

ПОСТРОЕНИЕ ДОВЕРИЯ ПРИ ИСПОЛЬЗОВАНИИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Получено: 20.04.2021 Доработано после рецензирования: 14.06.2021 Принято: 23.06.2021

УДК 004.8 JEL C80, D80, O30 DOI <https://doi.org/10.26425/2658-3445-2021-4-2-28-36>

Дашков Андрей Александрович

Канд. техн. наук, доц., ФГБОУ ВО «Государственный университет управления», г. Москва, Российская Федерация

ORCID: <https://orcid.org/0000-0002-9850-5319>

e-mail: aa_dashkov@guu.ru

Нестерова Юлия Олеговна

Студент магистратуры, ФГБОУ ВО «Государственный университет управления», г. Москва, Российская Федерация

ORCID: <https://orcid.org/0000-0001-8454-3704>

e-mail: iunester@yandex.ru

АННОТАЦИЯ

В XXI в. доверие становится категорией, разнообразно проявляющейся и влияющей на многие сферы деятельности человека, в том числе на экономику и бизнес. С развитием информационно-коммуникационных и сквозных технологий это влияние становится все более заметным. Особое место в цифровых технологиях занимает доверие человека при взаимодействии с системами искусственного интеллекта и машинного обучения. В этом случае доверие становится потенциальным камнем преткновения в сфере дальнейшего развития взаимодействия искусственного интеллекта и человека. Доверие играет ключевую роль в обеспечении признания в обществе, постоянного прогресса и развития искусственного интеллекта.

В статье с разных сторон рассмотрено доверие человека к системам искусственного интеллекта и машинного обучения. Проведена работа по структурированию существующих исследований по указанной тематике и выявлены наиболее важные способы создания доверия у потенциальных потребителей продуктов искусственного интеллекта. Исследовано отношение к искусственному интеллекту в разных странах, а также необходимость доверия у пользователей систем искусственного интеллекта, проанализировано влияние недоверия на бизнес. Выявлены факторы, имеющие решающее значение в формировании первоначального уровня доверия и развития непрерывного доверия к искусственному интеллекту.

КЛЮЧЕВЫЕ СЛОВА

Доверие, искусственный интеллект, доверие к искусственному интеллекту, факторы доверия, отношение к искусственному интеллекту, потребители, бизнес, этика

ДЛЯ ЦИТИРОВАНИЯ

Дашков А.А., Нестерова Ю.О. Построение доверия при использовании искусственного интеллекта // E-Management. 2021. Т. 4, № 2. С. 28–36.



ARTIFICIAL INTELLIGENCE TECHNOLOGIES IN MANAGEMENT

BUILDING TRUST WHEN USING ARTIFICIAL INTELLIGENCE

Received: 20.04.2021 Revised: 14.06.2021 Accepted: 23.06.2021

JEL C80, D80, O30

Andrey A. Dashkov

Cand. Sci. (Tech.), Assoc. Prof., State University of Management, Moscow, Russia

ORCID: <https://orcid.org/0000-0002-9850-5319>

e-mail: aa_dashkov@guu.ru

Yulia O. Nesterova

Graduate Student, State University of Management, Moscow, Russia

ORCID: <https://orcid.org/0000-0001-8454-3704>

e-mail: iunester@yandex.ru

ABSTRACT

In the XXI century, “trust” becomes a category that manifests itself in a variety of ways and affects many areas of human activity, including the economy and business. With the development of information and communication technologies and end-to-end technologies, this influence is becoming more and more noticeable. A special place in digital technologies is occupied by human trust when interacting with artificial intelligence and machine learning systems. In this case, trust becomes a potential stumbling block in the field of further development of interaction between artificial intelligence and humans. Trust plays a key role in ensuring recognition in society, continuous progress and development of artificial intelligence.

The article considers human trust in artificial intelligence and machine learning systems from different sides. The main objectives of the research paper are to structure existing research on this subject and identify the most important ways to create trust among potential consumers of artificial intelligence products. The article investigates the attitude to artificial intelligence in different countries, as well as the need for trust among users of artificial intelligence systems and analyses the impact of distrust on business. The authors identified the factors that are crucial in the formation of the initial level of trust and the development of continuous trust in artificial intelligence.

