсализм). При этом в социально чувствительных областях ценности Самопреодоления связаны с повышением доверия к ИИ, если его применение не подвергает опасности жизнь и здоровье людей [*Hadar-Shoval et al.,* 2024]. Если риск нарушения конфиденциальности воспринимается как следствие использования технологий, то повышается недоверие к таким системам. Различия были обнаружены по ценностям Самоутверждения, где DeepSeek преуменьшал ценности Власти и Достижений в сравнении с ChatGPT и Gemini. Такая особенность интерпретируется как коллективистская культурная тенденция, отличающаяся от этических систем западных моделей. Интересно, что ценностная согласованность между разными LLM-моделями определяет степень доверия ИИ-инструментов во взаимодействии друг с другом [*Sakamoto, Uchida, Ishiguro,* 2025]. Такие данные особенно важно рассматривать в контексте перспектив широкого внедрения ИИ, когда из взаимодействия будет полностью исключена модерирующая роль человека. Описанные исследования не направлены на изучение связи ценностей и доверия пользователей к ИИ. Однако становится очевидным, что рассмотрение содержания ценностных профилей конкретных нейросетевых инструментов (в том числе ценностная приоритезация, на основании которой происходит генерация ответов) необходима при оценке доверия пользователей к этим инструментам.

Ещё одна важнейшая цифровая область, наряду с ИИ, — сфера криптоинвестиций, представляющая большой экономический интерес для современного общества. При этом именно доверие и культурные ценности влияют на принятие криптовалют, а недоверие негативно связано с интересом к криптовалютам и готовностью их использовать [*Jalan et al.,* 2023]. Кроме того, доверие укрепляет и усиливает связи между осведомлённостью о криптовалюте и готовностью использовать современные финансовые системы [*Shahzad et al.,* 2024]. Исследования показывают, что ценности Открытости изменениям напрямую повышают осведомлённость о существовании криптовалют (вероятность выше в 3–13 раз), а ценности Самоутверждения повышают вероятность покупки на основании доверия к различным криптовалютам (вероятность выше 8–17 раз) [*Stanciu, Partsch, Lechner*, 2024]. В другом исследовании рассматривается связь между ценностями и стереотипами, связанными с банками и криптовалютами [*Hobeika, Liew, Rajan,* 2025]. Было показано, что ценности Самопреодоления (Гедонизм и Стимуляция), а также Безопасность влияют на стереотипы о цифровых финансовых технологиях и доверие к ним. При этом необходимо отметить, что в ряде исследований использование цифровых банковских услуг подразумевало взаимодействие с ИИ, что отчасти позволяет рассматривать такие результаты в контексте ИИ-систем.

Необходимо отметить, что существующие эмпирические работы, посвящённые изучению ценностей как факторов доверия к ИИ, носят фрагментарный характер. Более системный взгляд представлен в работах, посвящённых изучению совпадения ценностных профилей нейросетей и пользователей. Становится очевидным, что такое соотношение собственных ценностей и воспринимаемых ценностей ИИ вносит серьёзный вклад в формирование доверия к автоматизированным системам, позволяя в меньшей степени ориентироваться только на восприятие технических характеристик ИИ. В большинстве работ ценности рассматриваются как устойчивые предикторы доверия технологиям и принятия инноваций. При этом в контексте ИИ доверие зачастую рассматривается как модератор или медиатор связи ценностей в использовании нейросетей. Однако теоретические предпосылки, описанные выше, а также гипотетические обобщения и перспективы эмпирических исследований подчёркивают вклад базовых ценностей в развитие доверия ИИ. На этом основании формируется потребность в комплексных исследованиях ценностных ориентаций, прогнозирующих доверие к инструментам ИИ в разных сферах деятельности.

Моральные основания и доверие к ИИ

Теория моральных оснований (ТМО) была разработана Дж. Хайдтом с коллегами для изучения психологических факторов принятия политических решений в разных культурах, а также анализа социальных проблем [*Graham et al.,* 2011]. Данная теория получила широкое развитие в последние несколько десятилетий, обобщив данные психологических, антропологических и когнитивных исследований механизмов нравственной оценки, а также структуры моральной сферы личности. Ключевым положением ТМО является отсутствие единого принципа, на который опираются моральные суждения. Предполагается, что моральные основания представляют собой условно независимые модули, сформированные в ходе эволюции как адаптивные решения коллективных задач в большинстве культур [*Graham et al.,* 2013]. Именно эти механизмы обеспечивают первичную эмоционально-окрашенную оценку событий и поступков, которая в дальнейшем может обосновываться на когнитивном уровне, преимущественно в интересах межличностного взаимодействия. Первоначально теория обозначала пять базовых оснований морального выбора: забота, справедливость, лояльность группе, авторитет, чистота. Также в рамках ТМО подразумевались пять нарушений принципов, определённых каждым моральным основанием. Это — вред, несправедливость, предательство, подрыв авторитета, деградация. В более поздних редакциях теории было добавлено шестое основание — свобода/угнетение, отражающее чувствительное отношение к сохранению автономии [*Graham et al.,* 2018]. Описанные моральные основания формируют две ключевые группы: индивидуализирующие и сплачивающие. Индивидуализирующие основания морального выбора = это забота и справедливость, которые характеризуются ориентацией на защиту прав личности, автономию и благополучие, а также подчёркивают идеи равенства и универсальности прав человека. К сплачивающим моральным основаниям относят лояльность к группе, авторитет и чистоту — т.е. те принципы, которые направлены на поддержание целостности группы, формирование групповой идентичности, уважение к иерархии власти и коллективным нормам. Таким образом, теория моральных оснований позволяет рассматривать мораль как многомерную, культурно-исторически сформированную, систему интуитивных оценок поведения, событий и социальных явлений, представляя один из наиболее перспективных современных подходов к изучению морально-нравственной сферы.

Повсеместное внедрение различных нейросетевых инструментов неизбежно актуализирует множество этических сложностей, в том числе прозрачность процессов генерации, приватность данных, академический обман, распознание и использование продуктов взаимодействия с искусственным ИИ. Экономические исследования показывают, что этические проблемы (непрозрачность данных, справедливость, ответственность)

препятствуют широкому внедрению нейросетей, негативно влияют на экономический рост, снижают прибыль посредством ограничений моральной приемлемости, что в совокупности приводит к долгосрочному торможению роста ВВП, даже несмотря на оптимистичные сценарии роста производительности [*Gondauri,* 2025]. Перечисленные факторы, зачастую, становятся барьерами взаимодействия с ИИ у представителей различных социальных групп в силу их моральных убеждений. При этом приемлемость моральной оценки предопределяет степень доверия или недоверия пользователей к нейросетям. Теория моральных оснований позволяет прогнозировать вероятные реакции на этическое напряжение, возникающее из-за нравственной оценки действий и результатов работы ИИ, а также определять чувствительность человека к тому или иному этическому принципу.

Классическая теория моральных оснований часто рассматривается через содержательную характеристику нарушений каждого из принципов. В сфере ИИ особое внимание уделяется нарушению трёх базовых принципов — заботы, справедливости и чистоты. Теория моральных оснований описывает причинение вреда через разные виды насилия, к которым относятся физическое, эмоциональное, психологическое. В вопросах взаимодействия с ИИ причинение вреда проявляется в расистских или нецензурных ответах [*Maninger, Shank,* 2022], оценке действий военных [*Malle, Magar, Scheutz,* 2019], некорректном распределении медицинской [*Bigman, Gray,* 2018] или финансовой помощи [*Eubanks,* 2018]. Нарушение принципа справедливости цифровыми агентами обычно выражается в обмане и предоставлении ложной информации, алгоритмической предвзятости, основанной на несправедливом отборе людей или решений [*Eubanks,* 2018]. Обнаруженный факт предоставления ложной информации или несправедливых решений, исходящих от ИИ, значительно повышает недоверие к таким инструментам при дальнейшем взаимодействии. Наиболее характерные для нейросетей нарушения святости (чистоты) выражаются в продвижении сексуализированного контента [*Shank, Gott,* 2020], распространении и адаптации грубой, расистской, сексистской лексики, которую чат-боты могут перенимать у пользователей [*Shank, DeSanti,* 2018]. Однако необходимо отметить, что нарушение авторитета и лояльности в меньшей степени характерно для генеративных моделей. В первую очередь нужно учитывать, что каждый ИИ-инструмент действует под своим брендом, стараясь подчеркнуть свою лояльность к производителю, что обеспечивает ряд маркетинговых преимуществ перед конкурентами. Также важный фактор — низкая, на данном этапе, встроенность нейросетевых агентов в структуры власти, что не позволяет этим инструментам оказывать решающее влияние в ключевых вопросах [*Maninger, Shank,* 2022].

