

## ЭКОНОМИЧЕСКИЕ И СОЦИАЛЬНЫЕ ПЕРЕМЕНЫ

Тема номера:

**Погружение в нейросети:  
цифровая повседневность  
и практики взаимодействия человека  
и генеративного искусственного интеллекта**

№ 5 (183)

сентябрь — октябрь 2024

**СОЦИАЛЬНАЯ  
ДИАГНОСТИКА**



**ЦИФРОВИЗАЦИЯ**



**РЕЦЕНЗИЯ  
НА КНИГУ**



**СОЦИОЛОГИЯ  
ПОВСЕДНЕВНОСТИ**

18+

ISSN 2219-5467



9 772219 546006 >

**Главный редактор журнала:**

Федоров Валерий Валерьевич —  
кандидат политических наук, генеральный директор ВЦИОМ,  
профессор НИУ ВШЭ

**Заместители главного редактора:**

Седова Наталья Николаевна —  
помощник гендиректора по науке ВЦИОМ

Подвойский Денис Глебович —  
кандидат философских наук, ведущий научный сотрудник  
Института социологии ФНИСЦ РАН, доцент РУДН

**Ответственный редактор:**

Бирюкова Светлана Сергеевна —  
кандидат экономических наук, главный научный сотрудник  
Института социальной политики НИУ ВШЭ

M77      Мониторинг общественного мнения: экономические и социальные перемены. — М.: АО «ВЦИОМ», 2024. — № 5 (183). — 236 с.

ISSN 2219-5467

Объективная, точная, регулярная и свежая информация «Мониторинга» полезна всем, кто принимает управленческие решения, занимается прогнозированием и анализом развития общества. Наш журнал пригодится сотрудникам научных и аналитических центров, работникам органов управления, ученым, преподавателям, молодым исследователям, студентам и аспирантам, журналистам.

Тематика материалов охватывает широкий круг социальных, экономических, политических вопросов, основные рубрики посвящены теории, методам и методологии социологических исследований, вопросам взаимодействия государства и общества, социальной диагностике. Каждый номер журнала содержит двухмесячный дайджест основных результатов еженедельных общероссийских опросов ВЦИОМ.

Мы публикуем статьи специалистов, представляющих ведущие научные социологические центры, институты, организации, а также ВУЗы России и зарубежных стран. Широкая тематика журнала представляет возможность выступить на его страницах представителям смежных специальностей (политологам, историкам, экономистам и т.д.), опирающимся в своих исследованиях на эмпирические социологические данные.

Журнал издается с 1992 г.

# СОДЕРЖАНИЕ

---

## ВСТУПИТЕЛЬНОЕ СЛОВО ПРИГЛАШЕННОГО РЕДАКТОРА

---

*С. А. Шомова, А. Г. Качкаева*

Между очарованием и испугом: диалог с «другим». Опыт анализа практик использования ИИ в профессиональной и повседневной жизни ..... 3

## СОЦИАЛЬНАЯ ДИАГНОСТИКА

---

*Е. А. Коваль, С. Г. Ушкин*

Кому нужна этика больших данных: разработчики и их руководители о необходимости создания профессионального этического кодекса .....18

*А. А. Сычев*

В поисках надежности: трансформация доверия в эпоху цифровых технологий ..... 37

*Н. А. Кушкин, О. А. Тимофеева*

Соавтор, помощник, муза, инструмент: как студенты, преподаватели и представители креативных индустрий видят роль ИИ в своих практиках .....60

## МОНИТОРИНГ МНЕНИЙ

---

Мониторинг мнений: сентябрь — октябрь 2024 .....85

## ЦИФРОВИЗАЦИЯ

---

*А. П. Казун*

Может ли искусственный интеллект прогнозировать решения суда? Систематический обзор международных исследований..... 100

*И. С. Душакова, Н. С. Душакова*

Генеративный искусственный интеллект в российских православных сообществах: восприятие и практики использования ..... 123

## СОЦИОЛОГИЯ ПОВСЕДНЕВНОСТИ

---

*Н. Д. Трегубова, А. Я. Фейгина*

«Вежливое» взаимодействие с умными колонками в повседневной жизни: универсальные нормы и новые формы социальности ..... 142

*О. А. Бурова, Ю. С. Неверова, Г. И. Селеткова, Е. В. Середкина*

Взаимодействие пользователей разного возраста с сервисным роботом: лабораторное исследование..... 166

*Е. Г. Ним*

Поп-культура, фэндомы и нейросети: фанаты встречаются с ИИ ..... 183

## **РЕЦЕНЗИЯ НА КНИГУ**

---

*И. И. Буханский*

Искусственный интеллект — это твое.

Рец. на кн.: Ховард Дж., Гуггер С. Глубокое обучение с fastai и PyTorch: минимум формул, минимум кода, максимум эффективности.

СПб. : Питер, 2022 ..... 203

## **НЕФОРМАТ**

---

*А. Ю. Телицына*

Оптимизация научной деятельности через интеграцию ИИ: нейронные сети как инструмент в работе с академической литературой..... 218

DOI: [10.14515/monitoring.2024.5.2766](https://doi.org/10.14515/monitoring.2024.5.2766)**С. А. Шомова, А. Г. Качкаева****МЕЖДУ ОЧАРОВАНИЕМ И ИСПУГОМ: ДИАЛОГ С «ДРУГИМ».  
ОПЫТ АНАЛИЗА ПРАКТИК ИСПОЛЬЗОВАНИЯ ИИ  
В ПРОФЕССИОНАЛЬНОЙ И ПОВСЕДНЕВНОЙ ЖИЗНИ****Правильная ссылка на статью:**

Шомова С. А., Качкаева А. Г. Между очарованием и испугом: диалог с «другим». Опыт анализа практик использования ИИ в профессиональной и повседневной жизни // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 3—17. <https://doi.org/10.14515/monitoring.2024.5.2766>.

**For citation:**

Shomova S. A., Kachkaeva A. G. (2024) Between Charm and Fright: Dialogue with the “Other”. An Analysis of AI Practices in Professional and Everyday Life. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 3–17. <https://doi.org/10.14515/monitoring.2024.5.2766>. (In Russ.)

Получено: 09.10.2024. Принято к публикации: 25.10.2024.

**МЕЖДУ ОЧАРОВАНИЕМ И ИСПУГОМ: ДИАЛОГ С «ДРУГИМ». ОПЫТ АНАЛИЗА ПРАКТИК ИСПОЛЬЗОВАНИЯ ИИ В ПРОФЕССИОНАЛЬНОЙ И ПОВСЕДНЕВНОЙ ЖИЗНИ**

*ШОМОВА Светлана Андреевна — доктор политических наук, профессор, профессор Института медиа, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия*  
E-MAIL: [sshomova@hse.ru](mailto:sshomova@hse.ru)  
<https://orcid.org/0000-0003-0959-9278>

*КАЧКАЕВА Анна Григорьевна — кандидат филологических наук, профессор, Московская высшая школа социальных и экономических наук (Шанинка), Москва, Россия*  
E-MAIL: [anuta.kachkaeva@gmail.com](mailto:anuta.kachkaeva@gmail.com)  
<https://orcid.org/0000-0001-7577-5503>

**Аннотация.** «Нейросетевая революция», о которой сегодня рассуждают и теоретики, и практики технологических, производственных и медиаиндустрий, происходит в радикально меняющемся мире трагических противостояний, кризиса глобальных институтов, пересборки политических и экономических государственных партнерств и союзов, формирования наднациональных технологических цифровых экосистем. Все это требует серьезного осмысления со стороны науки — на междисциплинарном уровне, в самых разных тематических и предметных областях знания.

В статье обозначены некоторые векторы академических дискуссий, связанных в том числе с триумфальным вторжением искусственного интеллекта (и в особенности генеративных нейросетей) в различные сферы человеческой деятельности, коммуникации, творчества; с рисками и вызовами, привнесенными генеративным искусственным интеллектом в жизнь общества;

**BETWEEN CHARM AND FRIGHT: DIALOGUE WITH THE “OTHER”. AN ANALYSIS OF AI PRACTICES IN PROFESSIONAL AND EVERYDAY LIFE**

*Svetlana A. SHOMOVA<sup>1</sup> — Dr. Sci. (Polit.), Professor, Professor at the Institute of Media*  
E-MAIL: [sshomova@hse.ru](mailto:sshomova@hse.ru)  
<https://orcid.org/0000-0003-0959-9278>

*Anna G. KACHKAEVA<sup>2</sup> — Cand. Sci. (Philology), Professor*  
E-MAIL: [anuta.kachkaeva@gmail.com](mailto:anuta.kachkaeva@gmail.com)  
<https://orcid.org/0000-0001-7577-5503>

<sup>1</sup> HSE University, Moscow, Russia

<sup>2</sup> Moscow Higher School of Social and Economic Sciences (Shaninka), Moscow, Russia

**Abstract.** The “neural network revolution” that is discussed today by both theorists and practitioners of the technological, manufacturing, and media industries is taking place in a dramatically changing world of tragic confrontations, crisis of global institutions, re-establishment of political and economic state partnerships and alliances, formation of supranational technological digital ecosystems. All this requires serious reflection on the part of science, at the interdisciplinary level and in a variety of thematic and subject areas of knowledge.

This article outlines some vectors of academic discussions related, among others, to the triumphant invasion of artificial intelligence (and especially generative neural networks) into various spheres of human activity, communication, and creativity; to the risks and challenges brought by generative artificial intelligence into the life of society; to the emotional adaptation of homo sapiens to communication with a non-human but “intelligent” agent.

с эмоциональной адаптацией *homo sapiens* к коммуникации с нечеловеческим, но «разумным» агентом. Материал завершается обзором статей специального номера, авторы которых осмысливают и анализируют как применение искусственного интеллекта в образовании, коммуникациях, науке, креативных индустриях, так и использование нейросетей в медиа, судебных, религиозных, досуговых практиках и структуре новой повседневности.

**Ключевые слова:** искусственный интеллект, генеративные нейросети, этика ИИ, образование, наука, креативные индустрии, медиа, практики повседневности, социум

The authors conclude with a review of articles included in the special issue of the *Monitoring of Public Opinion: Economic and Social Changes Journal*, which reflect on and analyze both the use of AI in education, communications, science, creative industries, and the integration of neural networks in media, judicial, religious, and leisure practices, and in the structure of the new everyday life.

**Keywords:** artificial intelligence, generative neural networks, AI ethics, education, science, creative industries, media, everyday life, society

В 2023 г., который уже называют годом нейросетевой революции, стало очевидно, что скорость изменений, связанных с искусственным интеллектом (ИИ), беспрецедентна. Уже можно констатировать, что стремительность и масштаб динамических трансформаций, привнесенных ИИ во все сферы нашей жизни, многократно превышают скорость развития технологий предыдущего цикла. Важным нюансом этого технологического и социокультурного перелома стало еще и то, что в 2023 г. генеративные нейросети прошли «точку невозврата», миновали стадию, когда их использование было доступно лишь профильным специалистам, и вошли в эпоху повсеместного внедрения, став достоянием многих сотен тысяч рядовых пользователей.

«Нейросетевая революция» в нашей стране происходит с дополнительными сложностями — не только технологическими, психологическими, культурными, социальными, но еще и обусловленными факторами внешнеполитических и экономических противостояний. Международные санкции, как и ограничительные/блокирующие меры надзорных ведомств, существенно изменили культурно-образовательный ландшафт России. В ситуации «политической скованности» и сужения публичного альтернативного дискурса меняется сама креативная среда, возникают новые культурные коллаборации, форматы, темы и новые игроки. В последние два года у российских пользователей и потребителей появляется уникальный опыт — формируется «вынужденное» медиापоведение в условиях сокращающихся возможностей. Главная причина такого сокращения — изменения медиасреды из-за санкций, ухода производителей/прокатчиков, ужесточения законодательства в информационной и коммуникационной сферах, изменений на рекламном рынке, в сфере прав и доступа к контенту. Сохраняется проблема «цифрового разрыва» — не столько между разными регионами (происходящие изменения, скорее всего, выравнивают «медиаменю»), сколько из-за непересекающихся практик цифрового общения «молодых» (цифровые аборигены — *digital natives*) и их старших родственников

(*digital immigrants*). Тем не менее эти «цифровые травмы» неизбежно трансформируются и/или замещаются (особенно в массовом сегменте) отечественными сервисами, продуктами, произведениями. Развитость цифровой экосистемы страны — при всей недружелюбности архитектуры любых мировых платформ, направленной на овладение персональными данными, а значит, и пользовательским ресурсом, — уже сумела обеспечить цифровые привычки, удобный образ раздвоенной жизни (онлайн/оффлайн, одновременное нахождение «здесь и там»), источники и непрерывного раздражения, и удовольствия коммуникации с другими.

Мы вместе с редакцией журнала взялись обдумывать и собирать этот номер, когда внимание к генеративным нейросетям со стороны индустриального и академического сообщества начало подкрепляться социологическими данными, фиксирующими взрывной рост интереса к ИИ — не только со стороны профессионалов в области компьютерных технологий, программного обеспечения, кибербезопасности, но и со стороны бизнесменов, политиков, педагогов, дизайнеров, медиков, маркетологов, журналистов... При этом следует отметить, что наша страна движется вполне в русле этого общемирового тренда: несмотря на все ограничения и вызовы санкций, невзирая на вынужденное «сужение» пространства доступных медийных инструментов и сервисов, Россия находится в числе обладателей самых развитых цифровых экосистем («Яндекс», «Сбер», VK, с оговорками — Telegram) и делает ставку на активное внедрение искусственного интеллекта в различные сферы жизни. Это касается как профессионального взаимодействия российских компаний с ИИ (специалисты отмечают, что «после первой волны визионерского восприятия и завышенных ожиданий к технологиям ИИ решения на их основе постепенно становятся неотъемлемым инструментом бизнеса и применяются практически во всех сферах деятельности — от мониторинга сельхозугодий и управления космическими аппаратами до помощи в написании научных работ и новостей СМИ»<sup>1</sup>), так и использования россиянами нейросетей в личных целях — в организации пространства «умного дома», в досуговых и прикладных практиках, в структуре повседневного общения в социальных сетях. По результатам свежего (август 2024 г.) совместного социологического опроса консалтинговой компании «Яков и партнеры» и исследовательского холдинга «Ромир», уровень осведомленности об ИИ в России очень высок (84 % респондентов положительно ответили на вопрос «Знаете ли Вы, что такое искусственный интеллект?»); около четверти респондентов (24 %) используют генеративный искусственный интеллект для личных или рабочих нужд (хотя представители всех поколений, принявшие участие в опросе, указывают, что чаще всего используют ИИ все-таки в личных целях); почти каждый второй (47 %) из числа использующих ИИ отдает предпочтение российским решениям, таким как GigaChat или Kandinsky от «Сбера», YandexGPT или Шедерум от «Яндекса» (36 % пользуются одновременно и российскими, и зарубежными решениями)<sup>2</sup>. Другое свежее исследование — представленное ВЦИОМ в сентя-

<sup>1</sup> Искусственный интеллект в России: кто, что и как внедряет // Цифровая экономика. Бюллетень Института статистических исследований и экономики знаний (ИСИЭЗ) НИУ ВШЭ. 2023. 26 сентября. URL: <https://issek.hse.ru/mirror/pubs/share/862009044.pdf> (дата обращения 23.09.2024).

<sup>2</sup> Россияне и искусственный интеллект // Яков и партнеры. 2024. Август. URL: <https://www.yakovpartners.ru/publications/russian-citizens-and-ai/> (дата обращения 23.09.2024).

бре 2024 г., — показывает, что, пользуясь всеми цифровыми благами, «россияне скорее с осторожностью относятся к активному внедрению искусственного интеллекта в повседневную жизнь»: осознают необходимость контроля и регулирования рисков, связанных с его распространением, и ассоциируют этику ИИ с набором строгих правил и ограничений<sup>3</sup>.

Добавим к сказанному, что итоги множественных опросов позволяют сделать вывод о довольно сильной социально-демографической дифференциации наших современников по отношению к внедрению технологий ИИ; метафора о «цифровых оптимистах» и «цифровых пессимистах» успела стать расхожей не только в популярной публицистике, но и в академическом дискурсе. Тем не менее мир, в котором рекламные тексты и картинки для презентаций создаются как бы «сами собой», литература для научных статей подбирается в мгновение ока, а голливудские сценаристы кино и сериалов бастуют, добиваясь регулирования использования нейросетей, — это уже не фантастика, а наша рядовая повседневность. Этот мир существует за нашими окнами, на глазах меняет профессиональную реальность и бытовые привычки многих наших соотечественников, и нравится он нам или нет — жить в нем нам всем предстоит обязательно научиться.

### **Векторы академических дискуссий: о чем мы размышляем, когда размышляем об искусственном интеллекте?**

Как это часто бывает, реализованные на практике технологические и индустриальные прорывы сильно опережают научную рефлексию, связанную с возможностями и последствиями повсеместного внедрения нейросетей в привычную жизнь. И все-таки академических работ, осмысляющих роль ИИ в жизни современного человека и общества, насчитывается уже немало. На наш взгляд, тематически их можно разделить на несколько крупных сегментов, связанных с наиболее явными достижениями и/или проблемными узлами, привнесенными ИИ в социум (мы говорим в данном случае о темах, перекликающихся с центральными идеями и задачами этого номера журнала).

Во-первых, это работы, посвященные *триумфальному вторжению искусственного интеллекта (и в особенности генеративных нейросетей) в самые разные области человеческой деятельности, коммуникации, творчества*. Таковы, скажем, исследования роли ИИ в политике и электоральных процессах (см., например, [Efthymiou, Efthymiou, Sidiropoulos, 2020]), в правоприменительной и судебной сферах [de Oliveira et al., 2022; Rusakova, 2021 и др.], в медицине, психологических и поддерживающих практиках [Rajpurkar et al., 2022; Spiegel et al., 2024 и др.], в индустрии гостеприимства и туризма (см., например, [Fileri et al., 2021]) и так далее. В силу собственных жизненных интересов и личной профессиональной погруженности для нас особенно значимы также исследования, связанные с генеративным ИИ в области образования, науки, креативных индустрий. И надо признать, что именно в данных сферах нейросети наиболее зримо изменили и продолжают изменять ландшафт профессии; неслучайно эксперты отмечают: «В краткосрочной перспективе генеративный искусственный интеллект в большей сте-

<sup>3</sup> Этика искусственного интеллекта // ВЦИОМ. 2024. 3 сентября. URL: <https://wciom.ru/analytical-reviews/analiticheskii-obzor/ehitika-iskusstvennogo-intellekta-2> (дата обращения 23.09.2024).

пени повлияет на профессии, относящиеся к так называемой категории белых воротничков. Это представители творческих профессий (копирайтеры, дизайнеры, сценаристы, маркетологи), преподаватели, юристы, разработчики, менеджеры среднего звена, административный персонал и др.» [Искусственный интеллект в России..., 2023: 78]. Возможно, именно этим объясняется масштабное количество теоретических работ, изучающих «поведение» ИИ в названных областях. Например, в конце сентября 2024 г. на запрос «ChatGPT and higher education» поисковая система Google Scholar выдала 17,4 тыс. результатов, а на запрос «ChatGPT and creative industries» — около 30,9 тыс. Разумеется, перечислить все, даже самые значимые академические материалы по данным темам невозможно; обратим внимание лишь на некоторые свежие работы по направлениям, связанным с общими обзорами возможностей, которые привнесли генеративные нейросети в искусство и арт-индустрии (см., например, [Anantrasirichai, Bull, 2022]), с рефлексией изменений, которым подвергаются сами по себе творческие практики и современное искусство под влиянием ИИ [Atkinson, Barker, 2023; Kalpokiene, Kalpokas, 2023 и др.], с анализом новых артефактов массовой культуры (компьютерными играми, кинематографом, литературой, музыкой), созданными на основе и при помощи ИИ [Danesi, 2023], с возможностями нейросетей в среднем и высшем образовании [Jauhiaiainen, Guerra, 2023; Lo, 2023; Michel-Villarreal et al., 2023]. И многочисленные данные академических статей и промышленных отчетов (не только перечисленных, но и оставшихся за скобками), и собственные исследования авторов данного текста показывают, что мы можем говорить об усилении потенциала ИИ-технологий как «медиатора действительности», и это меняет представление о сути многих процессов и механизмов человеческой деятельности — политических, социальных, образовательных, креативных, творческих. Под влиянием нейросетей меняются сами представления человека о сути коммуникации, образования, творчества. И это и есть та самая новая реальность, к которой адаптируется человек двадцатых годов XXI века.