KEYWORDS

Trust, artificial intelligence, trust in artificial intelligence, trust factors, attitude to artificial intelligence, consumers, business, ethics

FOR CITATION:

Dashkov A.A., Nesterova Y.O. (2021) Building trust when using artificial intelligence. *E-Management*, vol. 4, no. 2, pp. 28–36. DOI <https://doi.org/10.26425/2658-3445-2021-4-2-28-36>



ВВЕДЕНИЕ

Проблема надежности и степени доверия человечества к инструментам искусственного интеллекта (далее – ИИ) сейчас стоит особо остро. Связано это с тем, что ИИ начал активно распространяться и проникать в различные сферы деятельности человека. Государство поддерживает и финансирует исследования и разработки в этой области [Потапова, Шклярчук, 2021]. Внедрение интеллектуальных систем в программные продукты, роботов, медицинское оборудование, транспорт становится все более распространенным. В связи с этим резонны вопросы: можем ли мы доверять ИИ; насколько решения, принятые ИИ, оправданы; кто несет ответственность за допущенные системой ИИ ошибки, которые могут повлечь не только финансовые, но и человеческие потери.

ПРОБЛЕМА ДОВЕРИЯ СИСТЕМАМ ИИ В РАЗНЫХ СТРАНАХ

В 2020 г. с 1 по 3 июля в Сколковском институте науки и технологий («Сколтех») при участии экспертов образованного на базе РВК Технического комитета по стандартизации ТК 164 «Искусственный интеллект» прошла международная онлайн-конференция, посвященная проблемам устойчивости, прозрачности и надежности ИИ¹. Ведущие мировые эксперты в области ИИ впервые объединились в таком формате для выработки единой позиции по вопросу доверия к результатам применения технологий ИИ. Данная конференция по вопросам доверия и надежности ИИ состоится и в 2021 г., что подтверждает актуальность поднимаемой в статье темы².

Проблема доверия к ИИ, кроме ученых, важна и бизнесу, который уже использует или планирует использовать подобные системы. Устранение слабых мест, связанных с доверием, поможет улучшить бизнес-процессы [Glikson and Woolley, 2020]. Высокий уровень доверия – это конкурентное преимущество, так как когда одной организации доверяют больше, чем конкурентам, то бизнес этой компании будет расти быстрее³.

Уровень доверия людей к ИИ исследуется во многих странах по всему миру. Например, Trust in Artificial Intelligence [Gillespie et al, 2021] – научное исследование, которое является первым глубоким погружением в понимание доверия граждан и ожиданий от использования ИИ во многих странах. Для этого исследования была опрошена репрезентативная выборка более чем из 1 200 граждан США, Канады, Германии, Великобритании и Австралии (общая выборка 6 054). Авторы [Gillespie et al, 2021] выявили во всех пяти странах четыре ключевых фактора, влияющих на доверие граждан к системам ИИ:

- убеждение в адекватности действующих правил и законов для обеспечения безопасности использования ИИ;
- влияние ИИ на рабочие места;
- понимание ИИ;
- воспринимаемое влияние ИИ на общество.

Из этих факторов первый, связанный с адекватностью действующих правил и законов, безусловно, является самым сильным.

Упомянутое выше исследование показало, что доверие к ИИ в пяти странах низкое, и граждане, как правило, настороженно или неоднозначно относятся к этим системам. Только около четверти (28 %) граждан готовы доверять системам ИИ в целом³. Двое из пяти граждан не желают делиться своей информацией или данными с системой ИИ, а третий не желает доверять выводам систем ИИ (например, рекомендациям или решениям).

Авторы [Gillespie et al, 2021] считают, что от правительства и бизнеса требуются следующие меры для укрепления общественного доверия к ИИ:

- оправдание ожиданий граждан относительно надежного ИИ;

¹ *Новости РВК* (2020). Российские эксперты разрабатывают критерии доверия к искусственному интеллекту // РВК, 15 июля. Режим доступа: <https://www.rvc.ru/press-service/news/company/156669/> (дата обращения: 19.04.2021).

² *Skoltech* (2021). Trustworthy AI transparency robustness sustainability // Trustworthy AI conference, July 5–7. Режим доступа: <https://events.skoltech.ru/ai-trustworthy> (дата обращения: 19.04.2021).