Тематика рассмотрения моральных основ доверия ИИ достаточно нова, однако она вызывает значительный интерес и у исследователей, и у экономических систем, готовых использовать ИИ-решения при выполнении своих задач и оптимизации рабочих процессов. Для изучения характера влияния двух моральных основ (восприятие вреда и восприятие несправедливости) была проведена серия эмпирических исследований [*Sullivan, de Bourmont, Dunaway,* 2022]. Основной целью этой работы было показать, каким образом формируется доверие между человеком и искусственными агентами. Было выявлено, что когнитивные оценки потенциального вреда и несправедливости с моральной точки зрения приводили к возрастанию тревоги и напрямую уменьшали доверие к системам искусственного интеллекта. Авторы подчёркивают, что нарушение значимых для человека моральных основ в цифровом взаимодействии с нейросетями (например, возможная предвзятость ИИ-инструментов или их ущерб) усиливает негативные эмоции и снижает доверие к таким системам.

Помимо изучения прямой взаимосвязи между моральными основаниями и доверием к ИИ в рамках данного обзора целесообразно также рассмотреть исследования моральных оснований как предикторов позитивного отношения и принятия цифровых технологий, что зачастую является следствием доверия к инновациям [*Nagy, Hajdú,* 2021;

Kauttonen, Rousi, Alamäki, 2025]. Так, в результате трёх масштабных исследований (*N*=2209) были обнаружены корреляции между политической идеологий/моральными основаниями и принятием технологических инноваций [*Claudy, Parkinson, Aquino,* 2024]. Для консерваторов были характерны сплачивающие моральные основания (лояльность, чистота, авторитет) и низкое доверие к цифровым новшествам, тогда как для либералов — индивидуализирующие (забота и справедливость) и более высокое доверие к технологиям. Несмотря на это, в вопросах, связанных с инвестициями в криптовалюту, были обнаружены совершенно противоположные эффекты: наибольшее доверие к крипто-инвестициям были выявлены у респондентов с доминирующими связывающими моральными основаниями, такими как авторитет и лояльность [*Banker, Park, Chan,* 2023]. Следовательно, различия в структуре моральных оснований служат значимыми предикторами в содержании технологического поведения и доверия инновациям в цифровой среде.

Исследования в сфере здравоохранения показывают важнейшую роль доверия цифровым медицинским инструментам в контексте оценки этической приемлемости электронной медицины, что как следствие приводит к возрастанию готовности использовать эти нововведения [*Ruelas-Villavicencio et al.,* 2025]. Нарушение приватности в данном контексте свидетельствует о девальвации моральных принципов свободы и справедливости и становится причиной снижения доверия пациентов цифровой медицине. Данные учёных из Финляндии демонстрируют, что наиболее высокие показатели доверия наблюдаются по отношению к неинвазивным инструментам мониторинга на основе ИИ [*Kauttonen, Rousi, Alamäki,* 2025], при использовании которых не возникает моральных дилемм, а ИИ не принимает никаких этических решений. Стоит также обратить внимание, что женщины демонстрировали большее недоверие и аккуратность по отношению к ИИ в системе здравоохранения. Этот эффект усиливался при рассмотрении таких модераторов, как уровень образования, установки к технологиям и условия использования.

В фокусе замены сотрудников, занятых в ключевых рабочих процессах, на ИИ-помощников также обнаруживается серия моральных дилемм. Качественное исследование включало в себя анализ интервью соискателей, которые взаимодействовали с нейросетевым HR-ассистентом [*Mirowska, Arsenyan,* 2025]. Обнаружено, что при процессе собеседования на новую должность нарушение всех шести моральных оснований косвенно приводило к снижению доверия ИИ-решениям (усиливались негативные реакции на ИИ, появлялось «ощущение неправильности»). При нарушении морального принципа «забота» ИИ воспринимался как «холодный, дегуманизированный», он общался с кандидатом как с числами, а при нарушении морального основания «справедливость» — снижалось доверие к честности процесса отбора. В ряде случаев респонденты отказывались участвовать в такого рода собеседованиях, аргументируя это нарушением авторитета человека.

Отдельный вопрос — доверие ИИ при решении этически сложных проблем. Согласно данным исследовательских отчётов, ключевое влияние на принятие решений при возникновении ситуаций необходимости решения моральных дилемм свою роль играют подсказки от автоматизированных систем [*Salatino et al.,* 2025]. Кроме того, ответы чат-ботов влияют не только на преодоление спорных ситуаций, но и на смену ощущения ответственности за совершённые действия, в том числе размывают субъектную позицию пользователя. Широко изучаемая тема — сравнение моральных убеждений, вложенных в различные генеративные модели, и то, насколько эти нравственные принципы влияют на генерируемые ответы. Несмотря на общую аккуратность и осторожность генерируемых ответов, при их сопоставлении обнаруживаются элементы лицемерного и непоследовательного поведения абстрактных моральных представлений ИИ и конкретных ситуаций нравственного выбора [*Nunes et al.,* 2024]. Такие проявления нарушений принципов справедливости и честности при ответах ИИ приводят к снижению к нему доверия, а также к готовности пользователей использовать конкретные

генеративные модели. Эмпирически подтверждено [*Bajpai, Sameer, Fatima,* 2025] и то, что на данном этапе развития технологий чат-боты не способны достичь присущего человеку уровня моральных рассуждений или комплексной оценки этически неочевидных дилемм. Хотя нельзя не учитывать, что рассмотрение данного тезиса подразумевает учёт переоценки людьми уникальности своего морального профиля, а это оказывается ключевым фактором недоверия к моральным оценкам ИИ [*Purcell, Bonnefon,* 2023], а также одним из факторов снижения доверия к инструментам ИИ в системе здравоохранения [*Longoni, Bonezzi, Morewedge* 2019].

Таким образом, существующие эмпирические работы показывают множество эмпирически подтверждённых взаимосвязей между моральными основаниями и доверием к ИИ. Совокупность индивидуализирующих и сплачивающих оснований морального выбора предсказывают доверие к различным нейросетевым инструментам. Тем не менее прогностичность развивающихся в этой сфере процессов во многом зависит от контекстуальных сценариев оценки действий ИИ, культурных и идеологических различий, а также методологических особенностей исследований. Основные предикторы доверия к ИИ в индивидуалистических культурах — моральные основания справедливости и заботы. Однако пренебрежение такими основаниями, как авторитет группы и святость (чистота) могут приводить к снижению доверия ИИ в консервативных сообществах. Становится очевидным, что влияние моральных оснований на оценку инструментов ИИ и доверие к ним необходимо рассматривать как многофакторный процесс, с учётом когнитивных представлений об ИИ-системах, ценностях, наличии установок на взаимодействие, а также сведений о контексте их применения. Отдельным дискуссионным вопросом остаётся этичность замены работников и оптимизация ряда процессов с помощью ИИ-ассистентов, а также моральная оценка и доверие сотрудников организации руководству, предпринимающему такие действия. Для разработчиков технологических

Таблица 1
Ценности и моральные основания как факторы доверия искусственному интеллекту