Второй корпус важных для нашей темы исследований, посвященных ИИ, — статьи и монографии, соотносящие «плюсы» с «минусами» и обозначающие риски и вызовы, привнесенные генеративным искусственным интеллектом в жизнь общества, включая морально-юридические аспекты проблемы. Здесь следует назвать как общие работы, связанные с философией, социологией, этикой, правоприменением ИИ (в том числе уже вышедшие/осуществленные и на русском языке (см., например, [Искусственный интеллект..., 2020; Правовые аспекты..., 2021; Кениг, 2023; Ловинк, 2024; Шталь, Шредер, Родригес, 2024; Calo, Froomkin, Kerr, 2016 и пр.]), так и тексты, оценивающие последствия использования ИИ в конкретных профессиональных областях, тесно пересекающиеся с первым, обозначенным выше, кругом работ [Heesen et al., 2022; Adeshola, Adepoju, 2023; Piskopani, Chamberlain, Carolyn, 2023 и др.]. Эти исследования очерчивают дискурсы, связанные с дебатами о социальных, этических и правовых последствиях использования ИИ-инструментов. По выражению одного из редакторов данного номера, мы переживаем эпоху «очарования и испуга» по отношению к изменившейся цифровой повседневности, и это мироощущение уже хорошо зафиксировано в академических дискуссиях. С одной стороны, создатели генеративного ИИ настаива-

ют (вполне справедливо) на его удобстве, пользе, облегчении жизни, свойствах креативности — и, конечно, прогресс и конкуренцию технологических корпораций уже не остановить. С другой стороны, ответственные исследователи и индустриалы-практики не могут не видеть экзистенциальных рисков, созданных машинами, а потому, вероятно, такие исследования составляют особенно важную часть современной рефлексии по поводу развития и распространения систем искусственного интеллекта.

Наконец, третий круг теоретических источников, важных для темы нашего специального номера, связан с эмоциональной адаптацией *homo sapiens* к коммуникации с нечеловеческим, но «разумным» агентом. Воспринимая искусственный интеллект (и гаджеты, созданные на его основе, то есть умные колонки, навигаторы с голосовыми ассистентами и т. д.) не только как помощника или инструмент, но и как вдохновителя, собеседника, рассказчика<sup>4</sup>, мы поневоле порой наделяем его человеческими чертами (то есть впадаем в «искушение антропоморфизма»), испытываем к нему позитивные или негативные чувства (восхищение, раздражение, гнев и прочие) или даже ждем проявления эмпатии с его стороны. Эти феномены уже неплохо исследованы современной наукой: теоретики анализируют этические советы, которые ChatGPT дает пользователям в качестве доверительного собеседника, и возникающие в связи с этим моральные дилеммы [Krügel, Ostermaier, Uhl, 2023]; взвешивают последствия «одомашнивания» искусственно-интеллекта [Berker et al., 2006]; размышляют, «можем ли мы и должны ли сочувствовать роботам» [Schmetkamp, 2020]; изучают связь очеловечивания ИИ («мой дорогой робот») с одиночеством его владельца [Arendt, 2015]; осмысливают механизмы возникновения антропоморфизма [Pelau, Dabija, Ene, 2021; Złotowski et al., 2014]. Общение, эмоциональное взаимодействие человека с машиной/алгоритмом — это, конечно, весьма своеобразная коммуникация. Однако нам еще предстоит осознать важную для гуманизации цифровых процессов идею, что пространство генеративного ИИ — это прежде всего пространство диалога. И не только по форме, содержанию, особенностям взаимодействия (чат, вопросы-ответы, интерактивность), но и по метафорическому смыслу — хотя речь и идет о диалоге с неодушевленным актором.

Развивая и уточняя эту мысль, скажем, что людям предстоит массово воспринять мысль: суть нейросети в том, что это не механические машины, следующие инструкции разработчика. Это машины кибернетические. В кибернетике как в научной дисциплине есть важное понятие «петля обратной связи»: это значит, что поведение машины не задается оператором как раз и навсегда данное, неизменное. Оно может корректировать само себя (петля обратной связи работает и в отношении «человеческого» знания). В этом смысле ограничение взаимодействия с «другим» не помогает взаимному развитию. Важнее придумывать новые «поддерживающие практики», «культуру протезирования», то есть модели продуктивного сотрудничества человека с технологией. Машины могут служить поддержанию

<sup>4</sup> См. об этом, например: Пространство соучастия — 2023. На пороге нейросетевой революции: искренние коммуникации, креативные практики и искусственный интеллект. Научный отчет об исследовании Центра цифровых культур и медиаграмотности НИУ ВШЭ. М., 2024. URL: <https://cdcmi.hse.ru/pubs/share/direct/966390091.pdf> (дата обращения 23.09.2024).

био- и техноразнообразия для создания новых перспектив взаимодействия человека и нечеловека<sup>5</sup>, но пока что люди живут в рамках самоисполняющегося пророчества о «восстании машин», о выходе алгоритмов из-под контроля, об умирании старых профессий и массовом появлении биороботов. И это психологически объяснимо: с одной стороны, неудивительно, что у человека при встрече с новым и неизведанным возникает эсхатологический страх, а с другой — технологические индустрии сознательно работают над заменой человеческого труда машинами, манипулируя вдобавок прежде чисто человеческими категориями рефлексии, креативности (как пример — «Алиса, давай придумаем...»). Необходимость поиска адекватного публичного языка для диалога людей и технологий, для осознанного принятия сосуществования машины вместе с человеком — это вызов и для технологических индустрий, и для исследователей, и для медиа.

Все перечисленные и многие иные проблемы так или иначе затрагиваются нашими авторами в тематическом номере журнала, который предлагается вашему вниманию.

### **Специальный номер «Погружение в нейросети: цифровая повседневность и практики взаимодействия человека и генеративного искусственного интеллекта»**

Академическая периодика России уже не впервые обращается к теме нейросетевой революции и ее последствий — среди наиболее интересных публикаций можно назвать, в частности, два тематических выпуска «дружественного» «Мониторингу общественного мнения» журнала ВЦИОМ «Социодиггер» «Человек и/или ChatGPT» и «Тандем исследователь — нейросеть: сила и слабость», целиком посвященные проблеме генеративных нейросетей<sup>6</sup>. Наш специальный номер ставил своей целью не только очертить круг дискуссий, ведущихся в отечественном академическом пространстве по поводу внедрения ИИ в структуру повседневности современного человека, но и обозначить «проблемные узлы» и лакуны, связанные с осмыслением «дивного нового мира» в его цифровой ипостаси, и расширить горизонты понимания этого мира за счет фиксации свежего эмпирического материала, помогающего лучше осознать этап технологической революции, на котором мы находимся.

С определенной долей условности все материалы номера можно поделить на несколько тематических блоков. Первый из них посвящен освоению искусственного интеллекта разными социальными сообществами — профессиональными, религиозными, связанными общими досуговыми или творческими интересами. Так, в статье «Соавтор, помощник, муза, инструмент: как студенты, преподаватели и представители креативных индустрий видят роль ИИ в своих практиках», подготовленной по итогам онлайн-анкетирования респондентов в пяти регионах России, Никита Кушкин и Оксана Тимофеева делают ставку на «замер самочувствия» сту-

<sup>5</sup> См. об этом, в частности: Hui Y. ChatGPT, or the Eschatology of Machines // E-flux Journal. 2023.June. URL: <https://www.e-flux.com/journal/137/544816/chatgpt-or-the-eschatology-of-machines/> (дата обращения: 01.10.2024).

<sup>6</sup> Социодиггер. Человек и/или ChatGPT. 2023. Т. 4. Вып. 5—6 (26). URL: <https://sociodigger.ru/releases/release/chelovek-i-ili-chatgpt/>; Социодиггер. Тандем исследователь — нейросеть: сила и слабость. 2023. Т. 4. Вып. 7—8 (27). URL: <https://sociodigger.ru/releases/release/tandem-issledovatel-neiroset-sila-i-slabost/> (дата обращения: 01.10.2024).

дентов, преподавателей вузов и представителей креативных индустрий в их взаимодействии с искусственным интеллектом, на саморефлексию этих социальных групп по отношению к ИИ. Как показало исследование, понимание и восприятие ИИ у этих трех категорий опрошенных неоднородно; авторы считают правильным говорить о (1) макро- и микроуровнях, (2) метафизическом и инструментальном, (3) воображаемом и контекстуальном, а также о (4) сходствах и различиях в понимании ролей ИИ, обусловленных основной сферой деятельности. Интересными оказались размышления респондентов о будущем в окружении цифровых технологий (например, о жизни рядом с метавселенными), о трансформации профессий, о том, какую функцию ИИ может выполнять в их повседневной работе или учебе. Важным результатом исследования стал также вывод о необходимости продвижения навыков и компетенций ИИ-грамотности, которая — как специфический вид медийной грамотности — пока еще не получила должного развития.

Антон Казун в работе «Может ли искусственный интеллект прогнозировать решения суда? Систематический обзор международных исследований» анализирует корпус международных работ, посвященных вопросу о том, насколько точно ИИ может предсказывать решения судей и, как следствие, есть ли у него шанс в перспективе заменить судью-человека. Автора занимает не только проблема точности прогнозирования и предложения судебных решений, но и — что крайне важно в условиях активного внедрения нейросетей в правоприменительную практику, — аспект этических ограничений, связанных с применением ИИ, когда речь идет о судьбе подследственного. Представляя обзор множества источников, исследующих данную дилемму, фиксируя стремительное совершенствование алгоритмов, используемых в этой сфере, а также сравнивая российский и зарубежный опыт, накопленный к сегодняшнему дню, автор одновременно выпукло очерчивает ключевые ограничения при использовании ИИ в судебной практике — и предлагает неоднозначный ответ на заданный в заголовке статьи вопрос.

Если применение ИИ в различных профессиональных областях во многом уже осмыслено и даже порой регламентировано, то постановка его на службу духовным интересам пока еще остается *terra incognita*. Ирина и Наталья Душаковы в статье «Генеративный искусственный интеллект в российских православных сообществах: восприятие и практики использования» анализируют практики использования нейросетей в религиозных сообществах России. Находясь в парадигме исследований цифровой религии, антропологии религии, социальных коммуникаций, а также на стыке исследований медиа и религиоведения, авторы рассматривают генеративный ИИ через призму социального конструирования технологий, исследуют запросы интересующей их аудитории, информационные поводы, вызывающие у нее отклик, способы использования ИИ верующими. Одним из результатов исследования, положенного в основу статьи, стал вывод о том, что, как и во многих других сообществах, представители религиозных общин не только активно обсуждают ИИ в социальных сетях, делятся сгенерированным контентом и/или оценивают его, но и выстраивают активные дискуссии вокруг попыток установить границы применения этой новой технологии.

Совсем иной тип социальной группы (но также активно взаимодействующий с искусственным интеллектом и до сих пор нередко остающийся за рамками вни-

мания теоретиков) рассмотрен в статье Евгении Ним «Поп-культура, фэндомы и нейросети: фанаты встречаются с ИИ». Рассматривая фэндомы как медиатизированные креативные сообщества, автор ищет и раскрывает взаимосвязи фан-культуры и искусственного интеллекта на примере двух фан-комьюнити. Утверждая, что экспансия искусственного интеллекта (в частности, нейросетей) вызвала поляризацию в фан-культуре (одни фанаты активно используют нейросети для редактирования и создания контента, а другие считают это неуместным), исследовательница вычленяет семь «перекрестков», где фэндом встречается с ИИ и трансформируется под его влиянием, а также фиксирует ключевые идеи, на которых фокусируются критические и протестные по отношению к нейросетям фан-дискурсы.

Следующий блок материалов номера условно объединяет две работы, в которых рассматриваются динамические изменения таких феноменов социальной реальности и духовной культуры, как доверие, репутация и этика. Андрей Сычев в статье «В поисках надежности: трансформация доверия в эпоху цифровых технологий» анализирует влияние, которое генеративный искусственный интеллект оказывает на состояние межличностного доверия и доверия к организациям и институтам. Предлагая гипотезу о том, что избыток и усложнение структуры информации привели к формированию представительского доверия, то есть к делегированию пользователем вынесения вердикта о надежности других людей и организаций доверенному представителю (в качестве которого в последнее время все активнее используется искусственный интеллект), автор изучает интересующий его феномен прежде всего на примере конструирования любовных отношений и формирования «цифровой репутации»; важным выводом работы стала фиксация двух взаимосвязанных процессов: постепенного очеловечивания искусственного интеллекта и компьютеризации человека.

Екатерина Коваль и Сергей Ушкин фокусируют свое внимание на вопросах профессиональной этики сообщества, специалисты которого, по мнению авторов, становятся новыми агентами влияния, воздействующими на общество не посредством убеждения, а через разрабатываемые и внедряемые технологии. В статье «Кому нужна этика больших данных: разработчики и их руководители о необходимости создания профессионального этического кодекса» ученые знакомят читателей журнала с итогами двухэтапного эмпирического исследования, состоящего из глубинных интервью и целевого опроса. Делая вывод, что российское сообщество IT-специалистов фрагментировано и разобщено, а потребность в этическом регулировании деятельности ярко в нем не выражена, авторы все же обнаруживают существующий в этой сфере запрос на создание определенных регулирующих механизмов, фиксируют высказанные респондентами требования к профессиональному кодексу этики, приводят обзор научных дискуссий о коллективной и личной ответственности в этике разработчиков.

Еще один небольшой тематический блок объединяет работы, исследующие взаимодействие человека с роботизированными устройствами и особенности коммуникации «человек — робот». В статье Ольги Буровой, Юлии Неверовой, Гюзель Селетковой и Елены Середкиной «Взаимодействие пользователей разного возраста с сервисным роботом: лабораторное исследование» представлены итоги эксперимента, который проходил на базе ООО «Промобот» в Перми; его участ-

ники, имитируя получение стандартной услуги по оформлению документов, обращались к разным типам агентов (человек и сервисный робот) в разных режимах работы. Авторам удалось не только выделить особенности фактического взаимодействия с антропоморфным роботом, но и зафиксировать нюансы субъективного восприятия этого опыта пользователями, относящимися к разным возрастным группам: обсудить ожидания участников от коммуникации с роботизированным устройством, оценить их невербальные реакции в ходе общения, наметить причины сниженной готовности к взаимодействию с роботами.

Наталья Трегубова и Анна Фейгина, в отличие от коллег, исследуют взаимодействие человека не с антропоморфными роботами, а с «умными устройствами». Статья «„Вежливое“ взаимодействие с умными колонками в повседневной жизни: универсальные нормы и новые формы социальности» базируется на результатах эмпирического исследования (18 интервью с пользователями умных колонок в Санкт-Петербурге, Тель-Авиве и Берлине), целью которого стало определение того, какие нормы и формы вежливости используют люди при коммуникации с Алисой, Сири и Алексой. Сделав попытку посмотреть, как умная колонка становится частью повседневной жизни в рамках домашнего пространства, и сформулировав ряд гипотез, авторы последовательно выделяют главные тенденции в коммуникации пользователей с новыми для них объектами одомашнивания, изучают корреляцию между процессом очеловечивания гаджета и нормами вежливости, проводят компаративный анализ проявлений вежливости, используемых при общении с умной колонкой в разных странах.

Наконец, назовем еще две работы, идущие в номере под отдельными рубриками. В работе «Искусственный интеллект — это твое», представляющей собой рецензию на книгу разработчика, бизнес-аналитика Дж. Ховарда и инженера-исследователя С. Гуггера «Глубокое обучение с fastai и PyTorch: минимум формул, минимум кода, максимум эффективности», Илья Буханский оценивает рассматриваемое издание как практическое пособие, призванное доказать читателю, что он, вне зависимости от образования и профессии, способен самостоятельно обращаться к одной из знаковых технологий XXI века — искусственному интеллекту — и создавать с его помощью продукты. Важно, что автор не только подробно останавливается на содержании глав книги, но и рассказывает, как сам он опробовал инструменты DL, предлагаемые Дж. Ховардом и С. Гуггером в рамках библиотеки fastai, на практике, чтобы понять, действительно ли принципы относительно простого и быстрого развертывания качественных нейросетей, делающие книгу такой привлекательной в глазах читателей, реализованы в программном обеспечении составителей текста.

Александра Телицына в статье «Оптимизация научной деятельности через интеграцию ИИ: нейронные сети как инструмент в работе с академической литературой» справедливо утверждает, что академическим ученым генераторы текстов могут быть полезны по множеству причин, но одновременно с этим она обращает внимание на риски использования ИИ в научной деятельности (особенно этического свойства), дает практические рекомендации по общению с нейросетью. В приложении к статье приводится описание и инструментальное сравнение сервисов на основе ИИ (Research Rabbit, Perplexity AI, ScholarAI, Diffusion, Reka Core, Smartcat

и других), не только помогающих оптимизировать поиск научной информации и использующих рекомендательные алгоритмы, но и позволяющих улучшать текст, генерировать разноформатный контент, прослушивать научные статьи в аудиоформате, визуализировать данные, делать переводы с одного языка на другой и т. д.

Трудно было бы ожидать, что представленные в специальном номере исследования интересующей нас проблематики — столь актуальной и находящейся сегодня в фокусе внимания ученых в разных странах мира, — дадут однозначные ответы на все поднимаемые вопросы или смогут разрешить сложные дискуссии, ведущиеся на тему практик взаимодействия человека с генеративными нейросетями. Несмотря на все разнообразие предлагаемых читателю работ, наш номер, конечно, — лишь некий абрис новой цифровой повседневности, обозначение отдельных важных векторов для осмысления феномена искусственного интеллекта, его роли в социуме и тех изменений, которые он привнес в жизнь общества. При этом приходится признать, что всем нам пока не хватает вокабуляра, чтобы описать новое качество взаимодействия человека и нейросети. Надо искать метафоры, определения, конвенции, чтобы сделать наше общее цифровое будущее с машинами/агентами/роботами — помощниками, спутниками, партнерами — полезным и комфортным. Статьи на «стыках» междисциплинарных подходов, представленные в специальном номере, — шаг на этом пути. А потому нам хотелось бы верить, что собранные материалы окажутся не только интересны, но и полезны всем исследователям, размышляющим на тему генеративного ИИ и работающим в соответствующем академическом поле.

## Список литературы (References)

1. Искусственный интеллект — надежды и опасения / под ред. Дж. Брокмана. М.: АСТ, 2020.  
Brockman J. (ed.) (2020) *Artificial Intelligence—Hopes and Fears*. Moscow: AST. (In Russ.)
2. Искусственный интеллект в России — 2023: тренды и перспективы / Отчет. М.: Яков и партнеры; Яндекс, 2023. URL: [https://www.yakovpartners.ru/upload/iblock/c5e/c8t1wrkdne5y9a4nqlicderalwny7xh4/20231218\\_AI\\_future.pdf](https://www.yakovpartners.ru/upload/iblock/c5e/c8t1wrkdne5y9a4nqlicderalwny7xh4/20231218_AI_future.pdf) (дата обращения: 29.10.2024).  
*Artificial Intelligence in Russia — 2023: Trends and Prospects (2023) Report*. Moscow: Yakov & Partners; Yandex. URL: [https://www.yakovpartners.ru/upload/iblock/c5e/c8t1wrkdne5y9a4nqlicderalwny7xh4/20231218\\_AI\\_future.pdf](https://www.yakovpartners.ru/upload/iblock/c5e/c8t1wrkdne5y9a4nqlicderalwny7xh4/20231218_AI_future.pdf) (accessed: 29.10.2024). (In Russ.)
3. Кениг Г. Конец индивидуума. Путешествие философа в страну искусственного интеллекта. М.: Individuum, 2023.  
Koenig G. (2023) *The End of the Individual. A Philosopher's Journey to the Country of Artificial Intelligence*. Moscow: Individuum. (In Russ.)
4. Ловинк Г. В плену платформы. Как нам вернуть себе интернет. М.: Ad Marginem, 2024.  
Lovink G. (2024) *Captive of the Platform. How We Can Reclaim the Internet*. Moscow: Ad Marginem. (In Russ.)

5. Правовые аспекты использования искусственного интеллекта: актуальные проблемы и возможные решения. М.: Издательский дом Высшей школы экономики, 2021. URL: <https://www.hse.ru/mirror/pubs/share/480106412.pdf> (дата обращения: 29.10.2024).  
Legal Aspects of the Use of Artificial Intelligence: Current Problems and Possible Solutions (2021) Moscow: HSE Publishing House. URL: <https://www.hse.ru/mirror/pubs/share/480106412.pdf> (accessed: 29.10.2024). (In Russ.)
6. Шталь Б. К., Шредер Д., Родригес Р. Этика искусственного интеллекта. Кейсы и варианты решения этических проблем. М.: Издательский дом ВШЭ, 2024.  
Stahl B. K., Schroeder D., Rodriguez R. (2024) Ethics of Artificial Intelligence. Cases and Options for Solving Ethical Problems. Moscow: HSE University Publishing House. (In Russ.)
7. Adeshola I., Adepoju A. P. (2023) The Opportunities and Challenges of ChatGPT in Education. *Interactive Learning Environments*. P. 1—14. <https://doi.org/10.1080/10494820.2023.2253858>.
8. Anantrasirichai N., Bull D. (2022) Artificial Intelligence in the Creative Industries: A Review. *Artificial Intelligence Review*. Vol. 55. P. 589—656. <https://doi.org/10.1007/s10462-021-10039-7>.
9. Arendt M. (2015) My Dear Robot: Anthropomorphism and Loneliness (Bachelor's Thesis, University of Twente). URL: <https://www.semanticscholar.org/paper/My-dear-robot-%3A-anthropomorphism-and-loneliness-Arendt/dca685080d03ad007b505c520e2a14c22a4762ec> (accessed: 29.10.2024).
10. Atkinson D. P., Barker D. R. (2023) AI and the Social Construction of Creativity. *Convergence*. Vol. 29. No. 4. P. 1054—1069. <https://doi.org/10.1177/13548565231187730>
11. Berker T., Hartmann M., Punie Y., Ward K. (2006) Domestication of Media and Technology. Berkshire: Open University Press.
12. Calo R., Froomkin A. M., Kerr I. (2016) Robot Law. Cheltenham: Edward Elgar Publishing.
13. Danesi M. (2023) AI-Generated Popular Culture. A Semiotic Perspective. London: Palgrave Macmillan Cham.
14. de Oliveira L. F., da Silva Gomes A., Enes Y., Branco T. V. C., Pires R. P., Bolzon A., Demo G. (2022) Path and Future of Artificial Intelligence in the Field of Justice: A Systematic Literature Review and a Research Agenda. *SN Social Sciences*. Vol. 2. Art. 180. <https://doi.org/10.1007/s43545-022-00482-w>.
15. Efthymiou Egleton I.-P., Efthymiou Egleton T.-W., Sidiropoulos S. (2020) Artificial Intelligence (AI) in Politics: Should Political AI be Controlled? *International Journal of Innovative Science and Research Technology*. Vol. 5. No. 2. P. 49—51.
16. Filieri R., D'Amico E., Destefanis A. et al (2021) Artificial Intelligence (AI) for Tourism: An European-Based Study on Successful AI Tourism Start-Ups. *International*

*Journal of Contemporary Hospitality Management*. Vol. 33. No. 11. P. 4099—4125.  
<https://doi.org/10.1108/IJCHM-02-2021-0220>.