³ *Капитал* (2020). Как раскрыть потенциал данных и доверия // Капитал, 27 ноября. Режим доступа: <https://kapital.kz/tehnology/91473/kak-raskryt-potentsial-dannykh-i-doveriya.html> (дата обращения: 19.04.2021).

- укрепление нормативной базы для управления ИИ;
- повышение грамотности населения в области ИИ.

ЭТИЧЕСКИЕ ПРИНЦИПЫ ФУНКЦИОНИРОВАНИЯ СИСТЕМ ИИ В СТРАНАХ ЕВРОПЫ

Европейская группа экспертов высокого уровня по ИИ 8 апреля 2019 г. представила руководящие принципы этики для надежного ИИ⁴. Они считают, что следует устанавливать доверительные отношения с ИИ и культивировать заслуживающий доверия ИИ. Согласно этому руководству, заслуживающий доверия ИИ обладать следующими характеристиками:

- законный (с соблюдением всех применимых законов и постановлений);
- этичный (с соблюдением этических принципов и ценностей);
- надежный (как с технической точки зрения, так и с учетом социальной среды).

Кроме того, в руководстве обозначены 7 ключевых требований, которым системы ИИ должны соответствовать, чтобы считаться заслуживающими доверия:

- системы ИИ должны расширять возможности людей, позволяя им принимать осознанные решения и защищая их основные права;
- техническая надежность и безопасность: системы ИИ должны быть устойчивыми и безопасными;
- конфиденциальность и управление данными;
- прозрачность: бизнес-модели данных, системы и ИИ должны быть прозрачными;
- разнообразие, отсутствие дискриминации и справедливость;
- благополучие общества и окружающей среды;
- подотчетность: необходимо создать механизмы, обеспечивающие ответственность и подотчетность за системы ИИ и их результаты.

МЕРЫ ПО ОБЕСПЕЧЕНИЮ ДОВЕРИЯ К СИСТЕМАМ ИИ В РОССИИ

В России также проводятся исследования об отношении граждан к технологиям ИИ. Проектный офис по реализации национальной программы «Цифровая экономика» Аналитического центра при Правительстве Российской Федерации и Всероссийский центр изучения общественного мнения 27 января 2020 г. представили результаты опроса россиян⁵.

Около половины россиян относятся к распространению технологий ИИ положительно. Статистика показывает, что 48 % населения выразили интерес, восхищение, воодушевление, доверие технологии, 31 % – нейтральное отношение, 7 % – удивление и иные эмоции, а 12 % высказали негативные эмоции⁴. В целом в России граждане готовы к применению систем ИИ и выражают к ним доверие. Однако процент негативного отношения может начать увеличиваться, а потому дальнейшее информирование граждан в целях повышения доверия должно продолжаться.

С 1 марта 2021 г. в России начал действовать новый стандарт ГОСТ Р 59276-2020 «Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения»⁶. В введении к стандарту отмечено, что доверие к системам ИИ является важнейшим условием, определяющим возможность их применения при решении ответственных задач обработки данных.

В ГОСТ Р 59276-2020 описаны основные способы обеспечения доверия к системам ИИ на трех уровнях архитектуры⁶:

- физическом: уровне сенсоров и исполнительных устройств;
- инфраструктурном: уровень, включающий аппаратные средства хранения, обработки и передачи информации, включая облачную инфраструктуру, а также системное программное обеспечение;

⁴ High-level expert group on artificial intelligence set up by the European Commission (2019). Ethics guidelines for trustworthy AI // Europa. Режим доступа: <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> (дата обращения: 15.04.2021).

⁵ Искусственный интеллект: угроза или возможность? // Всероссийский центр изучения общественного мнения. Режим доступа: <https://wciom.ru/analytical-reviews/analiticheskii-obzor/iskusstvennyj-intellekt-ugroza-ili-vozmozhnost> (дата обращения: 19.04.2021).

⁶ ГОСТ Р 59276-2020 «Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения» от 03.01.2021. Утвержден и введен в действие Приказом Федерального агентства по техническому регулированию и метрологии от 23 декабря 2020 г. № 1371-ст. Режим доступа: <https://docs.cntd.ru/document/1200177291> (дата обращения: 19.04.2021).