Социально-психологические факторы		**Связь с доверием к ИИ**	**Характеристика связи**
Ценности	Открытость изменениям (Openness to change)	Положительная	Ценности Открытости изменениям повышают доверие за счёт ориентации личности на новое, а также готовности к неопределённости. Искусственный интеллект рассматривается в качестве возможностей и инноваций
	Самоутверждение (Self-enhancement)	Положительная	Ценности данного блока повышают доверие в ситуациях использования искусственного интеллекта для расширения собственных возможностей, роста эффективности и достижения своих целей. Эффект усиливается при рассмотрении доверия к ИИ в профессиональных контекстах
	Самопреодоление (Self-transcendence)	Противоречивая	С одной стороны, эти ценности положительно связаны с повышением доверия в ситуациях, когда искусственный интеллект воспринимается в контексте социальной полезности обществу и характеризуется просоциальной направленностью. С другой стороны, наличие рисков причинения вреда или нарушения справедливости со стороны ИИ снижает доверие к нейросетям

Окончание табл. 1

Социально-психологические факторы		Связь с доверием к ИИ	Характеристика связи
Ценности	Сохранение (Conservation)	Отрицательная	Систематически снижает доверие из-за ориентации личности на стабильность, предсказуемость и соблюдение консервативных норм. Данные ценности связаны с восприятием искусственного интеллекта как источника угроз и неопределённости. При этом такая связь может модерироваться культурными особенностями
Моральные основания	Индивидуализирующие (Individualizing)	Противоречивая	Индивидуализирующие моральные основания (в особенности, забота и справедливость) формируют условное доверие, зависящее от оценки этичности ИИ: усиливают доверие при ожидании общественной пользы и справедливости, но резко снижают его при восприятии вреда или нарушения моральных норм
	Сплачивающие (binding)	Отрицательная	Сплачивающие моральные основания снижают доверие в ситуациях, когда использование ИИ нарушает групповую иерархию или подрывает авторитет, особенно в консервативных сообществах

Источник: составлено авторами на основании проведённого обзора литературы.

инноваций следует предусматривать механизмы адаптивной прозрачности, более чёткие алгоритмы конфиденциальности данных, а также возможности персонализации, снижающие этическое напряжение, которое может возникнуть между нейросетевым продуктом и моральными суждениями пользователей. Всё это повышает уровень доверия к используемым цифровым продуктам.

Результаты проведённого теоретического анализа современных исследований в контексте социально-психологических факторов доверия искусственному интеллекту обобщены в табл. 1.

Заключение

Таким образом, проведённый анализ литературы показывает, что социально-психологические факторы серьёзно влияют на формирование доверия к ИИ. Их необходимо учитывать наряду с когнитивными факторами, связанными с восприятием и оценкой нейросетей. В структуре социально-психологических факторов доверия к искусственному интеллекту наибольшую значимость имеют ценности и моральные основания.

При рассмотрении связи ценностей и доверия к искусственному интеллекту нами были обнаружены противоречивые результаты. С одной стороны, ценности Сохранения отрицательно связаны с доверием к нейросетевым инструментам. Отмечены значительная осторожность и скептицизм при взаимодействии с новыми технологиями. Ориентация на ценности Безопасность и Традиции сопрягается с возрастанием недоверия к ИИ и формированием негативных стереотипов о действиях автоматизированных систем. С другой стороны, такие взаимосвязи не всегда прогностичны, а их направление может модерироваться культурными особенностями. Например, на китайской выборке наблюдаются обратные тенденции, и приоритет ценностей Сохранения связан с высоким доверием ИИ. Ценности Открытости изменениям в большинстве случаев положительно связаны

с доверием к ИИ, за исключением ситуаций высоких рисков нарушения надёжности результатов работы, когда корреляция утрачивается, но не становится отрицательной. Ценности Самопреодоления также дают двойственный эффект при формировании доверия ИИ-технологиям. Если ИИ-инструменты направлены на достижение общественных благ и помощь людям в чувствительных областях, то пользователи склонны доверять таким технологиям. В то же время, если применение ИИ сопряжено с возрастанием рисков здоровью или конфиденциальности, то доверие снижается. Ценности Самоутверждения могут быть позитивно связаны с доверием ИИ в тех ситуациях, когда пользователи рассматривают ИИ-системы как эффективные инструменты достижения целей и расширения собственного потенциала.

Конгруэнтность между ценностями пользователей и воспринимаемыми ценностями ИИ — одна из наиболее изучаемых областей приложения ценностной теории Ш. Шварца к рассматриваемой сфере. Ценностная конгруэнтность выступает сильным положительным предиктором доверия к автоматизированным системам и значительно превосходит объяснительную силу воспринимаемых технологических предикторов. Следовательно, складывается тенденция высокой приоритезации адаптивной настройки LLM под ценностные профили пользователей при дальнейших разработках в сфере ИИ. Точечная корректировка ответов на основании ценностей пользователей значительно повышает доверие к таким инструментам, что приводит к снижению напряжения при внедрении инноваций и повышению экономической выгоды от использования ИИ-технологий.

В структуре моральных оснований обнаружены более устойчивые и однозначные связи индивидуализирующих и сплачивающих принципов и доверия к ИИ-технологиям. Наиболее чувствительными моральными основаниями при формировании доверия ИИ выступают забота/вред и справедливость/несправедливость. Пользователи, которые ставят в приоритет индивидуализирующие моральные основания, склонны доверять искусственному интеллекту, если это связано с помощью людям и улучшением их благополучия, но нарушение этих принципов (несправедливость или причинение вреда) приводит к снижению доверия ИИ-системам. Пользователи, ориентированные на сплачивающие моральные основания (авторитет, чистота, лояльность), в меньшей степени склонны доверять инструментам ИИ, особенно в социально-чувствительных сферах и личном использовании.

Перспективы исследований данной области обладают большим психологическим и экономическим потенциалом. Они должны быть сфокусированы вокруг создания динамической системной модели доверия ИИ, которая будет способна адаптироваться к стремительно меняющимся моральным и ценностным контекстам глобального цифрового сообщества.

ЛИТЕРАТУРА

Винокуров Ф.Н., Садовская Е.Д. (2023). Экспериментальное сравнение доверия искусственному интеллекту и человеку в экономических решениях // *Экспериментальная психология.* Т. 6. №2. С. 87–100. DOI: 10.17759/exppsy.2023160206.

Казун А.Д., Поршнев А.В. (2021). Кто верит в теории заговора? Факторы склонности к конспирологическому мышлению в России, Казахстане и Украине // *Мониторинг общественного мнения: экономические и социальные перемены*. №6. С. 549–565. DOI: 10.14515/monitoring.2021.6.1889.

Лукичев П.М. (2024). Принятие решений в современной экономике: искусственный интеллект *vs* поведенческая экономика // *Вопросы инновационной экономики.* Т.14 №3. С. 649–666. DOI: 10.18334/vinec.14.3.121070.

Макаров М.Ю. (2020). Влияние искусственного интеллекта на производительность труда // *Экономика и управление*. Т. 26. №5. С. 479–486. DOI: 10.35854/1998-1627-2020-5-479-486.

Манахова И.В., Маковская А.М. (2025). Человек и искусственный интеллект в дискурсе поведенческой экономики // *Вестник Московского университета. Серия 6. Экономика*. Т. 60. №3. С. 3–19. DOI: 10.55959/MSU0130-0105-6-60-3-1.

Петрунин Ю.Ю., Нуралиева Н.З. (2025). Доверие к технологиям генеративного искусственного интеллекта как зеркало доверия к институтам // *Государственное управление: электронный вестник*. №113. С. 22–30. DOI: 10.55959/MSU2070-1381-113-2025-22-30.

Татарко А.Н. (2017). Взаимосвязь базовых человеческих ценностей и электорального поведения // *Социальная психология и общество*. Т. 8. №1. С. 17–37. DOI: 10.17759/sps.2017080102.

Aghion P., Jones B.F., Jones C.I. (2017). Artificial Intelligence and Economic Growth // *NBER Working Papers*. No. 23928. DOI: 10.3386/w23928.