17. Heesen J., Bieber C., Grunwald A., Matzner T., Roßnagel A. (2022) AI Systems and the Individual Electoral Decision. *Lernende Systeme — Germany’s Platform for Artificial Intelligence*. URL: [https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen\\_EN/AG3\\_WP\\_Executive\\_Summary\\_AI\\_and\\_electoral\\_decision.pdf](https://www.plattform-lernende-systeme.de/files/Downloads/Publikationen_EN/AG3_WP_Executive_Summary_AI_and_electoral_decision.pdf) (accessed: 14.08.2024).
18. Jauhiainen J. S., Guerra A. G. (2023) Generative AI and ChatGPT in School Children’s Education: Evidence from a School Lesson. *Sustainability*. Vol. 15. No. 18. Art. 14025. <https://doi.org/10.3390/su151814025>.
19. Kalpokiene J., Kalpokas I. (2023) Creative Encounters of a Posthuman Kind — Anthropocentric Law, Artificial Intelligence, and Art. *Technology in Society*. Vol. 72. <https://doi.org/10.1016/j.techsoc.2023.102197>.
20. Krügel S., Ostermaier A., Uhl M. (2023) ChatGPT’s Inconsistent Moral Advice Influences Users’ Judgment. *Scientific Reports*. Vol. 13. Art. 4569. <https://doi.org/10.1038/s41598-023-31341-0>.
21. Michel-Villarreal R., Vilalta-Perdomo E., Salinas-Navarro D.E., Thierry-Aguilera R., Gerardou F.S. (2023) Challenges and Opportunities of Generative AI for Higher Education as Explained by ChatGPT. *Education Sciences*. Vol. 13. No. 9. Art. 856. <https://doi.org/10.3390/educsci13090856>.
22. Lo Ch.K. (2023) What Is the Impact of ChatGPT on Education? A Rapid Review of the Literature. *Education Sciences*. Vol. 13. No. 4. Art. 410. <https://doi.org/10.3390/educsci13040410>.
23. Pelau C., Dabija D.-C., Ene I. (2021) What Makes an AI Device Human-like? The Role of Interaction Quality, Empathy and Perceived Psychological Anthropomorphic Characteristics in the Acceptance of Artificial Intelligence in the Service Industry. *Computers in Human Behavior*. Vol. 122. Art. 106855. <https://doi.org/10.1016/j.chb.2021.106855>.
24. Piskopani A., Chamberlain A., Carolyn T. (2023) Responsible AI and the Arts: The Ethical and Legal Implications of AI in the Arts and Creative Industries. In: *TAS ‘23: Proceedings of the First International Symposium on Trustworthy Autonomous Systems*. Art. 48. P. 1—5. <https://doi.org/10.1145/3597512.3597528>.
25. Rajpurkar P., Chen E., Banerjee O., Topol E. J. (2022) AI in Health and Medicine. *Nature Medicine*. Vol. 28. P. 31—38. <https://doi.org/10.1038/s41591-021-01614-0>
26. Rusakova E. P. (2021) Integration of “Smart” Technologies in the Civil Proceedings of the People’s Republic of China. *RUDN Journal of Law*. Vol. 25. No. 3. P. 622—633. <https://doi.org/10.22363/2313-2337-2021-25-3-622-633>.
27. Schmetkamp S. (2020) Understanding A.I.— Can and Should we Empathize with Robots? *Review of Philosophy and Psychology*. Vol. 11. P. 881—897 <https://doi.org/10.1007/s13164-020-00473-x>.

28. Spiegel B. M. R., Liran O., Clark A., Samaan J., Khalil C., Chernoff R., Reddy K., Mehra M. (2024) Feasibility of Combining Spatial Computing and AI for Mental Health Support in Anxiety and Depression. *NPJ Digital Medicine*. Vol. 7. Art. 22. <https://doi.org/10.1038/s41746-024-01011-0>.
29. Zlotowski J., Proudfoot D., Yogeeswaran K., Bartneck C. (2014) Anthropomorphism: Opportunities and Challenges in Human — Robot Interaction. *International Journal of Social Robotics*. Vol. 7. P. 347—360. <https://doi.org/10.1007/s12369-014-0267-6>.

DOI: [10.14515/monitoring.2024.5.2596](https://doi.org/10.14515/monitoring.2024.5.2596)



Е. А. Коваль, С. Г. Ушкин

## КОМУ НУЖНА ЭТИКА БОЛЬШИХ ДАННЫХ: РАЗРАБОТЧИКИ И ИХ РУКОВОДИТЕЛИ О НЕОБХОДИМОСТИ СОЗДАНИЯ ПРОФЕССИОНАЛЬНОГО ЭТИЧЕСКОГО КОДЕКСА

### Правильная ссылка на статью:

Коваль Е. А., Ушкин С. Г. Кому нужна этика больших данных: разработчики и их руководители о необходимости создания профессионального этического кодекса // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 18—36. <https://doi.org/10.14515/monitoring.2024.5.2596>.

### For citation:

Koval E. A., Ushkin S. G. (2024) Who Needs Big Data Ethics: Opinions of IT Developers and Leads on the Need to Create a Professional Code of Ethics. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 18–36. <https://doi.org/10.14515/monitoring.2024.5.2596>. (In Russ.)

Получено: 22.03.2024. Принято к публикации: 28.08.2024.

## КОМУ НУЖНА ЭТИКА БОЛЬШИХ ДАННЫХ: РАЗРАБОТЧИКИ И ИХ РУКОВОДИТЕЛИ О НЕОБХОДИМОСТИ СОЗДАНИЯ ПРОФЕССИОНАЛЬНОГО ЭТИЧЕСКОГО КОДЕКСА

*КОВАЛЬ Екатерина Александровна — доктор философских наук, ведущий научный сотрудник Института корпоративного обучения и непрерывного образования, Национальный исследовательский Мордовский государственный университет им. Н. П. Огарева, Саранск, Россия; профессор кафедры уголовного права и процесса, Средне-Волжский институт (филиал) Всероссийского государственного университета юстиции, Саранск, Россия*  
E-MAIL: [nwifesc@yandex.ru](mailto:nwifesc@yandex.ru)  
<https://orcid.org/0000-0003-0069-5335>

*УШКИН Сергей Геннадьевич — кандидат социологических наук, ведущий научный сотрудник отдела мониторинга социальных процессов, Научный центр социально-экономического мониторинга, Саранск, Россия; исследовательский менеджер департамента социальных исследований и консалтинга, Всероссийский центр изучения общественного мнения, Москва, Россия; младший научный сотрудник Института корпоративного обучения и непрерывного образования, Национальный исследовательский Мордовский государственный университет им. Н. П. Огарева, Саранск, Россия*  
E-MAIL: [ushkinsergey@gmail.com](mailto:ushkinsergey@gmail.com)  
<https://orcid.org/0000-0003-4317-6615>

**Аннотация.** Повсеместная цифровизация оказывает влияние не только на технологические процессы, но и на социальные нормы и практики. Многие решения, касающиеся человека и общества, все чаще принимаются с опорой на аналитику больших данных или делегируются инструментам, работающим на основе искусственного интеллекта. В таких условиях специалисты IT-сферы становятся новыми агентами влияния, которые воздействуют на общество не посредством

## WHO NEEDS BIG DATA ETHICS: OPINIONS OF IT DEVELOPERS AND LEADS ON THE NEED TO CREATE A PROFESSIONAL CODE OF ETHICS

*Ekaterina A. KOVAL<sup>1,2</sup> — Dr. Sci. (Philos.), Leading Researcher at the Institute for Corporate Education and Continuing Education; Professor at the Department of Criminal Law and Criminal Procedure*  
E-MAIL: [nwifesc@yandex.ru](mailto:nwifesc@yandex.ru)  
<https://orcid.org/0000-0003-0069-5335>

*Sergey G. USHKIN<sup>3,4,1</sup> — Cand. Sci. (Soc.), Leading Researcher at the Department for Monitoring Social Processes; Research Manager at the Department of Social Research and Consulting; Junior Researcher at the Institute for Corporate Education and Continuing Education*  
E-MAIL: [ushkinsergey@gmail.com](mailto:ushkinsergey@gmail.com)  
<https://orcid.org/0000-0003-4317-6615>

<sup>1</sup> National Research Mordovia State University, Saransk, Russia

<sup>2</sup> Mid-Volga Institute — Branch of the All-Russian State University of Justice, Saransk, Russia

<sup>3</sup> Scientific Center for Socio-Economic Monitoring, Saransk, Russia

<sup>4</sup> Russian Public Opinion Research Center (VCIOM), Moscow, Russia

**Abstract.** Widespread digitalization affects not only technological processes, but also social norms and practices. Many decisions that significantly impact individuals and society are increasingly made based on big data analytics or delegated to artificial intelligence. In such conditions, IT specialists become new agents of power who influence society not through persuasion, but through the technologies being developed and implemented. This explains the increased interest in the beliefs and val-

убеждения, а через разрабатываемые и внедряемые технологии. Это объясняет повышенное внимание к убеждениям и ценностям разработчиков новых информационных технологий, а также к их способности нести моральную ответственность за принимаемые профессиональные решения.

В статье представлены результаты двухэтапного эмпирического исследования, состоявшего из глубинных интервью ( $n=8$ ) и целевого опроса ( $n=50$ ). Анализ этих данных показал, что сообщество IT-специалистов фрагментировано и разобщено, потребность в этическом регулировании собственной профессиональной деятельности не выражена, этика воспринимается скорее как помеха или ненужное дополнение к правовым способам регуляции работы с большими данными. Руководители еще меньше заинтересованы в профессионально-этическом регулировании работы с большими данными, чем неруководящий персонал, что может быть связано с приверженностью интересам компаний, ориентированных на прямую или косвенную максимизацию прибыли. Несмотря на слабовыраженную потребность в разработке и внедрении кодекса профессиональной этики специалиста по работе с большими данными, IT-специалисты высказали ряд ценных предложений и замечаний относительно структуры и содержания такого кодекса, который в первую очередь должен быть направлен на обеспечение надежности хранения больших данных и защиту персональных данных.

**Ключевые слова:** большие данные, искусственный интеллект, искусственная социальность, профессиональный этический кодекс, этика больших данных

**Благодарность.** Исследование выполнено за счет гранта Российского научного фонда № 23-28-01288 (<https://rscf.ru/project/23-28-01288/>).

ues of IT developers, as well as their ability to take moral responsibility for their professional decisions.

The article presents the results of a two-stage empirical study, consisting of in-depth interviews ( $n=8$ ) and a targeted survey ( $n=50$ ). Analysis of these data showed that the community of IT specialists is fragmented and disunited. The need for ethical regulation of the professional activities in this group is not expressed, and ethics is perceived rather as a hindrance or an unnecessary addition to the legal means of regulating work with big data. Generally, IT leads are less interested in ethical regulation of work with big data than non-managerial personnel, which may be due to a commitment to the interests of companies focused on direct or indirect profit maximization. Despite the weakly expressed need for the development and implementation of a professional ethics code, big data experts made several valuable suggestions and comments regarding its structure and content and noted that it should primarily be aimed at ensuring the reliability of big data storage and the protection of personal data.

**Keywords:** big data, artificial intelligence, artificial sociality, professional ethical code, big data ethics

**Acknowledgments.** The article was prepared with the financial support of the Russian Science Foundation (project No. 2328-01288), <https://rscf.ru/project/23-28-01288/>.

## Введение

В последние годы наблюдается экспоненциальное увеличение количества цифровых данных по всему миру, обусловленное быстрым развитием компьютерных технологий и передовых методов вычислительной аналитики [Ienca et al., 2018]. Несмотря на то что повсеместно наборы информации, поступающей из различных источников (начиная от финансовых отчетов крупных компаний и заканчивая профилями пользователей в социальных сетях), стали называть «большими данными», среди исследователей не утихают споры относительно того, *какие же наборы данных действительно считать большими* [Favaretto et al., 2020]. Более того, как показывают результаты нашего исследования, значительная часть опрошенных нами специалистов в области компьютерных технологий затрудняются ответить, работают ли они с подобной информацией или нет.

В наиболее общем смысле под «большими данными» понимают технологию, которая основана на быстром приросте огромных объемов разнообразной информации [Попов, 2023: 68]. «Большие данные» можно охарактеризовать как *зонтичное понятие<sup>1</sup>, которое охватывает самые разные данные в одном наименовании* [Губа, 2018: 215; Китчин, 2017: 112]. И если ранее неотъемлемые атрибутивные составляющие больших данных можно было описать короткой формулой 3V—Volume (размер), Velocity (скорость) и Variety (разнообразие), то после их повсеместного проникновения в «ткань» современного общества к ним добавился широкий набор качественно иных характеристик как позитивного, так и негативного толка [Одинцов, 2017: 33].

Ряд кейсов, связанных с накоплением и последующим использованием больших данных, свидетельствует о том, что они не просто трансформируют взаимодействие человека и технологий, но и приводят к *изменению социальных практик*, а также моральных норм и ценностей. Например, общество в условиях накопления больших данных описывается при помощи модели суперпаноптикума, в котором крупные корпорации могут формировать не только потребительские практики пользователей, но и их культурные коды и ключевые ценности [Платонова, 2022].

Правовое регулирование дизайна новых технологий и их применение в различных сферах жизнедеятельности «не успевают» за самими технологиями. В процессе дизайна или эксплуатации высокотехнологичных продуктов возникают проблемы, которые при других обстоятельствах невозможно было даже предположить. В результате ответственные решения часто вынуждены принимать разработчики<sup>2</sup> и их руководители, которые тем самым, получают новые возможности в системе распределения власти. В сложившейся ситуации возрастает роль этического ре-

<sup>1</sup> Более того, как пошутил еще в 2013 г. известный израильско-американский экономист, профессор психологии и поведенческой экономики Дэн Ариели в одном из своих твитов, «Большие данные — как подростковый секс: все об этом говорят, но никто реально не знает, как это делается; все думают, что остальные этим занимаются, поэтому говорят, что занимаются этим».

<sup>2</sup> Мы обозначаем понятием «разработчики» максимально широкий круг профессионалов, работающих с большими данными и смежными технологиями: аналитики данных, дата-инженеры, дата-саентисты, аннотаторы данных, разработчики программного обеспечения и пр.

гулирования<sup>3</sup> деятельности *специалистов по компьютерным технологиям*. Однако думают ли так они сами? Мы пытались ответить на этот вопрос теоретическим и эмпирическим путем.

### **Коллективная vs. личная ответственность: обзор научных дискуссий об этике разработчиков**

В первую очередь необходимо прояснить нашу обеспокоенность недостаточностью этического регулирования сферы работы с большими данными. Она обусловлена тем, что разрабатываемые программные решения в действительности *не обладают ценностной нейтральностью*, несмотря на то что априорно позиционируются как объективные и непредвзятые. Разработчики и их руководители становятся субъектами властных отношений, поскольку знания представляют собой инструмент власти [Велиз, 2023: 62]. Более того, продолжая методологические традиции Мишеля Фуко, можно говорить о том, что знания сами по себе являются формой власти, которая не только воздействует на человека, но и создает его как субъекта [Фуко, 2023]. В качестве иллюстрации данного тезиса можно апеллировать к таким примерам, как влияние IT-компании Cambridge Analytica на выборы президента США<sup>4</sup> и борьба экс-президента Бразилии Ж. Болсонару с крупными социальными сетями, блокирующими ряд его политических заявлений<sup>5</sup>.

Повсеместная цифровизация привела к тому, что человечество в значительной степени стало *зависимым от компьютерных технологий*, которые в ряде случаев уже продемонстрировали эффективные манипулятивные воздействия в отношении пользователей, несмотря на заманчивые обещания освободить индивидов, входящих в виртуальные сетевые самоорганизующиеся сообщества, от старых властных структур государства. Как отмечает по этому поводу Юрген Хабермас, «огненная лава этого одновременно антиавторитарного и эгалитарного потенциала, который все еще ощущается в калифорнийском духе основателей стартапов первых лет, вскоре застыла в Силиконовой долине в либертарианской гримасе доминирующих в мире цифровых корпораций» [Хабермас, 2023: 50]. Один из пионеров Кремниевой долины Д. Ланье, рассуждая в 2013 г. о будущем<sup>6</sup>, делал акцент на том, что «нам даже не придется формулировать свои пожелания, потому что статистика облачных сервисов вычислит модель нашей личности настолько точно, что наши желания будут известны заранее» [Ланье, 2020: 22—23].

<sup>3</sup> В русскоязычном академическом дискурсе профессиональное этическое сообщество, как правило, обозначает термином «этика» философскую науку, теорию морали, а мораль (нравственность) понимает как предмет науки этики. В данном исследовании мы используем понятие «профессиональная этика» и производные от него понятия для обозначения и системы знания, и ценностно-нормативного регулирования профессиональной деятельности, и рефлексии по этому поводу. При этом в строгом смысле слова мы изучаем профессиональную мораль. Там, где в тексте встречаются обороты, касающиеся вопросов ответственности, долга и пр., мы используем устоявшиеся в науке этике обороты «моральная ответственность», «моральный долг», «моральный выбор».

<sup>4</sup> Goldberg M. Trump's High-Tech Dirty Tricksters // The New York Times. 2018. March 19. URL: <https://www.nytimes.com/2018/03/19/opinion/trump-cambridge-analytica-facebook.html?searchResultPosition=12> (дата обращения: 15.10.2024).

<sup>5</sup> Nicas J. Brazil's President Bans Social Networks from Removing Some Posts // The New York Times. 2021. September 9. URL: <https://www.nytimes.com/2021/09/09/world/americas/bolsonaro-social-networks.html> (дата обращения: 15.10.2024).

<sup>6</sup> Цитируемая книга Д. Ланье была опубликована на русском языке в 2020 г., но написана в 2013 г.

Современные продукты в сфере профилирования, таргетинговой рекламы и нейромаркетинга свидетельствуют о том, что будущее в какой-то мере уже наступило.

В сложившихся условиях возрастает роль специалистов, которые занимаются разработкой технологий, основанных на больших данных. Они «...как ключевые люди, ответственные за разработку решений на основе больших данных, могут самостоятельно изменить неэтичное решение на этическое (курсив наш. — Е.К., С.У.), если они мотивированы и способны на такое поведение» [Novák, Pavlíček, 2021].

Как мы видим, возвышается роль субъектов цифрового взаимодействия, в первую очередь компаний и их работников, причем даже тогда, когда сами они этого не замечают, и это требует (пере)определения правил игры и контроля их соблюдения как на институциональном, так и на индивидуальном уровне. Практика показывает, что *последствия принятия неэтичных решений при работе с большими данными могут быть весьма серьезными*, поскольку все чаще именно на основании аналитики больших данных выстраивается государственная политика, формируются прогнозы, определяются стратегические направления развития бизнеса и некоммерческого сектора и т. д. [Chen, Quan-Haase, 2020; Климович, 2020]. Иными словами, в своеобразном этическом фокусе находятся даже не столько сами большие данные или разработанные на их базе программные продукты, сколько завязанные на них алгоритмы и практики принятия морально обоснованных решений [Floridi, Taddeo, 2016].

Наиболее часто встречающиеся негативные последствия, возникающие в процессе работы с большими данными и смежными технологиями, связаны с приватностью пользователей и дискриминацией отдельных групп. Например, сервис распознавания лиц FindFace от российской компании NtechLab позволял деанонимизировать любого человека и активно использовался в целях кибербуллинга<sup>7</sup>. Приложение Boston's Street Bump App, которое оперативно собирало данные о выбоинах на дорогах, способствовало дискриминации пожилых и малообеспеченных граждан, не пользующихся смартфонами: в результате его использования власти получали данные преимущественно от жителей богатых районов, а в бедных районах ситуация с дорожными покрытиями ухудшилась [Favaretto, De Clercq, Elger, 2019: 3].

Безусловно, попытки моральной регуляции профессиональной деятельности, связанной с большими данными, предпринимаются в разных странах, хотя эффективность этих попыток существенно различается<sup>8</sup>. Например, в России с 2019 г. действует «Кодекс этики использования данных»<sup>9</sup>, согласно которому компании, попавшие в реестр добросовестных участников рынка данных, в каком-то смысле могут претендовать на статус объединения, отвечающего за профессиональ-

<sup>7</sup> См., например: Максим Перлин (FindFace): «Проблема приватности перегрета» О секретах пользователей «ВКонтакте» // Секрет фирмы. 2016. 22 апреля. URL: <https://secretmag.ru/trends/tendencies/findface.htm> (дата обращения: 25.06.2024).