– прикладном: уровень прикладного программного обеспечения, реализующего алгоритмы интеллектуальной обработки данных.

В государственном стандарте также сказано, что доверие к системе ИИ достигается в том случае, если используются способы обеспечения доверия на всех стадиях жизненного цикла системы ИИ5.

ФАКТОРЫ, ВЛИЮЩИЕ НА ДОВЕРИЕ К ИИ

Профессор К. Сиау, заведующий кафедрой бизнеса и информационных технологий Университета науки и технологий штата Миссури, считает, что существует два типа доверия к ИИ: первоначальное и постоянное [Siau, Wang, 2018]. Он вывел факторы, влияющие на доверие (таблица).

Таблица. Особенности, влияющие на формирование доверия к искусственному интеллекту

Table. Features affecting the trust formation to artificial intelligence

Первоначальное формирование доверия	Постоянное развитие доверия
Производительность: – представительство (представление) – образ/восприятие – отзывы других пользователей	Производительность: – удобство использования и надежность – сотрудничество и коммуникация – безопасность и защита конфиденциальности – интерпретируемость
Процесс: – прозрачность – возможность судебного разбирательства	Цель: – замена «человеческих» рабочих мест – соответствие целям

Источник: [Siau & Wang, 2018, с. 51]

Рассмотрим подробно факторы, приведенные в таблице.

1. Представление.

Представление играет важную роль в первоначальном построении доверия. Автор считает, что чем больше робот похож на человека, тем легче людям установить с ним эмоциональную связь. Собака-робот – еще один пример представления ИИ, которому людям легче доверять.

2. Образ/восприятие.

Научно-фантастические книги и фильмы создали для систем ИИ плохой образ: люди могут предполагать, что ИИ выходит из-под контроля и может нам навредить. Соответственно, необходимо альтернативное освещение ИИ в СМИ, кино, книгах. Людям следует понять, что ИИ не враг, а помощник.

3. Отзывы других пользователей.

На сегодняшний день распространение отзывов происходит очень быстро благодаря сети «Интернет». Положительные отзывы сильно способствуют формированию доверия со стороны потенциальных потребителей.

4. Прозрачность.

Чтобы доверять приложениям ИИ, людям нужно понять, как они запрограммированы и какую функцию будут выполнять в определенных условиях. Эта прозрачность важна, и ИИ должен уметь объяснять/оправдывать свое поведение и решения. Одной из проблем машинного обучения и глубокого обучения является «черный ящик» в процессах принятия решений. Если объяснимость приложения ИИ плохая или отсутствует, это отрицательно влияет на доверие.

5. Возможность испытания.

Тестируемость означает возможность для людей получить доступ к приложению ИИ и попробовать его, прежде чем начать использовать. Тестирование позволяет улучшить понимание. Таким образом, предоставление потенциальным пользователям возможности опробовать новую технологию будет способствовать более высокому уровню первоначального доверия.

6. Удобство использования и надежность.

Производительность включает в себя компетентность ИИ в выполнении задач и выполнение этих задач последовательным и надежным образом. Приложение ИИ должно быть разработано так, чтобы работать с ним было легко и интуитивно. Не должно быть никаких неожиданных простоев или сбоев. Удобство использования и надежность способствуют постоянному доверию.

7. *Сотрудничество и коммуникация.*

Хотя большинство приложений ИИ разрабатывают для выполнения задач самостоятельно, наиболее вероятным сценарием в краткосрочной перспективе является то, что люди будут работать в партнерстве с машинами. От того, насколько легко и гладко будет осуществляться сотрудничество и коммуникация, зависит постоянное доверие.

8. *Общительность и связь.*

Так как люди – социальные существа, постоянное доверие может быть усилено общественной деятельностью. С собакой-роботом, которая может узнать своего хозяина и проявить привязанность, можно обращаться как с домашней собакой, устанавливая эмоциональную связь и доверие.

9. *Безопасность и защита частной жизни.*

Оперативная безопасность и безопасность данных – два важных фактора, влияющих на доверие к технологиям. Люди вряд ли будут доверять чему-то, что слишком рискованно для работы. Безопасность данных важна, потому что машинное обучение опирается на большие наборы данных, что делает конфиденциальность проблемой.