Anisha S.A., Sen A., Ahmad B. et al. (2025). Exploring Acceptance of Digital Health Technologies for Managing Non-Communicable Diseases Among Older Adults: A Systematic Scoping Review // *Journal of Medical Systems*. Vol. 49. Article 35. DOI: 10.1007/s10916-025-02166-3.

Awad E., Dsouza S., Kim R., Schulz J., Henrich J., Shariff A., Bonnefon J.F., Rahwan I. (2018). The Moral Machine experiment // *Nature*. Vol. 563. No. 7729. Pp. 59–64. DOI: 10.1038/s41586-018-0637-6.

Bai X., Yang L. (2025). Research on the influencing factors of generative artificial intelligence usage intent in post-secondary education: An empirical analysis based on the AIDUA extended model // *Frontiers in Psychology*. Vol. 16. Article 1644209. DOI: 10.3389/fpsyg.2025.1644209.

Bajpai S., Sameer A., Fatima R. (2025). Insights into Moral Reasoning of AI: A Comparative Study Between Humans and Large Language Models // *Journal of Media Ethics*. Pp. 1–15. DOI: 10.1080/23736992.2025.2553146.

Banker S., Park J., Chan E.Y. (2023). The moral foundations of cryptocurrency: Evidence from Twitter and survey research // *Frontiers in Psychology*. Vol. 14. Article 1128575. DOI: 10.3389/fpsyg.2023.1128575.

Bhat S. et al. (2024). Value alignment and trust in human-robot interaction: Insights from simulation and user study // *Discovering the frontiers of human-robot interaction: Insights and innovations in collaboration, communication, and control* / R. Vinjamuri (ed.). — Cham: Springer Nature Switzerland. Pp. 39–63. DOI: 10.1007/978-3-031-66656-8_3.

Bigman Y.E., Gray K. (2018). People are averse to machines making moral decisions // *Cognition*. Vol. 181. Pp. 21–34. DOI: 10.1016/j.cognition.2018.08.003.

Cheung J.C., Ho S.S. (2025). The effectiveness of explainable AI on human factors in trust models // *Scientific Reports*. Vol. 15. Article 23337. DOI: 10.1038/s41598-025-04189-9.

Claudy M.C., Parkinson M., Aquino K. (2024). Why should innovators care about morality? Political ideology, moral foundations, and the acceptance of technological innovations // *Technological Forecasting and Social Change*. Vol. 203. Article 123384. DOI: 10.1016/j.techfore.2024.123384.

Dietvorst B.J., Simmons J.P., Massey C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err // *Journal of experimental psychology. General*. Vol. 144. No. 1. Pp. 114–126. DOI: 10.1037/xge0000033.

Eubanks V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. — New York: St. Martin's Press.

Filippucci F. et al. (2024). The impact of Artificial Intelligence on productivity, distribution and growth: Key mechanisms, initial evidence and policy challenges // *OECD Artificial Intelligence Papers*. No. 15 / OECD Publishing. — Paris. DOI: 10.1787/8d900037-en.

Gerke S., Minssen T., Cohen G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare // *Artificial Intelligence in Healthcare* / A. Bohr, K. Memarzadeh (eds.). — Elsevier. Pp. 295–336. DOI: 10.1016/B978-0-12-818438-7.00012-5.

Glikson E., Woolley A.W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research // *Academy of Management Annals*. Vol. 14. No. 2. Pp. 627–660. DOI: 10.5465/annals.2018.0057.

Gondauri D. (2025). The impact of artificial intelligence on gross domestic product: A global analysis // arXiv preprint. ArXiv:2505.11989. DOI: 10.48550/arXiv.2505.11989.

Graham J., Nosek B.A., Haidt J., Iyer R., Koleva S., Ditto P.H. (2011). Mapping the Moral Domain // *Journal of Personality and Social Psychology.* Vol. 101. No. 2. Pp. 366–385. DOI: 10.1037/a0021847.

Graham J., Haidt J., Koleva S., Motyl M., Iyer R., Wojcik S., Ditto P.H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism // *Advances in Experimental Social Psychology*. Vol. 47. Pp. 55–130. DOI: 10.1016/B978-0-12-407236-7.00002-4.

Graham J., Haidt J., Motyl M., Meindl P., Iskiwitch C., Mooijman M. (2018). Moral Foundations Theory // *Atlas of Moral Psychology* / K. Gray, J. Graham (eds.). — New York, London: The Guilford Press. Pp. 211–222.

Gursoy D., Della Corte V., del Gaudio G., Crisci A., Xu, Y. (2025). Factors influencing elderly adoption of artificial intelligence robots for aging-in-place: Motivators, barriers, and emotional impact // *Work, Aging and Retirement.* Article waaf016. DOI: 10.1093/workar/waaf016.

Hadar-Shoval D., Asraf K., Mizrachi Y., Haber, Y., Elyoseph Z. (2024). Assessing the Alignment of Large Language Models With Human Values for Mental Health Integration: Cross-Sectional Study Using Schwartz's Theory of Basic Values // *JMIR Mental Health*. Vol. 11. Article e55988. DOI: 10.2196/55988.

Hobeika J, Liew C.Y, Rajan M.E.S (2025). Are cryptocurrencies influenced by stereotypes and values? Evidence of the role of banker stereotypes and human values on cryptocurrency acceptance // *International Journal of Bank Marketing*. Vol. 43. No. 8. Pp. 1627–1661, DOI: 10.1108/IJBM-04-2024-0181.

Horowitz M.C., Kahn L. (2021). What influences attitudes about artificial intelligence adoption: Evidence from U.S. local officials // *PLOS ONE.* Vol. 16. No. 10. Article e0257732. DOI: 10.1371/journal.pone.0257732.

Jalan A., Matkovskyy R., Urquhart A., Yarovaya L. (2023). The role of interpersonal trust in cryptocurrency adoption // *Journal of International Financial Markets, Institutions and Money*. Vol. 83. Article 101715. DOI: 10.1016/j.intfin.2022.101715.

Jobin A., Ienca M., Vayena E. (2019). The global landscape of AI ethics guidelines // *Nature Machine Intelligence*. Vol. 1. Pp. 389–399. DOI: 10.1038/s42256-019-0088-2.

Kai C., Ping W., Xiaomin J. (2026). AI anxiety and adoption intention in higher education based on an extended TAM-UTAUT and PLS-SEM analysis // *Scientific Reports*. Vol. 16. Article 3672. DOI: 10.1038/s41598-026-35823-9.

Kauttonen J., Rousi R., Alamäki A. (2025). Trust and Acceptance Challenges in the Adoption of AI Applications in Health Care: Quantitative Survey Analysis // *Journal of Medical Internet Research*. Vol. 27. Article e65567. DOI: 10.2196/65567.

Lammert D., Liu M., Betz S., Lammert J., Pfeffer J. (2026). Culturally-Aware Artificial Intelligence: Personal Values and Technology Acceptance among AI Researchers in China and Germany // *EAI Endorsed Transactions on Internet of Things*. Vol. 11. DOI: 10.4108/eetiot.10618.

Lep Ž., Trunk A., Babnik K. (2022). Value Orientations and Institutional Trust as Contributors to the Adoption of Online Services in Youth: A Cross-Country Comparison // *Frontiers in Psychology*. Vol. 13. Article 887587. DOI: 10.3389/fpsyg.2022.887587.

Li H., Wei X. (2025). Factors influencing older adults' adoption of AI voice assistants: Extending the UTAUT model // *Frontiers in Psychology*. Vol. 16. Article 1618689. DOI: 10.3389/fpsyg.2025.1618689.

Longoni C., Bonezzi A., Morewedge C.K. (2019). Resistance to medical artificial intelligence // *Journal of Consumer Research*. Vol. 46. No. 4. Pp. 629–650. DOI: 10.1093/jcr/ucz013.

Malle B.F., Magar S.T., Scheutz M. (2019). AI in the sky: How people morally evaluate human and machine decisions in a lethal strike dilemma // *Robotics and Well-Being. Intelligent Systems, Control and Automation: Science and Engineering* / A.M. Ferreira, S.J. Sequeira, S.G. Virk, M. Tokhi, E.E. Kadar (eds). — Vol. 95. — Cham: Springer International Publishing. Pp. 111–133. DOI: 10.1007/978-3-030-12524-0_11.