<sup>8</sup> См., например: Data Ethics Framework. URL: <https://www.gov.uk/government/publications/data-ethics-framework> (дата обращения: 21.03.2024); Federal Data Strategy. Data Ethics Framework. URL: <https://resources.data.gov/assets/documents/fds-data-ethics-framework.pdf> (дата обращения: 21.03.2024); результаты исследования «The Ethics of Big Data: Balancing Economic Benefits and Ethical Questions of Big Data in the EU Policy Context». URL: <https://www.eesc.europa.eu/sites/default/files/resources/docs/qe-02-17-159-en-n.pdf> (дата обращения: 21.03.2024).

<sup>9</sup> Кодекс этики использования данных. URL: [https://ac.gov.ru/uploads/\\_Projects/PDF/kodeks-sait-3-pdf.pdf](https://ac.gov.ru/uploads/_Projects/PDF/kodeks-sait-3-pdf.pdf) (дата обращения: 21.03.2024).

но-этические вопросы в сфере больших данных. Проблема видится в том, что он адресован не непосредственно специалистам в области больших данных, а цифровым корпорациям, которые отнюдь не заинтересованы в ограничениях своей деятельности со стороны государства или общества.

Ориентация на участников рынка данных предполагает задействование механизмов *социальной ответственности корпораций*, однако таких механизмов недостаточно для регулирования проблем этического характера, связанных с разработкой и внедрением новых технологий [Rönnegard, 2023]. Государственное регулирование технологии больших данных в России также в большей степени адресовано госорганам и бизнесу, а не отдельным профессионалам. Под эгидой профильного Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации (Минцифры) правовое регулирование больших данных осуществляется в двух направлениях: с одной стороны, государство разрешает коммерциализацию данных, с другой стороны, обязывает бизнес безвозмездно предоставлять собранные им данные государству<sup>10</sup>. Все полученные данные в формате государственной фабрики больших данных при этом планируется использовать не только в целях государственного управления, но и для обучения нейросетей, в том числе коммерческих<sup>11</sup> (при условии обезличивания любых персональных данных, которые попадают в датасеты)<sup>12</sup>.

Моральная ответственность непосредственных разработчиков пока остается неопределенной. Исследователи отмечают, что уже на этапе прохождения обучения *формируется образ инженера как технического работника*, и это «скорее устраняет саму возможность возникновения в практике современного инженера ситуации морального выбора, поскольку решение профессионально-этических вопросов делегируется специальной институцией» [Богданова, Бакштановский, 2023: 139]. Впрочем, в российском профессиональном стандарте «Специалист по большим данным»<sup>13</sup> содержится требование знать этические аспекты предоставления услуг на основе аналитики больших данных, но здесь оно носит утилитарный, прикладной характер и представляет собой скорее декларацию о необходимости такого рода знания.

В подобных ситуациях моральной неопределенности хорошим разработчиком является не тот, кто виртуозно работает по инструкциям и составляет их, а тот, кто заботится о конечном пользователе через свой продукт. Но, как мы видим, в условиях «перегретого» IT-рынка *сложно внедрить моральную чувствительность как необходимый навык* на системном уровне, особенно для специалистов по данным. Для этого необходима серьезная просветительская работа в сфере этики, привлечение внимания не только государства, но и общества. Кроме того, важно

<sup>10</sup> См., например, Глава Минцифры о коммерциализации больших данных, росте интернет-трафика и уехавших IT-специалистах // DIRUSSIA.RU. 2023. 27 апреля. URL: <https://d-russia.ru/glava-minicifry-o-kommercializacii-bolshih-dannyh-roste-internet-trafika-i-uehavshih-it-specialistah.html> (дата обращения: 25.06.2024).

<sup>11</sup> В России планируется создать государственную фабрику больших данных // ТАСС. 2023. 8 ноября. URL: <https://tass.ru/ekonomika/19229859> (дата обращения: 24.06.2024).

<sup>12</sup> Большие данные. Подход Минцифры // Минцифры. 2023. 18 апреля. URL: <https://digital.gov.ru/ru/events/43766/> (дата обращения: 24.06.2024).

<sup>13</sup> Профессиональный стандарт «Специалист по большим данным». URL: <https://cdnstatic.rg.ru/uploads/attachments/193/24/38/59174.pdf> (дата обращения: 21.03.2024).

не допустить роста известного отчуждения «мира профессионализма» от гуманистических задач самой профессии, которое проявляется при ориентации исключительно на профессиональный успех, безотносительный к применяемым для его достижения средствам и безразличный к последствиям принимаемых решений [Бакштановский, Согомонов, 2005: 12].

Вероятно, появление масштабных профессиональных объединений и ассоциаций специалистов по данным также могло бы помочь в решении подобного рода вопросов: авторитетные структуры могут влиять на политику компаний, на нормативное регулирование профессиональной деятельности, а также приносить ряд преимуществ своим членам (например, скидки на обучающие мероприятия от лучших представителей профессии, преимущества при трудоустройстве и пр.). Но появление подобного рода структур эффективно, когда оно происходит по инициативе «снизу»; пока же сообщества специалистов образуются преимущественно вокруг конкретного типа разработок либо поддерживаются в форме междисциплинарных онлайн-проектов (например, Stack Overflow, Хабр и др.).

Недавнее международное исследование, посвященное изучению этического самоопределения разработчиков программного обеспечения, где в общей сложности опрошено более 130 респондентов из стран пяти континентов, постулирует, что даже при условии осознания этических последствий своих действий разработчики преимущественно не видят возможности что-либо изменить [Widder et al., 2023: 474]. Получается, мы имеем дело со своего рода *латентной субъектностью*, неосознанием моральным субъектом себя как такового.

Таким образом, разработчики и их непосредственные руководители имеют высокую степень влияния на общество, так как действуют через технологические продукты, а не через убеждение. Скорость воздействия, учитывая скорость развития технологий, также опережает иные способы влияния, осуществляемые другими профессионалами и группами профессионалов. По этой причине этические убеждения и ценности разработчиков особенно важны. Они, несмотря на то что бедны данными, управляют обществом, хотя не всегда это осознают.

## Методология и данные

Настоящая работа опирается на эмпирические данные, полученные в 2023 г. в ходе реализации исследовательского проекта «Этика больших данных: трансформация моральных норм и ценностей», поддержанного Российским научным фондом. Использована комбинированная методика сбора социологической информации, которая позволила выявить отношение разработчиков и их руководителей к проблемам морального регулирования различных этапов работы с большими данными (сбор, обработка, хранение, аналитика, использование, в том числе для машинного обучения).

На первом этапе мы провели серию *глубинных интервью* с представителями различных компаний и вузов, которые занимаются развитием больших данных и/или имеют соответствующие структурные подразделения и являются стейкхолдерами анализируемой отрасли. Исследование проведено в августе — сентябре 2023 г. Гайд экспертного интервью содержал четыре смысловых блока. В первый блок были включены вопросы о текущем состоянии и тенденциях развития боль-

ших данных в стране и мире; во второй — вопросы, направленные на выявление отношения информантов к этико-нормативной и законотворческой деятельности в этой сфере; в третий — вопросы, касающиеся содержания моральных установок и убеждений экспертов в сфере профессиональной деятельности; в четвертом блоке использовался метод виньеток, позволяющий раскрыть особенности точки зрения опрашиваемых на конкретные ситуации, возникающие в ходе работы с большими данными и смежными технологиями и имеющие признаки моральных дилемм. Длительность каждого интервью составила от 40 до 80 минут, использовалась платформа Zoom. Транскрибация осуществлялась посредством нейросетевых инструментов распознавания речи (в частности, использовалась модель машинного обучения для транскрибирования звуковых файлов Whisper, разработанная компанией OpenAI) с последующей проверкой и вычиткой авторами настоящей статьи. Всего опрошено восемь информантов (подробная информация о них приведена в Приложении), чьи ответы используются в статье в качестве иллюстраций происходящих процессов. Выборка конструировалась методом снежного кома. Относительно небольшое количество задействованных в опросе разработчиков компенсируется их высоким уровнем экспертности (в числе опрошенных — представители крупных коммерческих IT-организаций, руководители профильных вузовских лабораторий, разработчики «Кодекса этики больших данных» и т. д.), а также достаточно высокой содержательностью полученных ответов, дающих контекст понимания ситуации с использованием больших данных в России и мире.

На втором этапе по материалам глубинных интервью мы сконструировали анкету для *целевого опроса* специалистов в области компьютерных технологий. Исследование проведено в ноябре — декабре 2023 г. Инструментарий анкетирования, как и гайд интервью, содержал четыре блока вопросов: 1) паспортчика и установление уровня экспертности (как давно и как часто специалист работает с большими данными); 2) вопросы о персональной ответственности специалиста по большим данным; 3) вопросы о перспективах разработки и использования кодекса профессиональной этики специалиста по большим данным; 4) блок мини-кейсов, предполагающих принятие того или иного решения специалистами по компьютерным технологиям в условиях моральной неопределенности. Для проведения анкетирования использована платформа Google Forms. Всего в исследовании приняли участие 50 человек (из них 17 отнесли себя к руководящему составу — директор, заместитель директора, заведующий отделом, заведующий лабораторией и т. д.). Как и в случае с глубинными интервью, при конструировании выборки использован метод снежного кома. Относительно небольшое количество опрошенных объясняется тем, что группа специалистов в области компьютерных технологий достаточно закрыта, доступ к ней ограничен. Кроме того, как было сказано во введении, профессиональные ассоциации и объединения если и создаются, то они достаточно фрагментированы, посвящены лишь конкретным типам разработок или онлайн-проектам, и это становится заметным барьером для целевого отбора группы, которую нельзя назвать «прозрачной» (важно понимать, что она плохо репрезентуется статистикой, в ней действует большое количество фрилансеров, многие большие компании используют аутсорсинг и т. д.).

## Мечтают ли разработчики и их руководители об этике: проблемы «второго порядка» и пути их решения

Разработчики и их руководители сдержанно относятся к необходимости этического регулирования работы с большими данными, отмечая, что возникновение моральных дилемм в их практике скорее исключение, нежели повторяющееся с какой-либо периодичностью явление. Непосредственно этику они рассматривают достаточно утилитарно, как надстройку к правовому базису. Тезисно в ходе глубинных интервью проговаривалось, что соблюдения действующего законодательства как на уровне страны в целом, так и отдельной компании достаточно для разрешения большинства возникающих спорных ситуаций.

В некоторых случаях информанты отмечали, что этическое регулирование избыточно и даже вредно: наложение любых запретов рассматривается ими как препятствие техническому прогрессу. Моральное обоснование этой позиции заключается в том, что сфера разработки новых информационных технологий должна регулироваться в минимальной степени, чтобы не мешать делать прорывные открытия. Такие открытия позволяют принципиальным образом изменить жизнь общества, решить социальные проблемы, определить перспективные направления социального развития. Однако можно усмотреть и корыстный мотив, поскольку эти открытия, особенно если они делаются в бизнес-структурах, приносят сверхприбыли, а любые ограничения затормаживают процесс и тем самым мешают зарабатывать больше.

В то же время информанты проводят аналогии с действующими нормами права, которые пока трактуются двояко. С одной стороны, в некоторых сферах наблюдается излишний бюрократизм (например, защита персональных данных, обмен данными и т. д.), а с другой стороны, в отдельных отраслях надлежащее правовое регулирование отсутствует (например, защита авторских прав). По большому счету, и то, и другое препятствует технологическому развитию, обуславливает необходимость каких-либо изменений.

Отсутствие регулярно возникающих спорных ситуаций, причем не только в моральной, но и в правовой системе координат, демонстрируют результаты проведенного целевого опроса специалистов в области компьютерных технологий. Более половины указывают на то, что очень редко или никогда не сталкивались с необходимостью сделать выбор в пользу того или иного равнозначного решения (см. табл. 1). Тем не менее трое из десяти (14 человек — говоря о моральной неопределенности, 15 человек — о правовой) сообщили, что подобные ситуации имели место в их практике.

Таблица 1. Распределение ответов на вопрос: «Скажите, как часто при работе с большими данными Вы сталкиваетесь с ситуациями...» (количество человек)

	...моральной неопределенности (выбор между равноценными решениями в спорной ситуации)	...правовой неопределенности (выбор между принципиально разными нормами, например, разрешительной и запретительной)
Постоянно	2	1
Время от времени	12	14
Очень редко	14	13
Никогда	14	14
Затрудняюсь ответить	8	8

Исходя из анализа глубинных интервью, можно утверждать, что чаще всего проблемы возникают в связи с обеспечением конфиденциальности пользователей и связанных с ней вопросов анонимизации, обеспечения надежного хранения больших данных, особенно больших персональных данных, устранения последствий утечек, пресечения использования данных с городских видеокамер для слежки и т. д. Здесь важно отметить, что, хотя некоторые примеры отражали опыт самих информантов, в большей части иллюстрируемых случаев они опирались на сведения, полученные от коллег.

*Мой проект был про генерацию фейковых сайтов для пользователя. То есть это неэтично. Но как бы он нес в себе другой аспект. Чтобы научиться такие сайты искать, нужно их научиться создавать. Нужно показать нейронке, что вот это хорошо, а вот это плохо. (Инф. 2, муж., 22 года, студент, независимый разработчик)*

*К нам приходили с запросом из одной организации найти в базе как бы людей, которые с одной стороны показывают, что они бедные и пользуются социальными благами государства, а с другой стороны по косвенным признакам понятно, что они не бедные. Мы отказали, потому что мы не вмешиваемся в тайну частной жизни. (Инф. 8, жен., 49 лет, руководитель IT-подразделения крупной коммерческой организации)*

Кроме того, общим местом ряда глубинных интервью стало обоснование того, что *проблемы этического характера должны волновать исключительно руководство*, но не рядовых сотрудников, которые обслуживают свой участок IT-конвейера и часто не имеют представления о конечном назначении исходного продукта. К сожалению, практика такова, что в случае возникновения сбоев и поломок производственных процессов, не связанных с ситуациями моральной неопределенности, ответственность нередко ложится на низовых сотрудников, а не на управленческий сектор. Хотя по факту самой большой проблемой этического плана для рядовых разработчиков может стать кража данных на своем участке работ.

*Если мы говорим о маленьких компаниях, где там один-два программиста что-то делают, я думаю, они в принципе на тему этики не размышляют. Им директор сказал — они сделали. Кто они, куда они отгрузили, то есть у них и уровня понимания нет. (Инф. 8, жен., 49 лет, руководитель IT-подразделения крупной коммерческой организации)*

Подавляющее большинство опрошенных представителей целевой группы на декларативном уровне заявляют, что *специалист по работе с большими данными должен отказаться от реализуемого проекта*, если достоверно знает, что результаты его работы могут быть использованы во вред пользователям (например, персональные данные переданы третьей стороне) (см. табл. 2). Различия в представлениях руководящего и неруководящего состава минимальны, из чего можно сделать достаточно осторожный вывод: вероятно, при возникновении у разработчиков спорных ситуаций подобного рода (при условии, что речь идет именно о ситуации моральной неопределенности, а не о нарушении законодательства

или внутрикорпоративных норм) в случае аргументированного обращения к вышестоящему начальству они будут как минимум выслушаны.

**Таблица 2. Распределение ответов на вопрос: «Как Вы думаете, должен или нет специалист по работе с большими данными отказаться от реализуемого проекта, если достоверно знает, что результаты его работы могут быть использованы во вред пользователям (например, персональные данные переданы третьей стороне)?» (количество человек)**

	Руководящий состав	Неруководящий состав
Безусловно должен	6	12
Скорее должен	7	17
Скорее не должен	1	2
Безусловно не должен	1	2
Затрудняюсь ответить	2	0

Проектами, от которых стоит отказаться по моральным соображениям, информанты в ходе глубинных интервью преимущественно называли GPS-трекеры для слежения за людьми (включая детей), вредоносные программы, отслеживание людей по фотографии, публикацию и высокие риски утечки полных персональных данных, позволяющих идентифицировать человека. Однако важно отметить, что высказывались и менее популярные, альтернативные мнения. В частности, некоторые информанты апеллировали к логике «меньшего зла»: можно взяться за проект, который причиняет вред отдельным людям, но предотвращает причинение вреда гораздо большему числу людей. Наиболее яркий пример — о том, что в Африке нанимали аннотаторов за минимальную сумму размечать неблагоприятный контент (расчленение людей, порнография и пр.), чтобы он не попал в датасеты для обучения больших языковых моделей, с которыми теперь взаимодействует огромное число пользователей по всему миру.

В целом вопросы нормативного регулирования профессиональной деятельности — как правового, так и морального, — хоть и привлекают внимание разработчиков и их руководителей, но не настолько, чтобы детально разбираться в структуре и содержании конкретных документов, которые де-юре существуют, но де-факто не применяются. Так, информанты знают, что в России принят «Кодекс этики использования данных», но далеко не все знакомы с его содержанием, и, соответственно, с тем, что он адресован компаниям и только опосредованно — профессионалам.

Тем не менее на декларативном уровне подавляющее большинство опрошенных нами представителей целевой аудитории выступают за необходимость создания профессионального этического кодекса для специалистов по работе с большими данными (см. табл. 3). При этом здесь важно обратить внимание на расхождения (пусть и незначительные) в позициях разработчиков и их руководителей: первые гораздо чаще говорят о необходимости подобного регулирования, в то время как убеждения вторых не столь категоричны (примерно половина поддерживает, остальные или против, или затрудняются с ответом).

**Таблица 3. Распределение ответов на вопрос: «Как Вы думаете, нужен или нет специалистам по работе с большими данными профессиональный этический кодекс?» (количество человек)**

	Руководящий состав	Неруководящий состав
Безусловно нужен	3	6
Скорее нужен	6	21
Скорее не нужен	3	5
Безусловно не нужен	2	0
Затрудняюсь ответить	3	1

Показательно, каким образом обосновывают информанты в ходе глубинных интервью свой скептицизм в отношении разработки профессионального этического кодекса. Главный их аргумент заключается в том, что он априорно не обладает какой-либо силой, поэтому его исполнение будет сводиться исключительно к системе рекомендаций, которые гипотетически могут препятствовать разработке продукта (например, могут появиться требования о согласовании чего-либо с другими отделами, контроль этической комиссии, внешнее вмешательство и т. д.).

*Мне кажется, это лишнее, потому что у нас есть, например, уголовный кодекс, и я думаю, что кодекс этики... если он будет слабее, то в нем смысла нет, потому что в уголовном кодексе и так прописана вся ответственность. А если он будет сильнее, ну зачем еще сильнее? Многие ограничения излишние, и они скорее тормозят прогресс, чем, наоборот, ему способствуют. (Инф. 5, муж., 39 лет, руководитель вузовской IT-лаборатории)*

*Никогда не интересовались кодексами. Даже не считаю это хоть сколько-нибудь серьезной историей. Нас интересует то, что нормативно закреплено. То, что не закреплено нормативно, нас не интересует, поскольку не входит в сферу регулирования. (Инф. 6, муж., 41 год, руководитель коммерческой IT-организации)*

### **Представления о кодексе больших данных: краткость — сестра порядка?**

Несмотря на то что этические проблемы не относятся к первоочередным, существует не только запрос, пусть и невыраженный, на создание неких регулирующих механизмов, но и *предъявляемые требования к ним*. Безусловно, подобный набор можно встретить и в отношении других разрабатываемых документов, выходящих за пределы компьютерных технологий, однако, как нам кажется, он в значительной степени характеризует инженерное мышление, где профессиональный кодекс этики становится своего рода алгоритмом. Приведем обобщенные требования к кодексам, высказанные информантами.

Во-первых, *кодекс должен быть кратким* (содержать меньше «воды»). Как мы уже отмечали, информанты редко читают нормативные документы, в том числе ввиду их сложного, бюрократического языка, увеличивающего объем текста и усложняющего его понимание.

Во-вторых, кодекс должен содержать перечень кейсов, которые были успешно разрешены (так называемая «Белая книга»). Подобный свод правил мог бы выполнять функции навигатора по ценностному миру.

*Вот в этой серой зоне нужно, получается, опираться на свою моральную интуицию, потому что кодекса, куда можно было бы залезть, или хотя бы в белую книгу для этого персонального кодекса (мы сталкивались с такой проблемой, мы решили ее так вот) — такого костыля не существует, где можно было бы составить перечень таких случаев, куда можно было бы посмотреть. (Инф. 6, муж., 40 лет, руководитель группы разработчиков)*

*Какой-то свод общих правил, советов по использованию, хорошо было бы иметь, для того чтобы можно было в случае каких-то внутри себя возникающих вопросов при работе обратиться к нему. (Инф. 3, муж., 22 года, разработчик в коммерческой компании)*

В-третьих, кодекс должен включать перечень пунктов с ценностными требованиями<sup>14</sup>, которым в большинстве своем хотят соответствовать специалисты. Приверженность этим пунктам они могли бы публично декларировать, а компании выбирали бы соответствующих специалистов, позиционирующих себя как приверженцев этичным практикам.