10. *Интерпретируемость.*

Большинство моделей машинного обучения с «черным ящиком» непостижимы. Для решения этой проблемы необходимо разработать интерпретируемые модели и обеспечить возможность для машины объяснять свои выводы или действия. Это может помочь пользователям понять обоснование результатов и процесс их получения.

11. *Замена человеческих рабочих мест.*

Искусственный интеллект может превзойти человеческую производительность во многих аспектах и заменить человеческих работников. Он будет продолжать расширять свои возможности и проникать в новые области. Беспокойство о том, что ИИ займет рабочие места и заменит людей, будет препятствовать доверию людей к ИИ. Например, те, чьи рабочие места могут быть заменены ИИ, могут не захотеть доверять ему. Низкоквалифицированные, повторяющиеся и опасные рабочие места относят к числу тех, которые, скорее всего, будут заняты машинами. Предоставление переподготовки и образования пострадавшим сотрудникам поможет смягчить это влияние на непрерывное доверие.

12. *Соответствие целям.*

Поскольку ИИ обладает потенциалом продемонстрировать и даже превзойти человеческий интеллект, понятно, что люди относятся к нему как к угрозе. Цели ИИ должны соответствовать заложенным в него задачам и не выходить за эти пределы, тогда доверие к ИИ будет стабильным.

КРИТИКА ПРАВОМЕРНОСТИ ПРИМЕНЕНИЯ ПОНЯТИЯ ДОВЕРИЯ ПО ОТНОШЕНИЮ К СИСТЕМАМ ИИ

Однако существует и альтернативное мнение о том, что системам ИИ не стоит доверять. Доктор Дж. Брайсон, доцент кафедры вычислительной техники Батского университета, считает, что никто не должен доверять ИИ [Bryson, 2018]. В своей статье она пишет, что доверие – это отношения между людьми, в которых доверяющая сторона, не зная наверняка, что будет делать доверенная сторона, верит любым данным обещаниям. Искусственный интеллект – набор методов разработки систем, которые позволяют машинам выполнять действия или получать знания из набора данных. Только другие методы разработки программного обеспечения могут быть равными ИИ, и поскольку они не умеют «доверять», то никто не может доверять ИИ. Дж. Брайсон считает, что людям и не нужно доверять системе ИИ, вместо этого необходимо знать, насколько вероятно, что такая система выполнит поставленную задачу и только эту задачу. Когда система, использующая ИИ, наносит ущерб, люди должны знать, что к ответственности необходимо привлекать людей, стоящих за этой системой, а не систему.

Подобного мнения придерживается и М. Райан [Ryan, 2020], сотрудник Королевского технологического института. По мнению этого автора, доверие как термин нельзя применять к ИИ. Доверие предполагает, что один может полагаться на другого, так как он действует по доброй воле по отношению к доверителю. Это главная причина, почему созданные человеком объекты, такие как ИИ, могут быть надежными, но не заслуживают доверия. В отношении доверия каждый должен нести ответственность за свои действия, чего ИИ не может

сделать, так как это набор методов и алгоритмов. Поэтому М. Райан предлагает сосредоточиться на надежности ИИ, а не степени доверия к нему, обращая внимание на то, что надежный ИИ возлагает бремя ответственности на тех, кто разработал, внедрил и использует эти технологии, а не перекладывает ее на систему.

Эти авторы правы в том, что важно, чтобы системы ИИ никогда не представлялись как ответственные или юридические лица сами по себе. Однако недоверие к системам ИИ повлечет за собой препятствие дальнейшему их развитию и использованию [Ferrario et al, 2020], а потому доверие необходимо развивать, не смотря на то, что ИИ является лишь технологией, а не человеком.

ВЫВОДЫ

Исходя из анализа существующих исследований можно выделить основные способы обеспечения доверия со стороны потенциальных пользователей систем ИИ. Доверие может обеспечить более высокую конкурентоспособность на рынке, поэтому для бизнеса важно обратить внимание на следующие факторы.

1. Осязаемость.

Существующие исследования подтверждают представление о том, что визуальное присутствие человекоподобных или животных повышает начальное доверие по сравнению с отсутствием визуализации.

2. Прозрачность.