Maninger T., Shank D.B. (2022). Perceptions of violations by artificial and human actors across moral foundations // *Computers in Human Behavior Reports*. Vol. 5. Article 100154. DOI: 10.1016/J.CHBR.2021.100154.

Massachusetts Institute of Technology. (2025). *The GenAI divide: State of Business 2025*. URL: ai_report_2025.pdf (access date: 26.02.2026).

Mayer R.C., Davis J.H., Schoorman F.D. (1995). An integrative model of organizational trust // *Academy of Management Review*. Vol. 20. No. 3. Pp. 709–734. DOI: 10.5465/amr.1995.9508080335.

Mehrotra S., Jonker C.M., Tielman M.L. (2021). *More Similar Values, More Trust? — The Effect of Value Similarity on Trust in Human-Agent Interaction*. Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society. Association for Computing Machinery. — New York. Pp. 777–783. DOI: 10.1145/3461702.3462576.

Mirowska A., Arsenyan J. (2025). «It Feels Wrong»: Understanding Reactions to Artificial Intelligence as a Decision-Maker in Selection Through the Lens of Moral Foundations Theory // *International Journal of Selection and Assessment*. Vol. 34. Article e70039. DOI: 10.1111/ijsa.70039.

Morselli D., Spini D., Devos T. (2012). Human Values and Trust in Institutions across Countries: A Multilevel Test of Schwartz's Hypothesis of Structural Equivalence // *Survey Research Methods*. Vol. 6. No. 1. Pp. 49–60. DOI: 10.18148/srm/2012.v6i1.5090.

Nagy S., Hajdú N. (2021). Consumer acceptance of the use of artificial intelligence in online shopping: Evidence from Hungary // *Amfiteatru Economic*. Vol. 23. No. 56. Pp. 155–173. DOI:10.24818/EA/2021/56/155.

Nunes J.L. et al. (2024). Are large language models moral hypocrites? A study based on moral foundations // *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.* Vol. 7. No. 1. Pp. 1074–1087. DOI: 10.48550/arXiv.2405.11100.

Panari C., Lorenzi G., Mariani M.G. (2021). The Predictive Factors of New Technology Adoption, Workers' Well-Being and Absenteeism: The Case of a Public Maritime Company in Venice // *International Journal of Environmental Research and Public Health*. Vol. 8. No. 23. Article 12358. DOI: 10.3390/ijerph182312358.

Purcell Z.A., Bonnefon J.F. (2023). Humans feel too special for machines to score their morals // *PNAS nexus*. Vol. 2. No. 6. Article pgad179. DOI: 10.1093/pnasnexus/pgad179.

Robinette P. et al. (2016). Overtrust of robots in emergency evacuation scenarios // *11th ACM/IEEE International Conference on Human-Robot Interaction* (HRI). — Christchurch, New Zealand. Pp. 101–108. DOI: 10.1109/HRI.2016.7451740.

Ruelas-Villavicencio A.L., Contreras-Yáñez I., Gómez-Ruiz R.P., Zagaglia Del Valle M.C., Malagón-Liceaga A., Pascual-Ramos V. (2025). Digital health literacy is linked to attitudes regarding the ethical aspects of digital health among patients with dermatologic comorbidities // *PLOS ONE*. Vol. 20. No. 9. DOI: 10.1371/journal.pone.0330916.

Sakamoto Y., Uchida T., Ishiguro H. (2025). Value-based large language model agent simulation for mutual evaluation of trust and interpersonal closeness // *Scientific Reports*. Vol. 15. Article 41653. DOI: 10.1038/s41598-025-25531-1.

Salatino A., Prével A., Caspar E. et al. (2025). Influence of AI behavior on human moral decisions, agency, and responsibility // *Scientific Reports*. Vol. 15. Article 12329. DOI: 10.1038/s41598-025-95587-6.

Schwartz S.H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries // *Advances in Experimental Social Psychology*. Vol. 25. No. 1. Pp. 1–65. DOI: 10.1016/S0065-2601(08)60281-6.

Schwartz S.H., Cieciuch J., Vecchione M., Davidov E., Fischer R., Beierlein C., Ramos A., Verkasalo M., Lönnqvist J.-E., Demirutku K., Dirilen-Gumus O., Konty M. (2012). Refining the theory of basic individual values // *Journal of Personality and Social Psychology.* Vol. 103. No. 4. Pp. 663–688. DOI: 10.1037/a0029393.

Segerer R. (2025). Cultural value alignment in large language models: A prompt-based analysis of Schwartz values in Gemini, ChatGPT, and DeepSeek // *arXiv preprint.* ArXiv:2505.17112. DOI: 10.48550/arXiv.2505.17112.

Shahzad M.F., Xu S., Lim W.M. et al. (2024). Cryptocurrency awareness, acceptance, and adoption: The role of trust as a cornerstone // *Humanities and Social Sciences Communications*. Vol. 11. Article 4. DOI: 10.1057/s41599-023-02528-7.

Shank D.B., DeSanti A. (2018). Attributions of morality and mind to artificial intelligence after real-world moral violations // *Computers in Human Behavior*. Vol. 86. Pp. 401–411. DOI: 10.1016/J.CHB.2018.05.014.

Shank D.B., Gott A. (2020). Exposed by AIs! People personally witness artificial intelligence exposing personal information and exposing people to undesirable content // *International Journal of Human–Computer Interaction*. Vol. 36. No. 17. Pp. 1636–1645. DOI: 10.1080/10447318.2020.1768674.

Shen H. et al. (2025). *ValueCompass: A framework for measuring contextual value alignment between human and LLMs* / Proceedings of the 9th Widening NLP Workshop. Pp. 75-86. DOI: 10.48550/arXiv.2409.09586.

Stanciu A., Partsch M., Lechner C.M. (2024). Basic human values and the adoption of cryptocurrency // *Frontiers in Psychology*. Vol. 15. Article 1395674. DOI: 10.3389/fpsyg.2024.1395674

Stein J.P., Messingschlager T., Gnambs T. et al. (2024). Attitudes towards AI: measurement and associations with personality // *Scientific Reports*. Vol. 14. Article 2909. DOI: 10.1038/s41598-024-53335-2.

Sullivan Y., de Bourmont M., Dunaway M. (2022). Appraisals of harms and injustice trigger an eerie feeling that decreases trust in artificial intelligence systems // *Annals of Operations Research*. Vol. 308. Pp. 525–548. DOI: 10.1007/s10479-020-03702-9.

Tang L., Ferronato P., Bashir M. (2023) Lecture Notes in Computer Science). Do Users' Values Influence Trust in Automation? // *Intelligent Human Computer Interaction. IHCI 2022. Lecture Notes in Computer Science.* / H. Zaynidinov, M. Singh, U.S. Tiwary, D. Singh (eds). — Cham: Springer. Vol. 13741. DOI: 10.1007/978-3-031-27199-1_30.

Ulfert A.S., Antoni C.H., Ellwart T. (2022). The role of agent autonomy in using decision support systems at work // *Computers in Human Behavior*. Vol. 126. Article 106987. DOI:10.1016/j.chb.2021.106987.

Vuori N., Burkhard B., Pitkäranta L. (2025). It's Amazing–But Terrifying!: Unveiling the Combined Effect of Emotional and Cognitive Trust on Organizational Member'Behaviours, AI Performance, and Adoption // *Journal of Management Studies*. Vol. 63. No. 2. Pp. 473–514. DOI: 10.1111/joms.13177.

Wen Y., Wang J., Chen X. (2025). Trust and AI weight: Human-AI collaboration in organizational management decision-making // *Frontiers in Organizational Psychology*. Vol. 3. Article 1419403. DOI: 10.3389/forgp.2025.1419403.