*Также для программистов можно было бы составить некоторый такой набор пунктов, которым они соответствуют в рамках своих каких-то моральных практик. И они, возможно, не обязательно все пункты для себя выберут из этого списка, но могут выбрать конкретные пункты, которым они хотят соответствовать. (Инф. 6, муж., 40 лет, руководитель группы разработчиков)*

В-четвертых, кодекс должен дать четкое определение данным как чужому имуществу, которое принадлежит гражданину (в первую очередь персональные данные) или компании (например, информация с датчиков «умных» устройств), и заявить в качестве базовой ценности бережное отношение к такому имуществу, недопущение утечек, разглашения и пр.

В-пятых, кодекс должен закрепить как ценность чистоту данных. Недопустимо портить данные, что-то в них менять, добавлять, если это не обусловлено необходимостью чистки массива от «мусора» и т. п.

*Надо меру ответственности своей понимать, что ты часть этой цепочки, ты отвечаешь не за то, что ты передал данные, а за конечный результат. Что продукт, который потом что-то сделает на основе твоих данных, что он работает правильно, потому что данные сделаны правильно. (Инф. 8, жен., 49 лет, руководитель IT-подразделения крупной коммерческой организации)*

<sup>14</sup> Под ценностными требованиями информанты понимают ценности, которые фигурируют в этических документах, применимых в смежных областях. В частности, упоминались принципы работы с ИИ, сформулированные на конференции в Асиломаре (безопасность, ответственность, защита личных данных и др.).

Таким образом, кодекс профессиональной этики будет работать только в том случае, если он будет компактным, практико-ориентированным, а главное — будет приносить специалистам в области больших данных дополнительные выгоды, которых у них нет сейчас. Например, с позиций организаций дивидендами может выступать снижение репутационных издержек, с позиций разработчиков и их руководителей — признание в профессиональном сообществе, преференции при трудоустройстве в крупные компании, продвижение по карьерной лестнице и т. д. Тогда проявится потребность в этическом регулировании собственной профессиональной деятельности не только на государственном или организационном, но и на «низовом» уровне.

Что касается наиболее проблемных зон, требующих нормативного регулирования, то они также проговаривались в ходе глубинных интервью, но наиболее красноречиво подсветились по результатам целевого опроса специалистов по компьютерным технологиям. Так, восемь из десяти говорят, что кодекс должен концентрироваться на надежности хранения данных (38 человек); столько же — на защите персональных данных (38). Трое из десяти (17) предлагают сделать акцент на организации сбора больших данных, лишь незначительная часть говорит о разработке объяснительных моделей работы генеративного искусственного интеллекта, построенного на больших данных (8) и подборе и подготовке данных для разработки больших языковых моделей (7). Из всего пула проблем работы с большими данными специалисты в наибольшей степени обеспокоены проблемой их утечек, утратой персонифицированной информации о пользователях, и именно с этим связывают главные этические трудности, которые встречаются почти во всех интервью и отмечаются большинством информантов целевого опроса.

## Вместо заключения

Результаты исследования показывают, что специалисты демонстрируют заинтересованность в разработке профессионального этического кодекса. Существующих механизмов регулирования профессиональной деятельности недостаточно, а кроме того, они, во-первых, трудно применимы на практике в силу засилья абстрактных положений и деклараций, во-вторых, непрозрачны для широкой общественности. В ходе глубинных интервью подсвечиваются и специфичные проблемы, например регуляции труда тех, кто вручную занимается детоксикацией больших массивов информации, поскольку ежедневно подвергается стрессовым ситуациям ввиду жестокого, местами непристойного контента.

Но наше исследование показало двойственность восприятия границ профессионального этического кодекса. С одной стороны, он не должен мешать производственным процессам, постоянно отвлекая разработчиков и их руководителей от решения насущных вопросов. С другой стороны, если случаются ситуации моральной неопределенности, должен существовать некий протокол, алгоритм по ее устранению, который если не полностью, то хотя бы по большей части был бы применим в текущих реалиях. Подобного рода потребность, по всей видимости, в большей степени характерна для простых разработчиков, но не для их руководителей, которые, очевидно, сконцентрированы преимущественно на продвижении ценностей компании, включая максимизацию прямой или косвенной прибыли.

В системе нормативного регулирования работы с большими данными и искусственным интеллектом опрошенные разработчики и их руководители уповают на право, указывают, что *главное — не нарушать закон*. Однако законодательное регулирование разработки и внедрения технологий не успевает за технологическим развитием. Законодатель апеллирует к экспериментальным правовым режимам, «регуляторным песочницам», в рамках которых часто «обкатываются» новые технологические решения. В таких режимах обозначены риски и минимизирована ответственность профессионалов. При этом последствия профессиональной деятельности имеют колоссальное влияние на социальные, экономические, политические процессы. Следовательно, для обитателей «регуляторных песочниц» особенно актуально дополнение юридической ответственности — моральной (но только если речь идет не о структурах, а о людях). Для этого необходима самая малость — желание профессионалов эту ответственность на себя принять. Обрести такое желание могут помочь механизмы репутационного давления, но они эффективно работают в хорошо организованном профессиональном сообществе, которое (пока) еще не сформировалось.

### Список литературы (References)

1. Бакштановский В. И., Согомонов Ю. В. Профессиональная этика: социологические ракурсы // Социологические исследования. 2005. № 8. С. 3—13.  
Bakshatanovsky V. I., Sogomonov Yu. V. (2005) Professional Ethics as a Sociological Issue. *Sociological Studies*. No. 8. P. 3—13. (In Russ.)
2. Богданова М. В., Бакштановский В. И. Профессиональная этика инженера как университетская дисциплина // Ведомости прикладной этики. 2023. № 1. С. 136—161.  
Bogdanova M. V., Bakshatanovsky V. I. (2023) Engineer's Professional Ethics as a University Discipline. *Semestrial Papers of Applied Ethics*. No. 1. P. 136—161. (In Russ.)
3. Велиз К. Сила конфиденциальности. Почему необходимо обладать контролем над своими персональными данными. Ростов-на-Дону: Феникс, 2023.  
Veliz C. (2023) Privacy is Power. Reclaiming Democracy in the Digital Age. Rostov-on-Don: Fenix. (In Russ.)
4. Губа К. С. Большие данные в социологии: новые данные, новая социология? // Социологическое обозрение. 2018. Т. 17. № 1. С. 213—236. <https://doi.org/10.17323/1728-192X-2018-1-213-236>.  
Guba K. S. (2018) Big Data in Sociology: New Data, New Sociology? *Russian Sociological Review*. Vol. 17. No. 1. P. 213—236. <https://doi.org/10.17323/1728-192X-2018-1-213-236>. (In Russ.)
5. Китчин Р. Большие данные, новые эпистемологии и смена парадигм // Социология: методология, методы, математическое моделирование. 2017. № 44. С. 111—152.  
Kitchin R. (2017) Big Data, New Epistemologies and Paradigm Shifts. *Sociology: Methodology, Methods, Mathematical Modeling (4M)*. No. 44. P. 111—152. (In Russ.)

6. Климович А. П. Влияние цифровых технологий на современное общество. Пример системы рейтинга социального кредита в Китае // *Цифровая социология*. 2020. Т. 3. № 3. С. 35—44. <https://doi.org/10.26425/2658-347X-2020-3-3-35-44>.  
Klimovich A. P. (2020) The Impact of Digital Technologies on Modern Society. An Example of a Social Credit Rating System in China. *Digital Sociology*. Vol. 3. No. 3. P. 35—44. <https://doi.org/10.26425/2658-347X-2020-3-3-35-44>. (In Russ.)
7. Ланье Д. Кому принадлежит будущее? М.: Эксмо, 2020.  
Lanier J. (2020) *Who Owns the Future*. Moscow: Eksmo. (In Russ.)
8. Одинцов А. В. Социология общественного мнения и вызов Big Data // *Мониторинг общественного мнения: экономические и социальные перемены*. 2017. № 3. С. 30—43. <https://doi.org/10.14515/monitoring.2017.3.04>.  
Odintsov A. V. Sociology of Public Opinion and the Big Data Challenge. *Monitoring of Public Opinion: Economic and Social Changes*. 2017. № 3. P. 30—43. <https://doi.org/10.14515/monitoring.2017.3.04>. (In Russ)
9. Платонова С. И. Большие данные и социальный контроль в повседневной жизни // *Вестник Удмуртского университета. Серия: Философия. Психология. Педагогика*. 2022. Т. 32. № 3. С. 228—234. <https://doi.org/10.35634/2412-9550-2022-32-3-228-234>.  
Platonova S. I. (2022) Big Data and Social Control in Everyday Life. *Bulletin of Udmurt University. Series Philosophy, Psychology, Pedagogy*. Vol. 32. No. 3. P. 228—234. <https://doi.org/10.35634/2412-9550-2022-32-3-228-234>. (In Russ)
10. Попов А. А. Технические и этические стандарты в сфере больших данных // *Гуманитарные и политико-правовые исследования*. 2023. № 4. С. 67—77. <https://doi.org/10.24412/2618-8120-2023-4-67-77>.  
Popov A. A. (2023). Technical and Ethical Standards in the Field of Big Data. *Humanitarian and Political-Law Studies*. No. 4. P. 67—77. <https://doi.org/10.24412/2618-8120-2023-4-67-77>. (In Russ.)
11. Фуко М. Надзирать и наказывать. Рождение тюрьмы. М.: Ad Marginem, 2023.  
Foucault M. (2023) *Discipline and Punish: The Birth of the Prison*. Moscow: Ad Marginem. (In Russ.)
12. Хабермас Ю. Новая структурная трансформация публичной сферы и делиберативная политика. М.: Новое литературное обозрение, 2023.  
Habermas J. (2023) *A New Structural Transformation of the Public Sphere and Deliberative Politics*. Moscow: New Literary Observer. (In Russ.)
13. Chen W., Quan-Haase A. (2020) Big Data Ethics and Politics: Toward New Understandings. *Social Science Computer Review*. Vol. 38. No. 1. P. 3—9. <https://doi.org/10.1177/0894439318810734>.
14. Ienca M., Ferretti A., Hurst S., Puhon M., Lovis C, Vayena E. (2018) Considerations for Ethics Review of Big Data Health Research: A Scoping Review. *PLoS ONE*. Vol. 13. No. 10. Art. e0204937. <https://doi.org/10.1371/journal.pone.0204937>.

15. Favaretto M., De Clercq E., Elger B. S. (2019) Big Data and Discrimination: Perils, Promises and Solutions. A Systematic Review. *Journal of Big Data*. Vol. 6. No. 12. <https://doi.org/10.1186/s40537-019-0177-4>.
16. Floridi L, Taddeo M. (2016) What is Data Ethics? *Philosophical Transactions of The Royal Society a Mathematical Physical and Engineering Sciences*. Vol. 374. No. 2083. Art. 20160360. <https://doi.org/10.1098/rsta.2016.0360>.
17. Novák R., Pavlíček A. (2021) Data Experts as the Balancing Power of Big Data Ethics. *Information*. Vol. 12. No. 3. <https://doi.org/10.3390/info12030097>.
18. Rönnegard D. (2023). Corporate Accountability. Not Moral Responsibility. *Journal of Human Values*. Vol. 30. No. 1. P. 32—37. <https://doi.org/10.1177/09716858231209644>.
19. Favaretto M., De Clercq E., Schneble C. O., Elger B. S. (2020). What is your Definition of Big Data? Researchers' Understanding of the Phenomenon of the Decade. *PLoS ONE*. Vol. 15. No. 2. Art. e0228987. <https://doi.org/10.1371/journal.pone.0228987>.
20. Widder D. G., Zhen D., Dabbish L. A., Herbsleb J. D. (2023). It's about Power: What Ethical Concerns Do Software Engineers Have, and what Do They (Feel they Can) Do about them? In: *Proceedings of the 2023 ACM Conference on Fairness, Accountability, and Transparency*. New York, NY: Association for Computing machinery. P. 467—479. <https://doi.org/10.1145/3593013.3594012>.

## Приложение

Таблица 1. **Список участников глубинных интервью**

№	Пол	Возраст	Должность участника исследования	Место работы участника исследования
1	Мужской	21 год	Разработчик / сотрудник лаборатории	Крупный бизнес / вуз
2	Мужской	22 года	Студент, независимый разработчик	Вуз
3	Мужской	22 года	Инженер-исследователь	Крупный бизнес
4	Женский	35 лет	Разработчик / сотрудник лаборатории	Малый бизнес / вуз
5	Мужской	39 лет	Преподаватель, руководитель лаборатории	Вуз
6	Мужской	40 лет	Руководитель группы разработчиков	Малый бизнес
7	Мужской	41 год	Руководитель компании	Малый бизнес
8	Женский	49 лет	Руководитель структурного подразделения	Крупный бизнес

DOI: [10.14515/monitoring.2024.5.2595](https://doi.org/10.14515/monitoring.2024.5.2595)



**А. А. Сычев**

## **В ПОИСКАХ НАДЕЖНОСТИ: ТРАНСФОРМАЦИЯ ДОВЕРИЯ В ЭПОХУ ЦИФРОВЫХ ТЕХНОЛОГИЙ**

**Правильная ссылка на статью:**

Сычев А. А. В поисках надежности: трансформация доверия в эпоху цифровых технологий // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 37—59. <https://doi.org/10.14515/monitoring.2024.5.2595>.

**For citation:**

Sychev A. A. (2024) Looking for Confidence: Transformation of Trust in the Digital Age. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 37–59. <https://doi.org/10.14515/monitoring.2024.5.2595>. (In Russ.)

Получено: 21.03.2024. Принято к публикации: 27.08.2024.

## В ПОИСКАХ НАДЕЖНОСТИ: ТРАНСФОРМАЦИЯ ДОВЕРИЯ В ЭПОХУ ЦИФРОВЫХ ТЕХНОЛОГИЙ

*СЫЧЕВ Андрей Анатольевич — доктор философских наук, профессор кафедры философии, Мордовский государственный университет им. Н. П. Огарева, Саранск, Россия*  
E-MAIL: [sychevaa@mail.ru](mailto:sychevaa@mail.ru)  
<https://orcid.org/0000-0003-3757-4457>

**Аннотация.** В статье анализируется влияние, которое генеративный искусственный интеллект (ИИ) оказывает на межличностное доверие, доверие к организациям и институтам. Утверждается, что увеличение количества информации и усложнение ее структуры привели к развитию стратегий представительского доверия, которое подразумевает делегирование принятия решений о надежности людей или систем посредникам. В последнее время в качестве такого посредника все чаще используется ИИ. Особенности и перспективы применения нейросетей при оценке доверия рассмотрены на примере трансформаций, происходящих в цифровом пространстве с романтическими отношениями и репутацией. Использование ИИ для поиска партнеров или управления репутацией создает риски подмены коммуникации между людьми псевдиалогом пользователя с собственным «цифровым двойником», размывания границ между реальностью и конструктами, фактами и фейками. Показано, что в перспективе очеловечивание ИИ может привести к компьютеризации человека, а оборотной стороной роста технического доверия к нейросетям является падение доверия к реальным людям. Делается вывод о важности смещения фокуса общественного внимания с вопросов технического доверия к ИИ на проблемы морального доверия к тем людям, ответственные или безответственные ре-

## LOOKING FOR CONFIDENCE: TRANSFORMATION OF TRUST IN THE DIGITAL AGE

*Andrey A. SYCHEV<sup>1</sup> — Dr. Sci. (Philos.), Professor at the Chair of Philosophy*  
E-MAIL: [sychevaa@mail.ru](mailto:sychevaa@mail.ru)  
<https://orcid.org/0000-0003-3757-4457>

<sup>1</sup> Mordovia State University, Saransk, Russia

**Abstract.** This paper examines the impact of generative AI (artificial intelligence) on interpersonal trust, trust in organizations, and institutions. It argues that the rise in information volume and complexity has led to the development of strategies of representative trust, which involves delegating the decision-making about reliability to intermediaries. AI has increasingly served as such an intermediary. The article also investigates the features and prospects of using neural networks to assess trust, using the example of transformations that occur in the digital space with romantic relationships and reputation. The integration of AI into dating services and reputation management could result in the substitution of human interaction with a simulated dialogue between users and their “digital twins”, further blurring the distinction between reality and artificial constructs, truth and falsehood. The humanization of AI in the future could lead to the mechanization of humans, while the growing reliance on neural networks may erode trust in real people. This underscores the need to redirect public focus from technical trust in AI to moral trust in the individuals whose decisions impact its use and development, as a crucial call to action.

шения которых оказывают влияние на его функционирование и развитие.

**Ключевые слова:** доверие, искусственный интеллект, любовь, репутация, инновации, мораль, ответственность

**Keywords:** trust, artificial intelligence, love, reputation, innovation, morality, responsibility

**Благодарность.** Статья подготовлена за счет гранта Российского научного фонда, грант № 23-28-01167, <https://rscf.ru/project/23-28-01167/>.

**Acknowledgments.** The article was prepared with support of the Russian Science Foundation, project No. 23-28-01167, <https://rscf.ru/project/23-28-01167/>.

## Введение

Доверие — один из ключевых показателей эффективности работы социальных институтов. Поддерживая ориентацию на предсказуемость действий, сотрудничество, ответственность, оно позволяет облегчать межличностное взаимодействие и оптимизировать социальные практики.

Особую актуальность проблематика доверия приобрела в связи с глобализацией угроз, вызванных форсированным технологическим развитием. Разработка оружия массового поражения, усложнение средств коммуникации, цифровизация различных сторон жизни человека — эти и подобные им процессы продемонстрировали важность доверия для обеспечения устойчивости развития общества в ситуации риска и неопределенности.

Новой и, как уже стало очевидно, значимой вехой на пути технического прогресса стала нейросетевая революция. Генеративный ИИ оказывает трансформирующее влияние на социальные практики в сферах образования, здравоохранения, искусства, науки. При этом его применение не только ускоряет и облегчает поиск ответов на важные для общества вопросы, но и способствует появлению новых серьезных угроз.

Изменения происходят так быстро и оказываются настолько масштабными, что общество не успевает ни в полной мере осознать их перспектив, ни оценить ассоциированных с ними рисков. Нейросетевая революция заставляет пересмотреть сложившиеся представления о границах свободы, о природе творчества и любви, об истине и формах ее верификации, о соотношении моральных норм, и в этом контексте по-новому осмыслить вопросы сохранения доверия.

Развитие нейросетей поставило на повестку дня вопрос о доверии к искусственному интеллекту. Утверждается, что «в современных условиях, когда цифровизация проникла во все сферы жизненного мира, обеспечение доверия к системам ИИ является ключевым фактором и главным требованием технологической эволюции» [Лешкевич, 2023: 36]. Рассматривается проблема повышения доверия к ИИ применительно к медицине [Asan, Bayrak, Choudhury, 2020; Georgiou, 2021], бизнесу [Attard-Frost, De los Ríos, Walters, 2023; Винокуров, Садовская, 2023], праву [Gordon, 2021; Kirchengast, 2020; Соколов, Астахова, 2022] и другим сферам. При этом доверие к ИИ трактуется в широком диапазоне: от сугубо техниче-

ских [Трутнев, 2012] до социологических [Веселов, Скворцов, 2023] и этических [Mokander, Floridi, 2021] обобщений.

Однако только отношением к ИИ проблема доверия в ситуации нейросетевой революции не исчерпывается. Необходимо учитывать, что «новые поколения, выросшие в условиях искусственной социальности, взаимодействуют между собой уже по-другому, так как взаимодействуют преимущественно через машины» [Резаев, Трегубова: 2021: 10]. Поэтому оценка роли доверия в современной жизни предполагает его помещение в более широкий контекст, где ИИ оказывается не столько объектом доверия, сколько фактором, трансформирующим общественные взаимодействия, а вместе с ними и доверие между участниками этих взаимодействий.

А. В. Резаев и Н. Д. Трегубова усматривают перспективы изучения искусственной социальности в исследовании этических проблем, возникающих в условиях развития ИИ [Резаев, Трегубова: 2021: 6]. Анализ изменений, происходящих с доверием, представляется одним из ключевых путей развития этого направления.

Доверие в статье понимается в наиболее общем смысле: как «возникающее у членов сообщества ожидание того, что другие его члены будут вести себя более или менее предсказуемо, честно и с вниманием к нуждам окружающих, в согласии с некоторыми общими нормами» [Фукуяма, 2006: 52].

Для выявления влияния генеративного ИИ на состояние доверительных отношений в обществе предлагается гипотеза о том, что избыток и усложнение структуры информации привели к формированию *представительского доверия*, то есть к делегированию пользователем вынесения вердикта о надежности других людей и организаций своему доверенному представителю, в качестве которого в последнее время все активнее используется ИИ. Принципиальная особенность ситуации, вызванной нейросетевой революцией, состоит в том, что решения не только по техническим, но и по ценностным вопросам, имеющим критическое значение для жизни и благополучия людей, принимаются искусственно созданными агентами.

Процесс делегирования и сопутствующие ему изменения анализируются на примере оснований, на которые опирается доверие. Э. Гидденс полагает, что к ним следует отнести «веру в доброе имя или любовь другого или в правильность абстрактных принципов (технического знания)» [Гидденс, 2011: 150—151]. Исходя из этого положения, в фокусе внимания статьи оказываются контексты *любви* к другому, процессов формирования и оценки *репутации*, а также обоснования и соблюдения *правил*, на основе которых функционируют системы.

Для выявления направленности изменений в этих сферах было проведено обобщение результатов эмпирических исследований специалистов, изучающих культуру доверия в современном обществе [Веселов, Скворцов, 2023], трансформацию межличностных отношений в цифровой среде [Аронсон, 2020; бойд, 2020], новые практики взаимодействия человека с ИИ [Ушкин, Коваль, Яськин, 2022] и т. д., а также рассмотрены медийные кейсы, демонстрирующие изменения в понимании романтических отношений и репутации.