Один из способов умерить нереально высокие ожидания от пользователей – обеспечить объяснение от носительно функциональности ИИ. Если пользователь понимает, как система ИИ пришла к подобным выводам и что она анализировала, то он может быть в ней уверен, а следовательно, и доверять.

3. Доверие к виртуальному ИИ.

Виртуальный агент с поддержкой ИИ присутствует физически (то есть нет первого пункта «осязаемость»), однако чат-бот или виртуальный помощник могут иметь собственное виртуальное представление. Искусственный интеллект может существовать на любом электронном устройстве и обладать такими функциями, как аватар, лицо, тело, голос, анимированные движения или даже проекция в реальном мире. Все это может увеличить степень доверия пользователей.

4. Надежность.

Надежность играет важную роль в доверии пользователей. Исследования, которые проверяли уровни доверия, обусловленные точностью и неудачами ИИ, выявили стабильный паттерн, указывающий на то, что ошибки встроенного ИИ наносят ущерб доверию. Ошибка влияет на доверие таким образом, что его трудно исправить, а потому следует тщательнее тестировать системы ИИ, формировать представительную обучающую выборку, а также разработать шкалу уверенности ИИ в предлагаемом решении, чтобы в случае низкой уверенности ИИ в правильности решения вопрос переходил в человеческие руки.

5. Репутация.

Эмоциональное доверие к встроенному ИИ может быть построено на основе репутации технологии и репутации организаций, связанных с ним. Если потребитель доверяет организации, стоящей за разработкой ИИ, то он имеет больший первоначальный запас доверия к технологии.

6. Соответствие заданным задачам.

Искусственный интеллект не должен выходить за рамки требуемых от него задач. Пользователь не хочет, а потому и не должен получать информацию сверх требуемой, так как это может уменьшить его доверие к системе. Соответственно, при разработке необходимо накладывать корректные ограничения на допустимую область применения системы ИИ.

7. Отзывы.

Следует мотивировать людей, которые довольны продуктом, писать положительные отзывы, и оперативно реагировать на каждый негативный отзыв. Таким образом, большое количество положительных отзывов будет формировать начальный уровень доверия у потенциальных потребителей.

8. Безопасность и конфиденциальность.

Пользователи должны знать, что их данные в полной безопасности, а потому бизнесу необходимо разрабатывать эффективные меры по защите персональных данных. Важным является обеспечение безопасности системы от взлома. Уменьшая степень риска пользователей при использовании системы ИИ, бизнес увеличивает их доверие.

Государству необходимо обратить внимание на следующие факторы для увеличения доверия к ИИ:

1. Укрепление нормативной базы.

Граждане должны понимать, что они и их права при использовании ИИ находятся под защитой государства. В случае происшествия им необходимо точно знать, куда можно обратиться и кто должен быть ответственным за решения, принятые системой ИИ.

2. Повышение грамотности населения в области ИИ.

Как показало исследование, только треть опрошенных россиян понимают сущность ИИ и могут дать ему определение. Неопределенность порождает недоверие, поэтому необходимо создавать программы повышения осведомленности граждан о целях и сферах применения ИИ.

3. Образ ИИ в глазах граждан.

Для увеличения первоначального доверия граждан к системам ИИ необходимо постепенно изменять укоренившийся образ «злого искусственного интеллекта». Для этого необходимо показывать ИИ в кино, литературе и средствах массовой информации как технологию, подконтрольную человеку, а не «роботов-захватчиков».

ЗАКЛЮЧЕНИЕ

В статье представлены некоторые существующие исследования о доверии к технологии искусственного интеллекта, исследовано отношение к искусственному интеллекту в разных странах, а также выявлены наиболее важные способы создания доверия у граждан.

Учитывая быстрое инвестирование и внедрение искусственного интеллекта, будет важно регулярно пересматривать общественное доверие и ожидания в отношении систем искусственного интеллекта по мере их эволюции с течением времени, чтобы обеспечить соответствие искусственного интеллекта ожиданиям общества.

Настоящая работа позволяет улучшить понимание доверия в области искусственного интеллекта, а также представляет основные рекомендации бизнесу для увеличения конкурентоспособности в результате первоначального и постоянного поддержания доверия потребителей.