Witte E.H., Stanciu A., Boehnke K. (2020). A New Empirical Approach to Intercultural Comparisons of Value Preferences Based on Schwartz's Theory // *Frontiers in Psychology*. Vol. 11. Article 1723. DOI: 10.3389/fpsyg.2020.01723.

Xiong Y. et al. (2023). More trust or more risk? User acceptance of artificial intelligence virtual assistant // *Human Factors and Ergonomics In Manufacturing*. Vol. 34. No. 3. Pp. 190–205. DOI: 10.1002/hfm.21020.

You Y., Hu Y., Yang W., Cao S. (2022). Research on the Influence Path of Online Consumers' Purchase Decision Based on Commitment and Trust Theory // *Frontiers in Psychology*. Vol. 13. Article 916465. DOI: 10.3389/fpsyg.2022.916465.

Zhang Q., Lee M.L., Carter S. (2022). You complete me: Human-ai teams and complementary expertise // *CHI Conference on Human Factors in Computing Systems*. Article 114. Pp. 1–28. DOI: 10.1145/3491102.3517791.

Zhang Q., Wang F., Liao G., Li M. (2025). How Does AI Trust Foster Innovative Performance Under Paternalistic Leadership? The Roles of AI Crafting and Leader's AI Opportunity Perception // *Behavioral Sciences*. Vol. 15. No. 8. Article 1064. DOI: 10.3390/bs15081064.

REFERENCES

Aghion P., Jones B.F., Jones C.I. (2017). Artificial Intelligence and Economic Growth // *NBER Working Papers.* No. 23928. DOI: 10.3386/w23928.

Anisha S.A., Sen A., Ahmad B. et al. (2025). Exploring Acceptance of Digital Health Technologies for Managing Non-Communicable Diseases Among Older Adults: A Systematic Scoping Review // *Journal of Medical Systems*. Vol. 49. Article 35. DOI: 10.1007/s10916-025-02166-3.

Awad E., Dsouza S., Kim R., Schulz J., Henrich J., Shariff A., Bonnefon J.F., Rahwan I. (2018). The Moral Machine experiment // *Nature*. Vol. 563. No. 7729. Pp. 59–64. DOI: 10.1038/s41586-018-0637-6.

Bai X., Yang L. (2025). Research on the influencing factors of generative artificial intelligence usage intent in post-secondary education: An empirical analysis based on the AIDUA extended model // *Frontiers in Psychology*. Vol. 16. Article 1644209. DOI: 10.3389/fpsyg.2025.1644209.

Bajpai S., Sameer A., Fatima R. (2025). Insights into Moral Reasoning of AI: A Comparative Study Between Humans and Large Language Models // *Journal of Media Ethics*. Pp. 1–15. DOI: 10.1080/23736992.2025.2553146.

Banker S., Park J., Chan E.Y. (2023). The moral foundations of cryptocurrency: Evidence from Twitter and survey research // *Frontiers in Psychology*. Vol. 14. Article 1128575. DOI: 10.3389/fpsyg.2023.1128575.

Bhat S. et al. (2024). Value alignment and trust in human-robot interaction: Insights from simulation and user study // *Discovering the frontiers of human-robot interaction: Insights and innovations in collaboration, communication, and control* / R. Vinjamuri (ed.). — Cham: Springer Nature Switzerland. Pp. 39-63. DOI: 10.1007/978-3-031-66656-8_3.

Bigman Y.E., Gray K. (2018). People are averse to machines making moral decisions // *Cognition*. Vol. 181. Pp. 21–34. DOI: 10.1016/j.cognition.2018.08.003.

Cheung J.C., Ho S.S. (2025). The effectiveness of explainable AI on human factors in trust models // *Scientific Reports*. Vol. 15. Article 23337. DOI: 10.1038/s41598-025-04189-9.

Claudy M.C., Parkinson M., Aquino K. (2024). Why should innovators care about morality? Political ideology, moral foundations, and the acceptance of technological innovations // *Technological Forecasting and Social Change*. Vol. 203. Article 123384. DOI: 10.1016/j.techfore.2024.123384.

Dietvorst B.J., Simmons J.P., Massey C. (2015). Algorithm aversion: people erroneously avoid algorithms after seeing them err // *Journal of experimental psychology. General*. Vol. 144. No. 1. Pp. 114–126. DOI: 10.1037/xge0000033.

Eubanks V. (2018). *Automating inequality: How high-tech tools profile, police, and punish the poor*. — New York: St. Martin's Press.

Filippucci F. et al. (2024). The impact of Artificial Intelligence on productivity, distribution and growth: Key mechanisms, initial evidence and policy challenges // *Artificial Intelligence Papers*. No. 15. OECD Publishing. Paris. DOI: 10.1787/8d900037-en.

Gerke S., Minssen T., Cohen G. (2020). Ethical and legal challenges of artificial intelligence-driven healthcare // *Artificial Intelligence in Healthcare* / A. Bohr, K. Memarzadeh (eds.). — Elsevier. Pp. 295–336. DOI: 10.1016/B978-0-12-818438-7.00012-5.

Glikson E., Woolley A.W. (2020). Human Trust in Artificial Intelligence: Review of Empirical Research // *Academy of Management Annals*. Vol. 14. No. 2. Pp. 627–660. DOI: 10.5465/annals.2018.0057.

Gondauri D. (2025). The impact of artificial intelligence on gross domestic product: A global analysis // *arXiv preprint*. ArXiv:2505.11989. DOI: 10.48550/arXiv.2505.11989.

Graham J., Nosek B.A., Haidt J., Iyer R., Koleva S., Ditto P.H. (2011). Mapping the Moral Domain // *Journal of Personality and Social Psychology.* Vol. 101. No. 2. Pp. 366–385. DOI: 10.1037/a0021847.

Graham J., Haidt J., Koleva S., Motyl M., Iyer R., Wojcik S., Ditto P.H. (2013). Moral foundations theory: The pragmatic validity of moral pluralism // *Advances in Experimental Social Psychology*. Vol. 47. Pp. 55–130. DOI: 10.1016/B978-0-12-407236-7.00002-4.

Graham J., Haidt J., Motyl M., Meindl P., Iskiwitch C., Mooijman M. (2018). Moral Foundations Theory // *Atlas of Moral Psychology* / K. Gray, J. Graham (eds.). — New York, London: The Guilford Press. Pp. 211–222.

Gursoy D., Della Corte V., del Gaudio G., Crisci A., Xu, Y. (2025). Factors influencing elderly adoption of artificial intelligence robots for aging-in-place: Motivators, barriers, and emotional impact // *Work, Aging and Retirement.* Article waaf016. DOI: 10.1093/workar/waaf016.

Hadar-Shoval D., Asraf K., Mizrachi Y., Haber, Y., Elyoseph Z. (2024). Assessing the Alignment of Large Language Models With Human Values for Mental Health Integration: Cross-Sectional Study Using Schwartz's Theory of Basic Values // *JMIR Mental Health*. Vol. 11. Article e55988. DOI: 10.2196/55988.

Hobeika J, Liew C.Y, Rajan M.E.S (2025). Are cryptocurrencies influenced by stereotypes and values? Evidence of the role of banker stereotypes and human values on cryptocurrency acceptance // *International Journal of Bank Marketing*. Vol. 43. No. 8. Pp. 1627–1661. DOI: 10.1108/IJBM-04-2024-0181.

Horowitz M.C., Kahn L. (2021). What influences attitudes about artificial intelligence adoption: Evidence from U.S. local officials // *PLOS ONE.* Vol. 16. No. 10. Article e0257732. DOI: 10.1371/journal.pone.0257732.

Jalan A., Matkovskyy R., Urquhart A., Yarovaya L. (2023). The role of interpersonal trust in cryptocurrency adoption // *Journal of International Financial Markets, Institutions and Money*. Vol. 83. Article 101715. DOI: 10.1016/j.intfin.2022.101715.

Jobin A., Ienca M., Vayena E. (2019). The global landscape of AI ethics guidelines // *Nature Machine Intelligence*. Vol. 1. Pp. 389–399. DOI: 10.1038/s42256-019-0088-2.