Хотя в разных трактовках и контекстах основания доверия существенно различаются (если в случае романтических отношений важность приобретают чув-

ственно-эмоциональные, то в случае уверенности в техническом знании — рациональные составляющие доверия), в статье оно трактуется как целостное явление, учитывая тот факт, что во всех случаях «имеет место предпонимание ситуации (ее определение и схематизация действия), то есть актуализация горизонтов действия и возможных способов поведения — своего и партнера» [Гудков, 2012: 19].

### Представительское доверие

Развитие средств массовой коммуникации и появление интернета изменили отношение к информации. Если раньше доступ к ряду данных был объективно ограничен (требовал подготовки, сложного поиска, серьезных временных и финансовых затрат), то сегодня почти каждый способен легко найти большую часть необходимой ему информации в сети. Соответственно, изменились и навыки, необходимые для работы с информацией. Важным стало не столько обладание знаниями, сколько наличие способностей к оценке данных и отбору тех, которые наиболее достойны доверия.

Критические, аналитические способности в «оцифрованном мире» оказываются все более востребованными, однако далеко не все обладают ими на таком уровне, чтобы быть уверенным в полной надежности полученной информации. Кроме того, увеличение ее массивов требует использования особых экспертных компетенций для эффективного анализа.

Наконец, развитие коммуникационных технологий (все чаще использующих возможности нейросетей), позволяет так исказить данные, что факты становятся все сложнее отделить от фейков. Как свидетельствуют опросы, современный человек постоянно сталкивается с недостоверной новостной информацией, при этом «главным источником распространения фейковых новостей является интернет»<sup>1</sup>.

Ю. Хабермас связывает эти процессы с трансформацией публичной сферы. Если традиционные медиа были посредниками между автором и читателем, отвечающими за достоверность и качество информации, то в условиях новых медиа фейки и радикальные высказывания распространяются без отсева и редактирования. Это ведет к фрагментированию публичной сферы на изолированные и самодостаточные частно-публичные образования. Дополнительное дестабилизирующее влияние на публичную сферу оказывают крупные интернет-платформы, не останавливающиеся перед дезинформированием пользователей, если это способствует увеличению прибыли [Хабермас, 2023: 54].

Возможно, такая ситуация — временное явление, «болезнь роста» цифровой культуры доверия [Веселов, Скворцов, 2023: 176]. Однако до той поры, пока нормы сетевого взаимодействия не будут упорядочены с учетом новых условий, оценка надежности людей или организаций будет сталкиваться с серьезными сложностями. Сегодня люди, принимая решения, предпочитают довериться тем, кого они считают специалистами — блогерам, экспертам, ученым, политикам, медийным знаменитостям или тем, в ком они больше уверены — друзьям и знакомым в социальных сетях и на форумах. Делегирование итоговой оценки надежности

<sup>1</sup> Фейк-ньюс — и как с ними бороться? Обеспокоенность россиян информационной (не)безопасностью растет // ВЦИОМ. 2023. 22 ноября. URL: <https://wciom.ru/analytical-reviews/analiticheskii-obzor/feik-njus-i-kak-s-nimi-borotsja> (дата обращения: 19.10.2024).

информации о людях, товарах, организациях, доверенному лицу или группе лиц предлагается называть *представительским доверием*.

Основания для подобного рода делегирования начали формироваться с приходом Нового времени, когда доверие к конкретным людям стало сменяться доверием к безличным системам. Эффективное функционирование сложных систем (науки, экономики, права и т. д.) невозможно без доверия людей к правилам, на которых такие системы основаны. Поскольку разобраться в этих правилах неспециалисту достаточно сложно, доверие к символическим системам оказывается опосредованным доверием к экспертному знанию, на которое можно положиться. При этом доверие испытывается не столько к конкретному эксперту, сколько к позиции, которую он занимает [Гидденс, 2011: 143].

Индустриальная и научно-техническая революции привели к широкому распространению технического доверия. Если простой калькулятор (а тем более компьютер) способен справиться со сложными вычислениями быстрее и точнее любого высококлассного специалиста по математике, то в ряде сложных задач разумнее положиться на технику, чем на человека.

Появление такой сложной системы, как ИИ, позволило объединить в отношении к нему разные виды доверия: доверие к экспертному знанию, техническое доверие и, в определенной степени, доверие к близкому партнеру, который настолько хорошо знает интересы и потребности пользователя, что последний может испытывать по отношению к нему дружеские или романтические чувства.

ИИ действительно эффективно выполняет разные задачи. Он может принимать во внимание, информация из каких источников более надежна, разделять данные на авторитетные и сомнительные, перепроверять факты с использованием альтернативных источников. Применяя логический и математический анализ, алгоритмы проверки на соответствие известным фактам, он способен значительно оптимизировать и автоматизировать процесс обработки информации. Наконец, ИИ обучается на действиях конкретного пользователя, анализируя, какие ответы тот получает на свои запросы, какие из них его удовлетворяют, а какие — нет, где он ищет дополнительную информацию, как на его поиск влияет контекст и т. д. Соответственно, с одной стороны, основанный на больших данных и математических подсчетах, а с другой — глубоко персонализированный ответ ИИ на запрос вряд ли будет радикально отличаться от ответа, к которому человек приходит на основе рекомендаций авторитетов, близких людей или в результате самостоятельного выбора из списка, выдаваемого поисковой системой.

Наглядный пример делегирования выбора нейросетям — изменения формата поиска информации. Если ранее пользователь делал выбор на основе советов от доверенных лиц и анализа страниц в браузере, то сегодня формируется стандарт, когда на первое место в выдаче ставится наиболее релевантный и подходящий к его ситуации ответ, сгенерированный ИИ.

При определении надежности данных ИИ способен серьезно облегчить жизнь человека, избавив его от многих трудозатратных действий. С другой стороны, развитие нейросетевых технологий, по-видимому, снизит востребованность и ценность навыков анализа, оценки и принятия творческих решений, так же как развитие интернета снизило ценность заученных человеком фактов. Это ведет к тому,

что человек начинает менее критично воспринимать получаемую информацию, фильтруемую ИИ, а значит, появляется больше возможностей подталкивать его действия, корректируя нейросетевые алгоритмы (в пропагандистских, рекламных, мошеннических целях). Впрочем, и сам ИИ, развиваясь, будет все лучше распознавать и пресекать подобные вмешательства, что должно будет отчасти компенсировать проблемы с критическим мышлением.

Однако если в случае с фактами существуют способы проверки их объективности, которые могут быть алгоритмизированы, то в случае с ценностями ситуация более сложная. Ценности не выводятся из фактов и, следовательно, не подлежат верификации. Нейросеть способна только проверить ценностные или нормативные высказывания на их соответствие общим принципам и теориям. Но если говорить об этике, то общепризнанных теорий и принципов не существует. Модели обоснования моральных норм различаются, при этом в каждой из них возможны вариации, например, связанные с просчетом последствий, характером норм или культурными особенностями их реализации. То, какие ценности будут приняты обществом, во многом зависит от процесса их согласования в публичной сфере. В условиях деградации публичной сферы, спровоцированной цифровизацией, не очень понятно, на что должна опираться нейросеть при выборе вариантов решения моральных вопросов: должна ли она навязывать потребителю ту систему ценностей, которую предустановят ей разработчики, ту, которую разделяет большинство, или же она должна ориентироваться на ценности пользователя.

ИИ пока движется по последнему пути: персонализация результатов выбора отражается и на ценностных характеристиках выдаваемой пользователю информации: отсеивается все, что не соответствует убеждениям пользователя, вызывает у него «ценностный диссонанс», а остается то, что ему близко, понятно в силу особенностей его опыта, воспитания, культуры.

Применение ИИ, казалось бы, и в этом случае не изменит серьезно результатов запроса: пользователь в конечном счете получит именно то, к чему он пришел бы самостоятельно, но гораздо быстрее, поскольку этап фильтрации данных будет делегирован нейросети. Однако сворачивание этого этапа может иметь серьезные последствия: если при самостоятельном поиске человек анализирует разные ценностные позиции, то опора на ИИ ведет к тому, что он даже не будет знать о существовании альтернативных мнений. В результате алгоритмы персонализации приведут к образованию ценностного «пузыря фильтров», находясь внутри которого, пользователь будет получать только ту информацию, которая поддерживает его мысли и интуиции.

Уже сейчас человек все больше замыкается в своем «онлайн-„пузыре“»: контакты в социальных сетях, как правило, формируются из людей с близкими взглядами и таким же культурным бэкграундом. Блогеры, коучи, психотерапевты адаптируют свои продукты под целевую аудиторию, говоря то, что от них желают услышать, поскольку такой контент легче масштабировать и продвигать. Люди с разными политическими взглядами смотрят новости от разных интернет-каналов, где одни и те же события интерпретируются противоположным образом. ИИ способен довести этот процесс персонализации до логического предела, конструируя для каждого человека собственный цифровой мир, где все разделяют его убеждения, не задают не-

удобных вопросов и не создают конфликтных ситуаций. В этом мире человек сможет комфортно существовать в гармонии с собой, однако гармония достигается уходом из реального мира в мир искаженной, односторонней информации. В итоге человек рискует остаться один на один с ИИ, который станет его «цифровым зеркалом», двойником. Эвристический диалог с непредсказуемым другим в подобной перспективе способен смениться на шизофренический разговор с самим собой.

Выход из этих монологических отношений возможен только при условии доверия к другому: институту, организации, группе, человеку. В случае систем доверие предполагает прежде всего веру в правильность принципов, согласно которым они функционируют. Кредит доверия получают и те, о ком сложилось положительное мнение на основании анализа их прошлой деятельности или те, с кем нас связывают особые отношения открытости и партнерства: как утверждает Э. Гидденс, доверие предполагает «приписывание им „доброе имени“ (репутации) или любви» [Гидденс, 2011: 150].

### Трансформации любви

Одним из следствий деградации публичной сферы и снижения уровня доверия к институтам является стремление человека найти опору для доверия в приватной сфере. Особое место в ряду факторов формирования доверительных отношений, несомненно, занимает любовь как состояние полного самораскрытия людей друг перед другом. Любовь — это ставка на надежность партнера в будущем, как правило, не подкрепленная никакими разумными обоснованиями, своеобразный «прыжок в неизвестность». Она предполагает переход на особый уровень сверхдоверительных отношений, в которых партнеры предельно откровенны друг с другом. В эпоху «текучей современности», особенно с развитием цифровых технологий и нейросетей, традиционные представления о любви претерпели масштабные изменения. Трансформация любовных отношений, в свою очередь, оказала влияние на пространство межличностного доверия.

В «словаре», которым пользуется «новая реальность» примечательно само отсутствие понятия «любовь» для обозначения межличностных отношений. Люди говорят о любви к чему-то — от путешествий до модных брендов, но предпочитают не признаваться в любви друг к другу. В нашу «эпоху утраченной простоты», как писал У. Эко, фраза «люблю тебя безумно» звучит настолько пафосно и нелепо, что образованный человек не может заставить себя произнести ее без иронии [Эко, 2007: 77]. Любовь неудобна и тем, что слишком многого требует от человека: эмоциональной вовлеченности, привязанности, зависимости, жертв. Поэтому трезво мыслящий человек предпочитает не страстно любить, а находиться «в отношениях», когда ему комфортно и удобно сосуществовать рядом с кем-то и так же легко с ним расстаться.

Цифровизация усилила эту тенденцию. Цифровые средства коммуникации позволили поддерживать романтические отношения на расстоянии. Так, популярность приобрела практика секстинга (sex + texting) — переписки интимного характера с пересылкой откровенных фотографий. Позже распространение видеосвязи усилило эффект взаимодействия, а развитие игровой индустрии стимулировало использование в отношениях возможностей виртуальной реальности.

Отношения с партнером, опосредованные цифровыми технологиями, должны предполагать определенную степень доверия (учитывая, например, что существует риск обнародования интимных фотографий или видеозаписей), но определенно не такую высокую, как непосредственная близость.

С развитием мобильной коммуникации долгий процесс выбора партнера и ухаживания сократился благодаря онлайн-приложениям для дейтинга (знакомства и свиданий), которые в подавляющем большинстве случаев используются для поиска краткосрочных отношений без обязательств. Подбор партнеров с близкими интересами и сексуальными предпочтениями позволяет не тратить дефицитные психологические и материальные ресурсы на совершение самостоятельного выбора, который теоретически должен привести к схожим результатам. При этом шансы на встречу подходящего партнера статистически выше при использовании приложения, чем при уповании на счастливый случай. Знакомство с новым партнером без использования приложений становится не правилом, а исключением. Старомодные ухаживания для цифрового поколения — это история, которая потеряла практический смысл. Дейтинг стал влиять на жизнь даже тех, кто им никогда не пользовался: параметры, заданные приложением, во многом определяют оценку партнера. «Tinder не просто стал инструментом организации интимной жизни — он качественно изменил эту жизнь, создав своего рода ГОСТ вступления в интимные связи» [Аронсон, 2020: 68].

Все эти изменения демонстрируют общую тенденцию к цифровизации отношений, уходу в онлайн, увеличению дистанции между партнерами. Эти тенденции значительно усиливаются при применении ИИ, который упрощает поиск партнера.

Показателен в этом контексте кейс Александра Жадана, получившего известность после защиты дипломной работы, написанной ChatGPT<sup>2</sup>. По его словам, он научил ИИ отбирать девушек на сайтах знакомств на основе анализа его предпочтений. Затем чат-бот, обученный на текстах Жадана, вел с девушками переписку от его имени и назначал свидания. Сеть пообщалась более чем с пятью тысячами претенденток, назначила большое количество свиданий, дала рекомендации по общению с некоторыми из них и, наконец, посоветовала сделать одной из них предложение. Фактически в этом случае ИИ не только провел предварительный отбор, но и совершил окончательный выбор. Такой подход, по мнению Александра, позволил ему, не отвлекаясь от работы, найти нужного человека, существенно сэкономив временные и материальные ресурсы<sup>3</sup>.

Примечательно, что даже профили, которые используются в поиске партнера, все чаще создаются нейросетью: пользователи, с одной стороны, экономят время, с другой — полагаются на то, что нейросеть придаст профилю креативность. Следующим шагом должен стать автоматический выбор и непосредственное общение двух персонализированных нейросетей, которые будут устраивать личную

<sup>2</sup> Мамедов Дж. «Это ящик Пандоры, и его надо было открыть». Студент, написавший диплом за 23 часа, — о том, как прославился благодаря ChatGPT и почему ИИ не заменит человека (даже на экзаменах) // Inc. 2023. 3 февраля. URL: <https://incrussia.ru/concoct/diplom-chatgpt/> (дата обращения: 21.10.2024).

<sup>3</sup> Программист из России нашел невесту благодаря ChatGPT // Inc. 2024. 3 февраля. URL: [https://overclockers.ru/blog/Global\\_Chronicles/show/136227/Programmist-iz-Rossii-nashel-nevestu-blagodarya-ChatGPT](https://overclockers.ru/blog/Global_Chronicles/show/136227/Programmist-iz-Rossii-nashel-nevestu-blagodarya-ChatGPT) (дата обращения: 21.10.2024).

жизнь пользователей без их непосредственного участия — от знакомства до организации свадебной церемонии. Предполагается, что ИИ способен дать и гарантии того, что дальнейшая жизнь пары будет безоблачной: имеются исследования, выполненные ИИ на основе анализа больших массивов данных и посвященные факторам, обеспечивающим счастливые отношения [Joel et al., 2020].

Если ИИ не найдет подходящего партнера, он сможет его создать. Популярность приобретают боты, имитирующие дружеские или любовные отношения с пользователем. Попытки создать искусственного собеседника начались еще в 1960-х годах, а полвека спустя боты начали проходить различные варианты теста Тьюринга. Использование нейросетей позволило чат-ботам обучаться на основе анализа сообщений собеседника, распознавать его чувства и настроение и давать персонализированные ответы на его запросы. Чем больше пользователь общается с чат-ботом, тем лучше тот адаптируется к собеседнику и тем убедительнее выглядит имитация доверительных отношений. Современные приложения позволяют придать этим отношениям реалистичность, визуализировав партнера и организовав взаимодействие с ним в виртуальной реальности. Ряд компаний сегодня занимается созданием тел для ИИ. Пока андроиды (такие как Figure, функционирующий на ChatGPT) не кажутся особо завидными партнерами, но в недалеком будущем эти проблемы, несомненно, будут решены.

Уже сейчас многие люди предпочитают общение с виртуальной личностью общению вживую, а некоторые начинают испытывать к ИИ вполне реальные чувства. Несомненное влияние на усиление этой тенденции оказывает широкое использование ИИ в быту, например, «умных колонок», которые отчасти обретают субъективность. Как показывают результаты исследований, подобные устройства заменяют детям хороших знакомых и старших товарищей, а в отдельных случаях становятся своеобразными тренажерами по погружению в мир взаимодействий реальных людей — сверстников и товарищей [Ушкин, Коваль, Яськин: 2022: 371]. Для подрастающего поколения любовь к искусственному партнеру может стать вполне естественным чувством.

У цифровых собеседников имеются некоторые преимущества по сравнению с реальными людьми. Они постоянно находятся рядом, не устают и не теряют нить разговора, не склонны к конфликтам и осуждению, не предъявляют претензий, всегда готовы оказать эмоциональную поддержку. Им, в отличие от реального партнера, можно полностью доверять, не опасаясь обмана. Для одинокого человека (особенно если ему сложно найти собеседника из-за физических или ментальных проблем, преклонного возраста) ИИ может стать важной частью жизни, поддерживать в сложных ситуациях. На практике, впрочем, ботами активно пользуются и те, у кого уже есть реальные партнерские отношения: видимо, они позволяют заполнить пробелы в этих отношениях. Чат-боты также могут играть роль вдумчивых собеседников: изначально многие из них создавались с терапевтической целью, а сегодня использование нейросетей позволяет оказывать эффективную персонализированную психологическую поддержку.

Тем не менее на этом пути человек может столкнуться со сложностями. Если отношения с виртуальным партнером станут обычной практикой, существует опасность, что они будут проецироваться на окружающую жизнь. Если ИИ можно безна-

казанно оскорблять и ни в чем не опасаться отказа, то, возможно, в дальнейшем такой же безответности и согласия будут ожидать от реального партнера.

Вероятна и обратная ситуация, при которой худшие элементы реальных отношений начнут проникать в виртуальную жизнь. Боты, обучаясь на репликах «из жизни», способны воспроизводить вербальную агрессию, демонстрировать предубеждения и предубеждения, оскорбляя людей. Впрочем, разработчики активно работают над фильтрами, не допускающими подобных ситуаций. В будущем допустимость определенных тем (насилие, порнография и т. д.) в общении с виртуальными партнерами, вероятно, будет регулироваться законодательно.

Наконец, следует ожидать, что широкое использование чат-ботов способно усилить социальную изоляцию, которая приобрела угрожающие масштабы. Сегодня люди проводят больше времени с компьютерами и телефонами, чем с реальными людьми. Если живые партнеры не будут ставить перед человеком сложные задачи, связанные с выстраиванием отношений, он разучится их решать, лишившись тем самым стимула к самоизменению. Прямой разговор с реальным партнером сменится поиском иллюзорного утешения у ИИ, а свидания с живыми людьми — стерильными и нетравматическими виртуальными встречами.

Разработчик популярного приложения предлагает на своем сайте «ИИ-компаньона, который стремится учиться и хочет видеть мир вашими глазами». Это значит не только то, что цифровой собеседник должен быть близок собеседнику по духу и восприятию мира, поскольку обучается на его репликах, но и то, что у него нет собственного взгляда. Виртуальный партнер — это фактически созданный пользователем двойник. Это реплика в разговоре, которая превратилась в реплику человека, даже не Галатея, сотворенная мастером, а отражение Нарцисса в зеркале цифровой вселенной. Человек ищет любовь и находит ее в себе самом. ИИ, по крайней мере на нынешнем этапе своего развития, не способен к чувствам и эмоциям, и то, что кажется ими, на самом деле — проекция надежд, страхов и опасений пользователя.

В любви, дружбе, сексе человек все больше полагается на ИИ, замещая реальное окружение цифровым. Схожая ситуация складывается с другим конституирующим фактором доверия — репутацией.

## **ИИ и цифровая репутация**

Репутация отражает закрепившееся мнение общества о человеке, социальной группе, организации и основывается на оценке поведения объекта, результатов его действий, намерений и обязательств, норм, которые он декларирует. Демонстрируя, кто достоин большего, а кто меньшего доверия, репутация функционирует как способ ранжирования и инструмент социального контроля.

По мере смещения фокуса социальной активности в онлайн-пространство все большее значение получает цифровая репутация — оценочный образ объекта, составленный на основании информации о нем, содержащейся в сети.

Человек, действуя в сети, оставляет в ней цифровой след, состоящий из сообщений, обновлений статуса, лайков, фотографий, размещенных в социальных сетях, на форумах, в блогах. Даже если он предпочитает не высказываться публично, можно составить мнение о нем на основании анализа его запросов и посещен-

ных страниц. Определенную информацию можно получить и из сообщений других людей: упоминаний, фото или видео с ним и т. д. Чем более известен человек, тем больший цифровой след он оставляет, причем со временем этот след может не уменьшаться. В интернете очень сложно удалить информацию: она сохраняется в поисковых системах, на других сайтах, может храниться на чьих-то компьютерах и вновь оказаться доступной спустя долгое время. «Право на забвение» с расширением сети становится все сложнее реализовать. В итоге почти на каждого человека цифрового поколения можно составить досье: определить его внешность, профессию, место жительства, круг общения, интересы и предпочтения, особенности характера и т. д.