СПИСОК ЛИТЕРАТУРЫ

- Потапова Е.Г., Шклярчук М.С. [ред.] (2021). *Этика и «цифра»: от проблем к решениям*. М.: РАНХиГС. 184 с.
- Ferrario A., Loi M., Vigano E. (2020). In AI we trust incrementally: A multi-layer model of trust to analyze human-artificial intelligence interactions // *Philosophy & Technology*. V. 33, I. 3. Pp. 523–539. <https://doi.org/10.1007/s13347-019-00378-3>
- Gillespie N., Lockey S., Curtis C. (2021). *Trust in artificial intelligence. A five country study*. The University of Queensland and KPMG Australia. 64 p. <https://doi.org/10.14264/e34bfa3>
- Glikson E., Woolley A. (2020). Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals* // *Academy of Management Annals*. V. 14, No. 2. Pp. 627–660. <https://doi.org/10.5465/annals.2018.0057>
- Ryan M. (2020). In AI we trust: Ethics, artificial intelligence, and reliability // *Science and Engineering Ethics*. V. 26. I. 5. Pp. 2749–2767. <https://doi.org/10.1007/s11948-020-00228-y>
- Siau K., Wang W. (2018). Building trust in artificial intelligence, machine learning, and robotics // *Cutter Business Technology Journal*. V. 31, No. 2. Pp. 47–53.

REFERENCES

- Ferrario A., Loi M. and Vigano E. (2020), “In AI we trust incrementally: A multi-layer model of trust to analyze human-artificial intelligence interactions”, *Philosophy & Technology*, vol. 33, issue 3, pp. 523–539. <https://doi.org/10.1007/s13347-019-00378-3>
- Gillespie N., Lockey S. and Curtis C. (2021), *Trust in artificial intelligence: A five country study*, The University of Queensland and KPMG, Australia. <https://doi.org/10.14264/e34bfa3>
- Glikson E. and Woolley A. (2020), “Human trust in artificial intelligence: Review of empirical research. *Academy of Management Annals*”, *Academy of Management Annals*, vol. 14, no. 2, pp. 627–660. <https://doi.org/10.5465/annals.2018.0057>
- Potapova E.G. and Shklyaruk M.S. [eds]. (2021), *Ethics and “digit”: from problems to solutions*, The Russian Presidential Academy of National Economy and Public Administration Publishing House, Moscow, Russia. (In Russian).

Ryan M. (2020), “In AI we trust: Ethics, artificial intelligence, and reliability”, *Science and Engineering Ethics*, vol. 26, issue 5, pp. 2749–2767. <https://doi.org/10.1007/s11948-020-00228-y>

Siau K. and Wang W. (2018), “Building trust in artificial intelligence, machine learning, and robotics”, *Cutter Business Technology Journal*, vol. 31, no. 2, pp. 47–53.

TRANSLATION OF FRONT REFERENCES

¹ RVC (2020), “Russian experts are developing criteria for trust in artificial intelligence”, *RVC News*, 15 July. Available at: <https://www.rvc.ru/press-service/news/company/156669/> (accessed 19.04.2021).

² Skoltech (2021), “Trustworthy AI transparency robustness sustainability”, *Trustworthy AI conference*, July 5–7. Available at: <https://events.skoltech.ru/ai-trustworthy> (accessed 19.04.2021).

³ Capital (2020), “How to unlock the potential of data and trust”, Capital, 27 November. Available at: <https://kapital.kz/technology/91473/kak-raskryt-potentsial-dannykh-i-doveriya.html> (accessed 19.04.2021).

⁴ High-level expert group on artificial intelligence set up by the European Commission (2019), Ethics guidelines for trustworthy AI, *Europa*. Available at: <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai> (accessed 15.04.2021).

⁵ “Artificial intelligence: threat or opportunity?”, *Russian Public Opinion Research Center (VCIOM)*. Available at: <https://wciom.ru/analytical-reviews/analiticheskii-obzor/iskusstvennyj-intellekt-ugroza-ili-vozmozhnost> (accessed 19.04.2021).

⁶ GOST R 59276-2020 “Artificial intelligence systems. Methods for ensuring trust. General” dated on January 3, 2021. Approved and put into effect by the Order of the Federal Agency on Technical Regulating and Metrology No. 1371-st dated on December 23, 2020. Available at: <https://docs.cntd.ru/document/1200177291> (accessed 19.04.2021).