Kai C., Ping W., Xiaomin J. (2026). AI anxiety and adoption intention in higher education based on an extended TAM-UTAUT and PLS-SEM analysis // *Scientific Reports*. Vol. 16. Article 3672. DOI: 10.1038/s41598-026-35823-9.

Kauttonen J., Rousi R., Alamäki A. (2025). Trust and Acceptance Challenges in the Adoption of AI Applications in Health Care: Quantitative Survey Analysis // *Journal of Medical Internet Research*. Vol. 27. Article e65567. DOI: 10.2196/65567.

Kazun A.D., Porshnev A.V. (2021). Who Believes in Conspiracy Theories? Factors Influencing Propensity for Conspiracy Thinking in Russia, Kazakhstan and Ukraine // *Monitoring of Public Opinion: Economic and Social Changes.* No. 6. Pp. 549–565. DOI: 10.14515/monitoring.2021.6.1889 (In Russ.).

Lammert D., Liu M., Betz S., Lammert J., Pfeffer, J. (2026). Culturally-Aware Artificial Intelligence: Personal Values and Technology Acceptance among AI Researchers in China and Germany // *EAI Endorsed Transactions on Internet of Things*. Vol. 11. DOI: 10.4108/eetiot.10618

Lep Ž., Trunk A., Babnik K. (2022). Value Orientations and Institutional Trust as Contributors to the Adoption of Online Services in Youth: A Cross-Country Comparison // *Frontiers in Psychology*. Vol. 13. Article 887587. DOI: 10.3389/fpsyg.2022.887587

Li H., Wei X. (2025). Factors influencing older adults' adoption of AI voice assistants: Extending the UTAUT model // *Frontiers in Psychology*. Vol. 16. Article 1618689. DOI: 10.3389/fpsyg.2025.1618689.

Longoni C., Bonezzi A., Morewedge C.K. (2019). Resistance to medical artificial intelligence // *Journal of Consumer Research*. Vol. 46. No. 4. Pp. 629–650. DOI: 10.1093/jcr/ucz013.

Lukichev P.M. (2024). Decision-making in the modern economy: artificial intelligence vs. behavioral economics // *Issues of Innovation Economics.* Vol. 14. No. 3. Pp. 649–666. DOI: 10.18334/vinec.14.3.121070 (In Russ.).

Makarov M.Yu. (2020). The Impact of Artificial Intelligence on Productivity // *Economics and Management*. Vol. 26. No. 5. Pp. 479–486. DOI: 10.35854/1998-1627-2020-5-479-486 (In Russ.).

Malle B.F., Magar S.T., Scheutz M. (2019). AI in the sky: How people morally evaluate human and machine decisions in a lethal strike dilemma // *Robotics and Well-Being. Intelligent Systems, Control and Automation: Science and Engineering* / A.M. Ferreira, S.J. Sequeira, S.G. Virk, M. Tokhi, E.E. Kadar (eds). Vol. 95. — Cham: Springer International Publishing. Pp. 111–133. DOI: 10.1007/978-3-030-12524-0_11.

Manakhova I.V., Makovskaya A.M. (2025). Human and artificial intelligence in the discourse of behavioral economics // *Lomonosov Economics Journal.* Vol. 60. No. 3. Pp. 3–19. DOI: 10.55959/MSU0130-0105-6-60-3-1 (In Russ.).

Maninger T., Shank D.B. (2022). Perceptions of violations by artificial and human actors across moral foundations // *Computers in Human Behavior Reports*. Vol. 5. Article 100154. DOI: 10.1016/J.CHBR.2021.100154.

Massachusetts Institute of Technology. (2025). *The GenAI divide: State of Business 2025*. URL: ai_report_2025.pdf (access date: 26.02.2026).

Mayer R.C., Davis J.H., Schoorman F.D. (1995). An integrative model of organizational trust // *Academy of Management Review*. Vol. 20. No. 3. Pp. 709–734. DOI: 10.5465/amr.1995.9508080335.

Mehrotra S., Jonker C.M., Tielman M.L. (2021). *More Similar Values, More Trust? — The Effect of Value Similarity on Trust in Human-Agent Interaction.* Proceedings of the 2021 AAAI/ACM Conference on AI, Ethics, and Society. Association for Computing Machinery. — New York. Pp. 777–783. DOI: 10.1145/3461702.3462576.

Mirowska A., Arsenyan J. (2025). «It Feels Wrong»: Understanding Reactions to Artificial Intelligence as a Decision-Maker in Selection Through the Lens of Moral Foundations Theory // *International Journal of Selection and Assessment*. Vol. 34. Article e70039. DOI: 10.1111/ijsa.70039.

Morselli D., Spini D., Devos T. (2012). Human Values and Trust in Institutions across Countries: A Multilevel Test of Schwartz's Hypothesis of Structural Equivalence // *Survey Research Methods*. Vol. 6. No. 1. Pp. 49–60. DOI: 10.18148/srm/2012.v6i1.5090.

Nagy S., Hajdú N. (2021). Consumer acceptance of the use of artificial intelligence in online shopping: Evidence from Hungary // *Amfiteatru Economic*. Vol. 23. No. 56. Pp. 155–173. DOI:10.24818/EA/2021/56/155.

Nunes J.L. et al. (2024). Are large language models moral hypocrites? A study based on moral foundations // *Proceedings of the AAAI/ACM Conference on AI, Ethics, and Society.* Vol. 7. No. 1. Pp. 1074–1087. DOI: 10.48550/arXiv.2405.11100.

Panari C., Lorenzi G., Mariani M.G. (2021). The Predictive Factors of New Technology Adoption, Workers' Well-Being and Absenteeism: The Case of a Public Maritime Company in Venice // *International Journal of Environmental Research and Public Health*. Vol. 8. No. 23. Article 12358. DOI: 10.3390/ijerph182312358.

Petrunin Y.Y., Nuralieva N.Z. (2025). Trust in Generative Artificial Intelligence as a Mirror of Institutional Trust // *Public Administration:* E-journal. Vol. 113. Pp. 22–30. DOI: 10.55959/MSU2070-1381-113-2025-22-30 (In Russ.).

Purcell Z.A., Bonnefon J.F. (2023). Humans feel too special for machines to score their morals // *PNAS nexus*. Vol. 2. No. 6. Article pgad179. DOI: 10.1093/pnasnexus/pgad179.

Robinette P. et al. (2016). *Overtrust of robots in emergency evacuation scenarios* // 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI). — Christchurch, New Zealand. Pp. 101–108. DOI: 10.1109/HRI.2016.7451740.

Ruelas-Villavicencio A.L., Contreras-Yáñez I., Gómez-Ruiz R.P., Zagaglia Del Valle M.C., Malagón-Liceaga A., Pascual-Ramos V. (2025). Digital health literacy is linked to attitudes regarding the ethical aspects of digital health among patients with dermatologic comorbidities // *PLOS ONE*. Vol. 20. No. 9. DOI: 10.1371/journal.pone.0330916.

Sakamoto Y., Uchida T., Ishiguro H. (2025). Value-based large language model agent simulation for mutual evaluation of trust and interpersonal closeness // *Scientific Reports*. Vol. 15. Article 41653. DOI: 10.1038/s41598-025-25531-1.

Salatino A., Prével A., Caspar E. et al. (2025). Influence of AI behavior on human moral decisions, agency, and responsibility // *Scientific Reports*. Vol. 15. Article 12329. DOI: 10.1038/s41598-025-95587-6.

Schwartz S.H. (1992). Universals in the content and structure of values: Theoretical advances and empirical tests in 20 countries // *Advances in Experimental Social Psychology*. Vol. 25. No. 1. Pp. 1–65. DOI: 10.1016/S0065-2601(08)60281-6.

Schwartz S.H., Cieciuch J., Vecchione M., Davidov E., Fischer R., Beierlein C., Ramos A., Verkasalo M., Lönnqvist J.-E., Demirutku K., Dirilen-Gumus O., Konty M. (2012). Refining the theory of basic individual values // *Journal of Personality and Social Psychology.* Vol. 103. No. 4. Pp. 663–688. DOI: 10.1037/a0029393.