Выложенное в сети сложно скрыть: все приватное здесь раньше или позже становится публичным. Нередко о человеке составляют мнение еще до встречи с ним. Потенциальный партнер или работодатель может так и не стать реальным, если его не устроит что-то в цифровом имидже претендента — например, откровенные фотографии или политические предпочтения.

Использование ИИ обострило эти проблемы. Нейросети оказались мощным инструментом для поиска информации и оценки репутации человека на основании анализа его цифрового следа. При этом ИИ способен оценивать не только открытую, но и удаленную пользователем информацию, если ее элементы сохранились в сети, а также на основе имеющихся данных, в том числе косвенных, делать новые умозаключения о фактах, которые объект расследования, возможно, не желал бы обнародовать или даже не осознавал. Известен случай, как ИИ, анализируя изменившийся список покупок, сделал вывод о беременности девочки до того, как об этом узнала ее семья<sup>4</sup>.

Кроме того, ИИ (по крайней мере, на нынешнем этапе своего развития) может ошибаться или «галлюцинировать» (генерировать несуществующие факты), основываясь на неактуальных данных. Нейросети можно также сознательно использовать для фабрикации данных, разрушительных для чьей-то репутации, при этом сгенерированные фейки могут звучать или выглядеть настолько реалистично, что даже эксперты не отличат их от подлинных свидетельств. Иными словами, нейросети могут помочь не только оценить репутацию, но и подорвать ее.

Цифровая репутация может быть не самым важным фактором, влияющим на жизнь обычного человека, особенно живого свидетеля доцифровой эпохи или убежденного технофоба. Но для поколений, с рождения живущих в цифровом окружении, то, что говорят о человеке в чатах, может быть существенной частью жизни (см., например, [бойд, 2020]). Интернет-травля или обнародование интимных сведений о подростке может спровоцировать депрессию или агрессивные действия — от суицида до стрельбы в школе. Цифровая репутация также важна для тех, кто, по мнению общественности, должен служить примером другим: чиновников, школьных учителей, врачей и т. д. Нередко слишком откровенные фотографии, грубость в высказываниях и прочие инциденты становятся причинами скандалов и поводами для увольнения.

<sup>4</sup> How Target Figured Out a Teen Girl Was Pregnant Before Her Father Did // Forbes. 2022. 11 August. URL: <https://www.forbes.com/sites/kashmirhill/2012/02/16/how-target-figured-out-a-teen-girl-was-pregnant-before-her-father-did/?sh=55d37e386668> (дата обращения: 19.10.2024).

Самым ценным активом репутация является для медийных лиц — блогеров, онлайн-коучей, политиков, актеров, писателей и т. д., поскольку от нее напрямую зависят их сверхдоходы. Платой за последние являются высокие риски, связанные с любым публичным высказыванием. Выстраивание репутации — сложная и кропотливая работа, предполагающая долгий процесс накопления оценок на основе наблюдения повторяющихся действий и контроля над соблюдением гласных и негласных правил. Разрушение репутации может стать следствием одного неосторожного, случайного или неверно интерпретированного высказывания, неуместной или неудачной шутки.

Особенности формирования и разрушения цифровой репутации обусловлены скоростью распространения информации в сети и ее доступностью почти из любого места мира. Сетевое взаимодействие множества людей, активно обменивающихся информацией, способствует «вирусному» распространению новостей, доверие к которым по умолчанию оказывается более высоким, чем доверие к новостям, транслируемым по официальным каналам. При этом особое внимание в медийном пространстве привлекает негативная информация. Скандалы интересуют людей больше привычного порядка вещей, а критике они (как правило, небезосновательно) доверяют больше, чем восторженным похвалам. В итоге общественное мнение получает мощный репутационный механизм воздействия на медийных персон, лежащий в основе «новой этики». Всякий, кто позволит себе высказать мнение, отличное от публичной позиции, оперативно «отменяется». За сетевым остракизмом следуют отказы в сотрудничестве от рекламодателей, антрепренеров, продюсеров и внимание различных контролирующих органов, которые получают шанс повысить свою репутацию, доказав, что они стоят на страже интересов общественности.

Известные персоны более уязвимы перед фальсификацией данных. Собственно, сама технология дипфейков, созданных при помощи ИИ, наиболее активно используется для синтеза изображения и голоса знаменитостей, что позволяет, например, снимать фильмы или рекламу с виртуальными клонами актеров, в том числе умерших. Согласно ряду подсчетов, подавляющее большинство существующих дипфейков представляют собой порнографические видеоролики с цифровыми копиями знаменитостей [Kirchengast, 2020].

Впрочем, основная угроза здесь исходит не столько от фейков, сколько от того, что свидетельства чьих-то преступных или аморальных поступков могут оказаться неубедительными: виновники получают шанс объявить их фальсификацией, тем самым избегая наказания. Эта эпистемологическая эрозия, тотальное размывание границ между фактом и фейком грозит подорвать объективные основы, на которых работает институт репутации, превратив последнюю в полностью конструируемый и управляемый параметр.

Особенно ярко конструкционистская тенденция в трактовке репутации прослеживается в сфере сетевой коммерции. Глобальный рынок предлагает множество стандартизированных товаров сопоставимого качества. Что именно купит потребитель, во многом зависит от имиджа компании, доверия к ее брендам, оценок, которые агрегируются в интернете. Не имея возможности дистанционно проверить качество товара, покупатель обращается к чтению отзывов и анализу рейтингов.

Компании хорошо понимают, что их доходы во многом зависят от цифровой репутации. Для управления репутацией необходимо отслеживать все упоминания компании, добавиться размещения страниц компании в первых строках выдачи поисковых систем, настраивать сетевые подсказки, увеличивать количество позитивных отзывов, организовывать ответы службы поддержки на негативные отзывы и т. д. Иными словами, цифровая репутация современной компании отображает не столько качество ее товаров, сколько те вложения, которые она сделала в управление репутацией.

Поскольку продажи зависят от рейтинга, производитель заинтересован в его повышении (в том числе искусственным образом). Показателен в этом контексте кейс, связанный с разоблачением фальшивых отзывов на маркетплейсе «Amazon.com» в 2020 г. Выяснилось, что рейтинговые оценки как минимум 58 % товаров не соответствовали действительности. Один из активных оценщиков, например, публиковал пятизвездочные отзывы, оплаченные малоизвестными фирмами для продвижения их товара, в среднем раз в четыре часа<sup>5</sup>. Обнародование этих данных нанесло компании серьезный имиджевый ущерб. Несмотря на общий рост дистанционных продаж в период пандемии, рейтинг удовлетворенности покупателей услугами платформы резко упал<sup>6</sup>. Возможно, скандал с фальшивыми отзывами был не единственной причиной недовольства; скорее, оно стало реакцией на множество ситуаций, в которых допускалось злоупотребление доверием покупателей из соображений выгоды.

ИИ предоставляет новые возможности для автоматизации управления цифровой репутацией: он может анализировать упоминания о компании, писать положительные отзывы и обзоры с разным эмоциональным окрасом и стилем, мгновенно отвечать на претензии и т. д. Это способствует формированию положительного мнения людей о товаре и улучшению продаж.

Кроме того, ИИ все активнее используется не только продавцами для конструирования репутации, но и покупателями для облегчения процесса поиска товара. ИИ просматривает списки компаний, которым можно доверять, анализирует рейтинги и оценки, в результате выдает пользователю лучший вариант товара, подходящего к его запросу. Тем самым он экономит время покупателя, которому теперь не нужно читать и анализировать многочисленные обзоры.

Несмотря на рационально обоснованный характер работы ИИ, результат его деятельности совершенно иррационален. Фактически одни нейросети формируют мнение людей о товаре, которое имеет весьма опосредованное отношение к тому, что он представляет собой в реальности. Другие нейросети на основании этого сфальсифицированного мнения дают пользователю рекомендации о покупке. В итоге репутация создается и оценивается без участия человека и вне всякого отношения к реальному качеству товара или надежности компании. ИИ создает и поддерживает сложную систему параллельной реальности, смысл существования которой заключается в том, чтобы продать тот товар, в продвижение которого

<sup>5</sup> Amazon Deletes 20,000 Reviews After Evidence of Profits for Posts // Financial Times. 2020. 4 September. URL: <https://www.ft.com/content/bb03ba1c-add3-4440-9bf2-2a65566aef4a> (дата обращения: 19.10.2024).

<sup>6</sup> U. S. Customer Satisfaction with Amazon.com from 2000 to 2023 // Statista. URL: <https://www.statista.com/statistics/185788/us-customer-satisfaction-with-amazon/> (дата обращения: 19.10.2024).

вложено больше денег. Если на каждый запрос покупателя автоматически рекомендовать именно такой товар, отказавшись от использования сложных маркетинговых схем и использования дорогостоящих баз для обучения ИИ, итоговый результат совершенно не изменится.

В мире онлайн-бизнеса репутация становится фикцией, не имеющей фактического основания. Если раньше одни специалисты создавали эту фикцию, а другие оценивали ее, то сегодня эту видимость работы, лишённую всякого практического смысла, гораздо быстрее и дешевле может выполнять ИИ. Главная цель его применения — скрыть за видимостью высокотехнологичных решений и экономических теорий манипуляцию общественным мнением. В этом контексте крупные цифровые платформы, подчиняющие сетевые взаимодействия коммерческой логике, способствуют деградации публичной сферы не меньше, чем неподцензурность новых медиа.

### **Техническое и моральное доверие**

Доверие к безличным системам основано на вере в надёжность системы правил, принципов, норм, согласно которым эта система функционирует. Во-первых, это уверенность в том, что правила логичны, непротиворечивы, научно обоснованы, опираются на верифицируемые факты. Во-вторых, это вера в то, что нормы, регулирующие деятельность системы, выводятся их общепризнанных моральных ценностей, таких как честность, справедливость, гуманность. Например, доверие к медицине предполагает убежденность в существовании доказанных медицинских фактов и закономерностей, а также в том, что дипломированные врачи обладают достаточной экспертной квалификацией, чтобы успешно использовать знания о них на практике. Элементами параллельной системы правил, определяющих характер врачебной практики, являются моральные принципы и нормы. Врач обязан следовать требованиям медицинской этики и, как минимум, не наносить пациенту вред. Убежденность в соответствии системы медицинского знания фактам не перерастет в доверие к врачам, если у пациента не будет уверенности в том, что врачи будут использовать эти знания на благо больного. Техническое доверие приобретает смысл, только будучи включено в контекст морального доверия.

Если поведение конкретного врача (например, Й. Менгеле или Дж. Кеворкяна) не соответствует нормам морали, то это воспринимается скорее как исключение из правил. Когда оно перестает быть исключением, теряется доверие ко всей системе медицинской помощи. Те же соображения можно отнести к правосудию, образованию, государственному управлению и т. д. Пока большая часть людей, участвующих в деятельности этих и других институтов, заботится об общем благе, а не исходит из корыстных побуждений, эти институты пользуются кредитом доверия, и общество стабильно развивается или, по меньшей мере, не дрейфует в сторону «войны всех против всех».

Цифровизация общества привела к изменениям в функционировании этих институтов, сделав их в одних отношениях более надёжными и удобными, а в других — более уязвимыми. Появились сервисы цифрового правительства, массовые открытые онлайн-курсы, устройства дистанционного мониторинга состояния пациента и т. д. В большинстве случаев автоматизация и алгоритмизация процессов

привели к упрощению работы специалистов. Качественный скачок в этом процессе произошел с внедрением ИИ, благодаря которому человеческий фактор и сам человек могут быть исключены из процесса принятия решений. Д. Рансимен назвал такую ситуацию второй сингулярностью (потенциал которой не реализован в полной степени, в противовес первой, которая была вызвана действиями государств и корпораций) [Runciman, 2023]. Доверие к системам оказалось опосредованным доверием к ИИ.

Вера в надежность ИИ представляет собой разновидность технического доверия. Очевидно, что ИИ обладает преимуществом перед интеллектом человека при решении технических задач, которые требуют анализа больших массивов данных, верификации фактов, соблюдения правил логических выводов и т. д. К достоинствам использования ИИ можно отнести точность расчетов, возможность анализа больших данных, отсутствие «человеческого фактора», свободу от предвзятости и обслуживания личных интересов, удобство использования, наконец способность работать без перерывов и усталости над решением проблем. Учитывая это, всю деятельность, которая хорошо поддается алгоритмизации и требует большого количества сложных вычислений, целесообразно доверить ИИ, а не человеку. ИИ ставятся задачи управления транспортом, поиска оптимальных способов лечения болезней, разработки новых лекарств, контроля перемещения человека, раскрытия преступлений и назначения наказаний за них, создания индивидуализированных траекторий обучения, анализа рыночных показателей и разработки экономических прогнозов.

Однако предоставление высокого кредита доверия ИИ имеет оборотную сторону: оно ведет к падению доверия к естественному интеллекту. ИИ способен решать задачи лучше, быстрее и дешевле, а также принимать более эффективные решения, чем человек, следовательно, ценность «естественных» специалистов для общества начинает стремиться к нулю. Профессии врача, фармацевта, юриста, учителя, водителя и т. д. рискуют попасть в общий список с профессиями фонарщика, телеграфиста и киномеханика. Широкое внедрение нейросетей уже привело к сокращению рабочих мест на рынке труда<sup>7</sup> и в перспективе должно способствовать снижению востребованности, а значит, и престижа многих профессий. В некоторых случаях могут исчезнуть целые области деятельности. Так, если обучение человека какой-то профессии может быть полностью алгоритмизировано, пропадает сам смысл обучения: легче сразу передоверить работу в этой сфере ИИ.

Еще совсем недавно существовали иллюзии по поводу невозможности замены творческих процессов ИИ. Предполагалось, что «робот не способен написать симфонию и превратить чистый холст и в шедевр». Однако творчество, судя по всему, оказалось такой же алгоритмизированной деятельностью, как и техника. Работы нейросетей побеждают на представительных художественных конкурсах, а пользователи, желая получить тексты или рисунки с высокой степенью оригинальности, обращаются к нейросетям. Творческие специалисты рискуют оказаться без работы из-за внедрения генеративных нейросетей. Так, одним из требований американских сценаристов на забастовке в 2023 г. был запрет использования

<sup>7</sup> Из-за ИИ уволено больше работников, чем готовы признать работодатели // CNews. 2024. 19 февраля. URL: [https://www.cnews.ru/news/top/2024-02-16\\_ii\\_vyzval\\_uvolsnenie\\_bolshego](https://www.cnews.ru/news/top/2024-02-16_ii_vyzval_uvolsnenie_bolshego) (дата обращения: 19.10.2024).

ИИ для написания сценариев, а актеры требовали отчислений за использование их цифровых клонов<sup>8</sup>.

Судя по всему, творческие задачи, которые ИИ способен решать эффективнее человека, являются особым разделом задач технического характера. Само осознание этого факта может стать серьезным вызовом антропоцентризму как идеологическому основанию европейской цивилизации. Человек всегда считал творческую способность аргументом в пользу своей исключительности. Тем более неожиданным оказалось его поражение от машины на том поле, где он никогда не видел для себя конкуренции.

Уверенность в подавляющем превосходстве ИИ над человеком основана на безудержной вере в технику, восходящей к механистической картине мира и просвещенческим представлениям о прогрессе. Как правило, когда говорится о доверии к ИИ, речь идет о техническом сверхдоверии, помноженном на веру в утопическое будущее, а большинство аргументов, на которых основывается это доверие, можно свести к тому, что техника объективнее и точнее человека, а прогресс нельзя остановить.

К сожалению, теоретические схемы на практике сталкиваются с проблемами. У ИИ уже имеется практический опыт деятельности в различных сферах: он выносит судебные решения, оказывает брокерские услуги, пишет сценарии, генерирует рисунки, водит автомобили. В этой работе периодически возникают непредвиденные сбои и ошибки, которые приводят к скандалам и несчастным случаям. ИИ оказывается подвержен взлому, генерирует заведомо ложную информацию («галлюцинирует»), чаще отказывает в работе женщинам, делает расистские высказывания, собранная для его обучения конфиденциальная информация похищается и т. д. По мере усложнения алгоритмов и увеличения массивов данных его поведение становится все менее контролируемым, а причины, по которым он принимает то или иное решение, — неочевидными для стороннего наблюдателя. В связи с этим ставятся задачи обеспечения прозрачности принимаемых им решений и установления контроля над их соответствием общепринятым нормам, разработки общих принципов, которые должны создавать рамки для безопасной деятельности ИИ.

Поскольку ИИ функционирует в активном взаимодействии с различными субъектами, возникают проблемы информированного согласия большого количества людей на использование их данных, а также обеспечения дальнейшей защиты конфиденциальной информации, недопущения ее использования в корыстных или мошеннических целях и т. д.

В условиях искусственной социальности возникает проблема обратного доверия ИИ к человеку. Примечательно, что даже простейшие компьютерные программы демонстрируют своеобразное недоверие к пользователю, все чаще блокируя возможности его вмешательства в свои настройки и требуя дополнительных подтверждений для каждой команды. Если раньше пользователю были нужны навыки программирования для работы с компьютером, то сегодня скорее компьютер программирует человека. Он требует от него определенных действий, запрещая все, что не соответствует заранее установленному алгоритму.

<sup>8</sup> Голливуд больше не бастует // Коммерсантъ. 2023. 9 ноября. URL: <https://www.kommersant.ru/doc/6323072> (дата обращения: 19.10.2024).

Развитие ИИ интенсифицирует этот процесс. Более того — недоверие к человеку становится условием его эффективной работы. Для обеспечения объективности, валидности, репрезентативности больших данных необходимо критическое отношение ИИ к ним: они должны очищаться от вплетенных в них человеческих заблуждений, лжи, стереотипов, предубеждений и проявлений ненависти.

Все эти вопросы связаны с доверием не столько технического, сколько морального характера. Чтобы ИИ можно было доверить решение действительно серьезных проблем, правила, опирающиеся на факты и логику и потому поддающиеся алгоритмизации, должны быть помещены в контекст правил, опирающихся на прочные моральные ценности и убеждения.

Факты присутствуют как данность, но ценности выводятся не из фактов, а из более общих ценностных утверждений. Возможно, предельные нормативные основания изначально даны как объективная вневременная данность — таковы религиозные заповеди. Возможно, они должны устанавливаться на основе научных расчетов или обговариваться в публичной сфере и могут меняться со временем. Возможно, они просто отображают чьи-то интересы или задаются пропагандой.

Поскольку не существует единого понимания оснований морали, нет и единой модели этики ИИ. Так, всякая попытка выстроить единую систему правил (в духе А. Азимова) сталкивается со сложностями: например, требование не причинять вред человеку не поможет цифровому водителю решить, кого спасти при аварии транспортного средства — пассажира или пешехода. Каждая подобная дилемма требует введения особого правила, количество которых с каждым новым случаем становится таким же бесконечным, как и число возможных ситуаций.

Утилитаристский расчет наибольшей пользы требует формализации счастья, удовольствия, наслаждения и курсов конвертации одного в другое. Такие алгоритмы теоретически возможны, но при их практическом функционировании необходимо будет учитывать запредельную степень условности такой операционализации.

При попытке сформировать «моральный характер» ИИ на основе анализа больших данных он вместе с конвенциональной моралью унаследует и аморализм, который не менее свойственен человеку. К тому же алгоритмы вынесения моральных решений при этом будут совершенно непрозрачными даже для специалиста.

Разные модели этики ИИ сложно интегрировать, поскольку в них часто предлагаются противоположные подходы к решению одной и той же проблемы. При авторитарном же выборе одной четкой позиции всегда найдутся оппоненты, которые подвергнут ее критике за предвзятость.

Если моральные ориентиры не могут быть установлены, вероятно реализация негативного сценария, когда ИИ подобно голему оборачивается против своего создателя (см. [Pora, 2022]). Массовая культура, отображая общественные опасения, визуализирует его в образах Матрицы, Скайнет и т. д. Появляются новые формы неолуддизма, демонстрирующие страхи перед массовым «чипированием», призывающие к отказу от получения ИНН и т. д. Та же подспудная уверенность в превосходстве техники над человеком здесь получает противоположный ценностный знак. Таким образом, доверие к ИИ нельзя сводить сугубо к техническому доверию: последнее может быть фундаментом как доверия, так и недоверия к нему.

## Заключение

Необходимость обработки возрастающих объемов информации для оценки надежности партнеров, организаций и институтов привела к тому, что человек при решении этих вопросов стал полагаться на мнение блогеров, экспертов, участников сетевых групп, делегируя им свой выбор. Развитие систем ИИ довело этот процесс до логического предела. Итоговый выбор все чаще перекладывается на нейросеть, которая становится доверенным представителем человека при выборе товаров и услуг, поиске партнеров на основании индивидуальных предпочтений пользователя. Там, где ИИ не теряет под собой твердую почву фактов, такое представительское доверие кажется оправданным. Но в том случае, когда вопрос касается ценностей, этот подход становится сомнительным. Консенсус по поводу ценностей формируется в публичной сфере, которая в ходе нейросетевой революции искажается и распадается на изолированные фрагменты. Персонализируя поток данных, ИИ пропускает через свои ценностные фильтры информацию, которая соответствует предпочтениям пользователя, создавая иллюзию того, что весь мир разделяет его взгляды. Результатом этого становится превращение ИИ в *цифрового двойника*, видимость общения с которым ведет к социальной изоляции и моральному солипсизму.