Segerer R. (2025). Cultural value alignment in large language models: A prompt-based analysis of Schwartz values in Gemini, ChatGPT, and DeepSeek // *arXiv preprint.* ArXiv:2505.17112. DOI: 10.48550/arXiv.2505.17112.

Shahzad M.F., Xu S., Lim W.M. et al. (2024). Cryptocurrency awareness, acceptance, and adoption: The role of trust as a cornerstone // *Humanities and Social Sciences Communications*. Vol. 11. Article 4. DOI: 10.1057/s41599-023-02528-7.

Shank D.B., DeSanti A. (2018). Attributions of morality and mind to artificial intelligence after real-world moral violations // *Computers in Human Behavior*. Vol. 86. Pp. 401–411. DOI: 10.1016/J.CHB.2018.05.014.

Shank D.B., Gott A. (2020). Exposed by AIs! People personally witness artificial intelligence exposing personal information and exposing people to undesirable content // *International Journal of Human–Computer Interaction*. Vol. 36. No. 17. Pp. 1636–1645. DOI: 10.1080/10447318.2020.1768674.

Shen H. et al. (2025). *ValueCompass: A framework for measuring contextual value alignment between human and LLMs* / Proceedings of the 9th Widening NLP Workshop. Pp. 75–86. DOI: 10.48550/arXiv.2409.09586.

Stanciu A., Partsch M., Lechner C.M. (2024). Basic human values and the adoption of cryptocurrency // *Frontiers in Psychology*. Vol. 15. Article 1395674. DOI: 10.3389/fpsyg.2024.1395674.

Stein J.P., Messingschlager T., Gnambs T. et al. (2024). Attitudes towards AI: measurement and associations with personality // *Scientific Reports*. Vol. 14. Article 2909. DOI: 10.1038/s41598-024-53335-2.

Sullivan Y., de Bourmont M., Dunaway M. (2022). Appraisals of harms and injustice trigger an eerie feeling that decreases trust in artificial intelligence systems // *Annals of Operations Research*. Vol. 308. Pp. 525–548. DOI: 10.1007/s10479-020-03702-9.

Tang L., Ferronato P., Bashir M. (2023). Do Users' Values Influence Trust in Automation? // *Intelligent Human Computer Interaction. IHCI 2022. Lecture Notes in Computer Science* / H. Zaynidinov, M. Singh, U.S. Tiwary, D. Singh (eds). — Cham: Springer. Vol. 13741. DOI: 10.1007/978-3-031-27199-1_30.

Tatarko A.N. (2017). The relationship of basic human values and voting behavior // *Social Psychology and Society.* Vol. 8. No. 1. Pp. 17–37. DOI: 10.17759/sps.2017080102 (In Russ.).

Ulfert A.S., Antoni C.H., Ellwart T. (2022). The role of agent autonomy in using decision support systems at work // *Computers in Human Behavior*. Vol. 126. Article 106987. DOI:10.1016/j.chb.2021.106987.

Vinokurov F.N., Sadovskaya E.D. (2023). Whom We Trust More: AI-driven vs. Human-driven Economic Decision-Making // *Experimental Psychology.* Vol. 16. No. 2. Pp. 87-100. DOI: 10.17759/exppsy.2023160206 (In Russ.).

Vuori N., Burkhard B., Pitkäranta L. (2025). It's Amazing–But Terrifying!: Unveiling the Combined Effect of Emotional and Cognitive Trust on Organizational Member'Behaviours, AI Performance, and Adoption // *Journal of Management Studies*. Vol. 63. No. 2. Pp. 473–514. DOI: 10.1111/joms.13177.

Wen Y., Wang J., Chen X. (2025). Trust and AI weight: Human-AI collaboration in organizational management decision-making // *Frontiers in Organizational Psychology*. Vol. 3. Article 1419403. DOI: 10.3389/forgp.2025.1419403.

Witte E.H., Stanciu A., Boehnke K. (2020). A New Empirical Approach to Intercultural Comparisons of Value Preferences Based on Schwartz's Theory // *Frontiers in Psychology*. Vol. 11. Article 1723. DOI: 10.3389/fpsyg.2020.01723.

Xiong Y. et al. (2023). More trust or more risk? User acceptance of artificial intelligence virtual assistant // *Human Factors and Ergonomics In Manufacturing*. Vol. 34. No. 3. Pp. 190–205. DOI: 10.1002/hfm.21020.

You Y., Hu Y., Yang W., Cao S. (2022). Research on the Influence Path of Online Consumers' Purchase Decision Based on Commitment and Trust Theory // *Frontiers in Psychology*. Vol. 13. Article 916465. DOI: 10.3389/fpsyg.2022.916465.

Zhang Q., Lee M.L., Carter S. (2022). You complete me: Human-ai teams and complementary expertise // *CHI Conference on Human Factors in Computing Systems*. Article 114. Pp. 1–28. DOI: 10.1145/3491102.3517791.

Zhang Q., Wang F., Liao G., Li M. (2025). How Does AI Trust Foster Innovative Performance Under Paternalistic Leadership? The Roles of AI Crafting and Leader's AI Opportunity Perception // *Behavioral Sciences*. Vol. 15. No. 8. Article 1064. DOI: 10.3390/bs15081064.

Самойлов Олег Михайлович
ORCID: 0009-0005-1509-7225
osamoilov1@gmail.com

Oleg Samoilov
PhD student, Research assistant, Center for Socio-Cultural Studies, National Research University Higher School of Economics (Moscow)
ORCID: 0009-0005-1509-7225
osamoilov1@gmail.com

Татарко Александр Николаевич
ORCID: 0000-0001-7557-9107
tatatrko@yandex.ru

Alexander Tatarko
ORCID: 0000-0001-7557-9107
Doctor of Psychology, Professor, Department of Psychology, Faculty of Social Sciences, Chief Researcher, Center for Socio-Cultural Studies, National Research University Higher School of Economics (Moscow)
tatatrko@yandex.ru

SOCIO-PSYCHOLOGICAL FACTORS OF TRUST IN ARTIFICIAL INTELLIGENCE: STATE OF RESEARCH

Abstract. The article presents a theoretical review of literature from the last ten years devoted to the analysis of socio-psychological factors of trust in artificial intelligence. The widespread adoption of automated artificial intelligence systems, which is associated with expected economic growth, reduced resource costs, and the optimization of various work processes, often faces user distrust in new tools and a reluctance to transform traditional work processes. The combination of factors that reduce trust in artificial intelligence leads to low economic efficiency in the implementation of innovations, despite the wide range of technical possibilities. In addition to the importance of considering cognitive and affective factors of trust in artificial intelligence, socio-psychological aspects that determine the ethical and value acceptability of using automated AI systems are of particular significance. Theoretical analysis revealed that individualizing moral foundations are positively associated with trust in artificial intelligence in situations where the use of AI systems benefits society. Users who prioritize binding moral foundations demonstrate greater distrust of artificial intelligence and are less willing to delegate some tasks to automated assistants. The values of Openness to Change and Self-transcendence are more positively associated with trust in AI and digital innovation, excluding situations of high risk to life or social injustice. The values of Self-Enhancement are also positively associated with trust in AI tools, but mainly in situations where artificial intelligence simplifies the achievement of the user's goals or expands human capabilities to do so. There is a heterogeneous structure of relationships between Conservation values and trust in artificial intelligence, due to cultural characteristics. However, it should be noted that Conservation values are most often considered predictors of distrust in automated AI systems. The importance of considering value congruence between users and perceived AI profiles is discussed. For AI system developers, there is a need to pay special attention to the possibilities of adaptive personalized adjustment of the value profiles of generative models to users, which will lead to more effective human-machine interaction. The article highlights the areas of future research in this field as part of the development of a systemic model of trust in artificial intelligence.

Keywords: *trust, artificial intelligence, socio-psychological factors, values, moral foundations.*
JEL: O33, D83, M15, A13, Z13.