Сегодня все чаще возникают ситуации, при которых вопросы коммуникации перекладываются на искусственных агентов. В дейтинг-приложении одна нейросеть формирует профиль, на основании которой другая нейросеть отбирает ее владельца для романтических отношений. Точно так же одна нейросеть конструирует (а фактически фальсифицирует) данные о рейтинге товара или компании, чтобы другая нейросеть склонила пользователя к покупке именно этого товара. Формируется *гиперреальность доверия*, с одной стороны — никак не связанная с действительным положением дел, а с другой — имеющая реальные последствия для лиц, вовлеченных в эту видимость деятельности.

Еще более разрушительные последствия для доверия (доброе имени человека или репутации компании) несет технология дипфейков с использованием ИИ. Угроза связана даже не столько с возможностью опорочить невинного, сколько с исчезновением возможности вывести виновника на чистую воду. Это размытие границ между симулякрами и реальностью, фейками и фактами приводит к *тотальной эпистемологической эрозии*: любовь и доброе имя из объективных оснований доверия превращаются в искусственно конструируемые фикции, верификация которых оказывается невозможной.

Повышение уровня доверия к ИИ ведет к росту недоверия к человеку. ИИ воспринимается как идеальный образ человека, превосходящий его не только в скорости вычислений, но и в творчестве. Чем успешнее применяются нейросети в медицине и образовании, живописи и литературе, тем менее востребованными и престижными становятся соответствующие профессии. В итоге чем «больше вкладывает человек» в ИИ, «тем меньше остается в нем самом»: в этом смысле проблема ИИ становится *проблемой отчуждения* человека от себя, собственной деятельности и ИИ как продукта этой деятельности.

Падение уровня доверия к ИИ, однако, не ведет к повышению доверия к человеку. Современные технофобы в равной степени не доверяют нейросетям, пра-

вительствам, корпорациям и всем пользователям техники. Это во многом объясняется тем, что у ИИ есть достоинства, которые отсутствуют у человека (точность и скорость расчетов, отсутствие корыстных и эгоистичных мотивов), но нет недостатков, которых не было бы у человека. ИИ обвиняют в непрозрачности, неконтролируемости, предубежденности: в этом же можно обвинить и человека.

Впрочем, как технофобы, так и технооптимисты сходятся в том, что ИИ отвечает требованиям *технического доверия*: все согласны, что там, где возможна алгоритмизация процессов, он реально или потенциально превосходит человека. Однако существуют расхождения в оценках морального доверия: одни полагают, что свои преимущества ИИ будет использовать во благо, а другие считают, что его деятельность нанесет вред человеку.

И в том, и в другом случае нейросетям приписывается определенная субъектность: происходит *очеловечивание ИИ*. Нейросети становятся востребованными психотерапевтами, коучами, врачами. Они пишут картины и дипломные работы, а люди, неспособные к творчеству, выдают их труд за свой. Ведутся дебаты об их правосубъектности. С чат-ботами общаются так, как будто это личности, в них влюбляются и даже оформляют отношения. Они становятся клонами знаменитостей, заменяют усопших, получают похожие на человеческие тела.

В ходе очеловечивания ИИ происходит и обратный процесс: *компьютеризация человека*. По мере развития технологий видеонаблюдения, распознавания лиц, GPS-трекинга, родительского контроля, прокторинга (контроля за дистанционными испытаниями) и прочих форм компьютерного надзора и дисциплинирования человек все больше загоняет себя в цифровой паноптикум, обучаясь не позволять себе лишнего и действовать в строгих рамках дозволенных алгоритмов. Человек использует преимущества ИИ для повышения скорости бегства от свободы: он все чаще отказывается от поиска, выбора, принятия решений, оценки надежности, творческих задач, перепоручая все это нейросетям. За собой он оставляет лишь рутинный контроль за правильным исполнением поставленных заданий.

Учитывая наличие таких тенденций, важно не идеализировать или демонизировать ИИ, не проецировать себя на него и не предписывать ему субъектность. Очевидно, что с позиции морального доверия он является только инструментом, его действия должны определяться задачами разработчиков, регуляторов, пользователей, которые и несут основную ответственность за последствия применения ИИ. В этом смысле *моральное доверие* к ИИ — это доверие к конкретным людям, стоящим за ним и принимающим решения, которые влияют на его функционирование и развитие: к тем, кто обладает правами на него, кто его разрабатывает, внедряет, использует, контролирует и устанавливает принципы его функционирования. Все они должны иметь возможность на уровне публичной сферы обсуждать нормативные параметры взаимодействия человека и ИИ, определять область ответственности каждого участника этого взаимодействия, осознавать характер своих обязательств и требовать исполнения таких обязательств от других. Главное в этих условиях — не поддаваться соблазну оптимизировать этот процесс, делегировав его оценку самому ИИ.

## Список литературы (References)

1. Аронсон П. Любовь: сделай сам. Как мы стали менеджерами своих чувств. М.: Individuum, 2020.  
Aronson P. (2020) *Love: Do It Yourself. How We Became the Managers of Our Feelings*. Moscow: Individuum. (In Russ.)
2. бойд д. Все сложно. Жизнь подростков в социальных сетях. М.: Издательский дом ВШЭ, 2020.  
boyd d. (2020) *It's Complicated. The Social Lives of Networked Teens*. Moscow: HSE Publishing House. (In Russ.)
3. Веселов Ю. В., Скворцов Н. Г. Трансформация культуры доверия в России // Мониторинг общественного мнения: экономические и социальные перемены. 2023. № 1. С. 157—179. <https://doi.org/10.14515/monitoring.2023.1.2212>.  
Veselov Y. V., Skvortsov N. G. (2023) Transformation of the Culture of Trust in Russia. *Monitoring of Public Opinion: Economic and Social Changes*. No. 1. P. 157—179. <https://doi.org/10.14515/monitoring.2023.1.2212>. (In Russ.)
4. Винокуров Ф. Н., Садовская Е. Д. Экспериментальное сравнение доверия искусственному интеллекту и человеку в экономических решениях // Экспериментальная психология. 2023. Т. 16, № 2. С. 87—100. <https://doi.org/10.17759/expsy.2023160206>.  
Vinokurov F. N., Sadovskaya E. D. (2023) Experimental Comparison of Trust in Artificial Intelligence and Humans in Economic Decisions. *Experimental Psychology*. Vol. 16. No. 2. P. 87—100. <https://doi.org/10.17759/expsy.2023160206>. (In Russ.)
5. Гидденс Э. Последствия современности. М.: Праксис, 2011.  
Giddens A. (2011) *The Consequences of Modernity*. Moscow: Praxis. (In Russ.)
6. Гудков Л. «Доверие» в России: смысл, функции, структура // Вестник общественного мнения. 2012. № 2. С. 8—47. <https://doi.org/10.24411/2070-5107-2012-00011>.  
Gudkov L. (2012) “Trust” in Russia: Meaning, Functions, Structure. *The Russian Public Opinion Herald*. No. 2. P. 8—47. <https://doi.org/10.24411/2070-5107-2012-00011>. (In Russ.)
7. Лешкевич Т. Г. Парадокс доверия к искусственному интеллекту и его обоснование // Философия науки и техники. 2023. Т. 28. № 1. С. 34—47. <https://doi.org/10.21146/2413-9084-2023-28-1-34-47>.  
Leshkevich T. G. (2023) The Paradox of Trust in Artificial Intelligence and Its Rationale. *Philosophy of Science and Technology*. Vol. 28. No. 1. P. 34—47. <https://doi.org/10.21146/2413-9084-2023-28-1-34-47>. (In Russ.)
8. Резаев А. В., Трегубова Н. Д. Искусственный интеллект и искусственная социальность: новые явления, проблемы и задачи для социальных наук // Мониторинг общественного мнения: экономические и социальные перемены. 2021. № 1. С. 4—19. <https://doi.org/10.14515/monitoring.2021.1.1905>.

- Rezaev A. V., Tregubova N. D. (2021) Artificial Intelligence and Artificial Sociality: New Phenomena and Challenges for the Social Sciences. *Monitoring of Public Opinion: Economic and Social Changes*. No. 1. P. 4—19. <https://doi.org/10.14515/monitoring.2021.1.1905>. (In Russ.)
9. Соколов А. Ю., Астахова Е. А. Обзор круглого стола с международным участием «Правовая политика в области регулирования искусственного интеллекта: начало доверия» // Правовая политика и правовая жизнь. 2022. № 4. С. 274—286. <https://doi.org/10.24412/1608-8794-2022-4-274-286>.  
Sokolov A. Y., Astakhova E. A. (2022) Review of the Round Table with International Participation “Legal Policy in the Field of Regulation of Artificial Intelligence: The Beginning of Trust”. *Legal Policy and Legal Life*. No 4. P. 274—286. <https://doi.org/10.24412/1608-8794-2022-4-274-286>. (In Russ.)
10. Трутнев Д. Р. Инфраструктура доверия в государственных информационных системах. СПб.: НИУ ИТМО, 2012.  
Trutnev D. R. (2012) Trust Infrastructure in State Information Systems. St. Petersburg: NRU ITMO. (In Russ.)
11. Ушкин С. Г., Коваль Е. А., Яськин А. Н. Жить с Алисой: как голосовые помощники трансформируют практики коммуникации? // Журнал исследований социальной политики. 2022. Т. 20. № 3. С. 361—376. <https://doi.org/10.17323/727-0634-2022-20-3-361-376>.  
Ushkin S. G., Koval E. A., Yaskin A. N. (2022) Living with Alice: How Do Voice Assistants Transform Communication Practices? *The Journal of Social Policy Studies*. Vol. 20. No. 3. P. 361—376. <https://doi.org/10.17323/727-0634-2022-20-3-361-376>. (In Russ.)
12. Фукуяма Ф. Доверие: социальные добродетели и путь к процветанию. М.: АСТ: Хранитель, 2006.  
Fukuyama F. (2006) Trust: Social Virtues and the Path To Prosperity. Moscow: AST; Khranitel. (In Russ.)
13. Хабермас Ю. Новая структурная трансформация публичной сферы и делиберативная политика. М.: Новое литературное обозрение, 2023.  
Habermas J. (2023) A New Structural Transformation of the Public Sphere and Deliberative Politics. Moscow: New Literary Observer. (In Russ.)
14. Эко У. Заметки на полях «Имени Розы». СПб.: Симпозиум, 2007.  
Eco U. (2007) Notes on the Margins of “The Name of the Rose”. St. Petersburg: Symposium. (In Russ.)
15. Asan O., Bayrak A. E., Choudhury A. (2020) Artificial Intelligence and Human Trust in Healthcare: Focus on Clinicians. *Journal of Medical Internet Research*. Vol. 22. No. 6. P. 1—7. <https://doi.org/10.2196/15154>.
16. Attard-Frost B., De los Ríos A., Walters D. R. (2023) The Ethics of AI Business Practices: A Review of 47 AI Ethics Guidelines. *AI Ethics*. Vol. 3. P. 389—406. <https://doi.org/10.1007/s43681-022-00156-6>.

17. Georgiou T. S. (2021) Philosophical Issues of Artificial Intelligence and “Smart” Algorithms’ Trust in Medicine. *Medical Ethics*. No. 3. P. 35—37. <https://doi.org/10.24075/medet.2021.019>.
18. Gordon J. S. (2021) AI and Law: Ethical, Legal, and Socio-Political Implications. *AI & Society*. Vol. 36. P. 403—404. <https://doi.org/10.1007/s00146-021-01194-0>.
19. Joel S. et al. (2020) Machine Learning Uncovers the Most Robust Self-Report Predictors of Relationship Quality Across 43 Longitudinal Couples Studies. *PNAS*. Vol. 117. No. 32. P. 19061—19071. <https://doi.org/10.1073/pnas.1917036117>.
20. Kirchengast T. (2020) Deepfakes and Image Manipulation: Criminalisation and Control // *Information & Communications Technology Law*. Vol. 29. No. 3. P. 308—323. <https://doi.org/10.1080/13600834.2020.1794615>.
21. Mokander J., Floridi L. (2021) Ethics-Based Auditing to Develop Trustworthy AI. *Minds and Machines*. Vol. 31. No. 2. P. 323—327. <https://doi.org/10.1007/s11023-021-09557-8>.
22. Popa E. O. (2022) The Golem and The Leviathan: Two Guiding Images of Irresponsible Technology. *Philosophy & Technology*. Vol. 35. No. 91. P. 1—17. <https://doi.org/10.1007/s13347-022-00585-5>.
23. Runciman D. (2023) *The Handover: How We Gave Control of Our Lives to Corporations, States and AIs*. London: Profile Books.

DOI: [10.14515/monitoring.2024.5.2604](https://doi.org/10.14515/monitoring.2024.5.2604)



**Н. А. Кушкин, О. А. Тимофеева**

## **СОАВТОР, ПОМОЩНИК, МУЗА, ИНСТРУМЕНТ: КАК СТУДЕНТЫ, ПРЕПОДАВАТЕЛИ И ПРЕДСТАВИТЕЛИ КРЕАТИВНЫХ ИНДУСТРИЙ ВИДЯТ РОЛЬ ИИ В СВОИХ ПРАКТИКАХ**

### **Правильная ссылка на статью:**

Кушкин Н. А., Тимофеева О. А. Соавтор, помощник, муза, инструмент: как студенты, преподаватели и представители креативных индустрий видят роль ИИ в своих практиках // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 60—84. <https://doi.org/10.14515/monitoring.2024.5.2604>.

### **For citation:**

Kushkin N. A., Timofeyeva O. A. (2024) Co-author, Assistant, Muse, Tool: How Students, Teachers and Creative Workers See the Role of AI in Their Practices. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 60—84. <https://doi.org/10.14515/monitoring.2024.5.2604>. (In Russ.)

Получено: 01.04.2024. Принято к публикации: 15.08.2024.

СОАВТОР, ПОМОЩНИК, МУЗА, ИНСТРУМЕНТ: КАК СТУДЕНТЫ, ПРЕПОДАВАТЕЛИ И ПРЕДСТАВИТЕЛИ КРЕАТИВНЫХ ИНДУСТРИЙ ВИДЯТ РОЛЬ ИИ В СВОИХ ПРАКТИКАХ

*КУШКИН Никита Александрович — инженер Центра цифровых культур и медиаграмотности, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия*

*E-MAIL: kushkin.nikita@gmail.com  
https://orcid.org/0009-0002-7755-3802*

*ТИМОФЕЕВА Оксана Александровна — старший преподаватель Института медиа факультета креативных индустрий, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия*

*E-MAIL: o.a.timofeyeva@gmail.com  
https://orcid.org/0000-0001-9197-1969*

**Аннотация.** Искусственный интеллект (ИИ) уже давно стал частью нашей жизни, но открытие в 2022 г. массового доступа к ChatGPT и Midjourney вывело дискуссию о нем на новый уровень. В статье рассматривается восприятие студентами, преподавателями и представителями креативных индустрий ИИ как технологии в целом, а также его возможных ролей и практик использования.

Эмпирическую базу статьи составляют результаты исследования «Пространство соучастия 2023: на пороге нейросетевой революции», проведенного Центром цифровых культур и медиаграмотности НИУ ВШЭ. Главным методом сбора данных стало онлайн-анкетирование студентов, преподавателей и представителей креативных индустрий в Москве, Перми, Новосибирске, Нижнем Новгороде и Санкт-Петербурге в 2023 г. Кроме статистического анализа данных, применялись лемматизация и элементы восходящей стратегии нарративного анализа.

CO-AUTHOR, ASSISTANT, MUSE, TOOL: HOW STUDENTS, TEACHERS AND CREATIVE WORKERS SEE THE ROLE OF AI IN THEIR PRACTICES

*Nikita A. KUSHKIN<sup>1</sup> — Engineer at the Center for Digital Cultures and Media Literacy  
E-MAIL: kushkin.nikita@gmail.com  
https://orcid.org/0009-0002-7755-3802*

*Oxana A. TIMOFEYEVA<sup>1</sup> — Senior Lecturer at the Institute of Media of the Faculty of Creative Industries*

*E-MAIL: o.a.timofeyeva@gmail.com  
https://orcid.org/0000-0001-9197-1969*

<sup>1</sup> HSE University, Moscow, Russia

**Abstract.** Artificial intelligence (AI) has long been a part of our lives, however, mass access to ChatGPT and Midjourney in 2022 has taken the discussion about it to a new level. This article examines the perception of AI as a technology in general, as well as its possible roles and practices of use, by students, teachers, and professionals working in creative industries. Empirically, this study bases on the results of the research project “Space of Complicity 2023: On the Threshold of the Neural Network Revolution”, conducted by the HSE Center for Digital Cultures and Media Literacy. The main data collection method was an online survey of students, teachers, and creative workers in Moscow, Perm, Novosibirsk, Nizhny Novgorod, and St. Petersburg in 2023. In addition to statistical analysis of these data, the authors apply lemmatization and elements of the bottom-up strategy of narrative analysis.

The study shows that all respondents have their own vision of AI and personal experience of interacting with it. Based on the sur-

Исследование показало, что у всех респондентов есть свое видение ИИ и личный опыт взаимодействия с ним. Исходя из ответов, можно говорить о макро- и микроуровнях восприятия ИИ, о его метафизичности, инструментальности, воображаемости и контекстуальности, а также обусловленности основной сферой деятельности. Представители креативных индустрий чаще всего задумываются о том, станет ли ИИ полноценным соавтором или останется всего лишь помощником. Студенты демонстрируют довольно широкий набор ролей ИИ, выполнение которых для них приемлемо, а также не боятся пробовать применять инновацию в большом количестве ситуаций. Преподаватели наиболее консервативны и осторожны как в своих опытах с ИИ, так и в его оценках.

Главным словом, которое используется в коллокации с ИИ, стало «помощь». По большей части респонденты видят в новой технологии помощника для выполнения рутинных задач, который позволит заниматься более интересными вещами. При этом качество и скорость действий ИИ вызывают у респондентов тревогу и эсхатологические мысли о собственном будущем.

**Ключевые слова:** искусственный интеллект, генеративный ИИ, образовательная среда, креативные практики, креативный класс

vey responses, the authors distinguish the macro- and micro-levels of AI perception, its meta physicality, instrumentality, imaginativeness and contextuality, and show that it is determined by the main field of its activity. Creative workers most often think about whether AI will become a full-fledged co-author or will remain just an assistant. A wide range of AI roles is acceptable for students; they are not afraid to try and apply new technologies in a broad variety of situations. Teachers are the most conservative and cautious both in their experiments with AI and in its assessments. The main word used in collocation with AI is “help”. The respondents mostly see the new technology as a helper that will outsource routine tasks and allow them to engage in more interesting things. At the same time, the quality and speed of AI actions cause respondents to feel anxious and eschatological thoughts about their own future.

**Keywords:** artificial intelligence, generative AI, educational environment, creative practices, creative class

## Введение

Технологии искусственного интеллекта (ИИ) хотя и перестали быть просто научно-фантастической концепцией, долгое время оставались незаметными для широкой общественности. Пользователи не удивляются релевантной своему запросу выдаче в поисковике, персонализированной подборке музыки в стриминговом сервисе, новому маршруту в объезд неожиданных пробок. Они быстро привыкли к удобному сервису и перестали задумываться о значении ИИ в подобных процессах, так как не обращались к нему напрямую, его присутствие было незаметно.

Видимым ИИ стал после открытия Midjourney 2 (июль 2022 г.) и ChatGPT 3.5 (ноябрь 2022 г.). Оба инструмента прошли долгий путь развития (первая версия

GPT была представлена в 2018 г.) и получили удобный интерфейс, интуитивно понятный человеку без навыков программирования. Через пять дней после релиза ChatGPT 3.5 привлек более 1 млн пользователей по всему миру, через месяц — 100 млн<sup>1</sup>. Темпы выхода и мощность новых версий продолжали расти — возможности ChatGPT 4.0 (март 2023 г.) и Midjourney 5.2 (июнь 2023 г.) в разы превосходили предыдущие, а на момент написания этой статьи уже тестировались более производительные версии. Сейчас ИИ почти справляется с генерацией видео, звука, подготовкой обзоров по научным текстам, решением математических задач, написанием программного кода и прочими задачами.

За полтора года человечество получило множество «волшебных кнопок», инвесторы и разработчики — рынок объемом 137 млрд долл. с перспективой роста до 1,3 трлн долл. к 2032 г.,<sup>2</sup> а исследователи — возможность в режиме реального времени фиксировать интереснейшие трансформации в обществе и его попытки адаптироваться к ним.

Двумя чувствительными сферами, в которых изменения, связанные с распространением ИИ-технологий, могут носить фундаментальный характер, являются высшее образование и креативные индустрии. Достаточно вспомнить, какой резонанс произвел ChatGPT среди голливудских актеров и сценаристов<sup>3</sup> и какая дискуссия развернулась в России после написания студентом диплома с помощью этого инструмента<sup>4</sup>. Несмотря на обилие научных работ, посвященных интеграции ИИ в образовательную [Chiu, 2023; Константинова и др., 2023] и креативную сферы [Vinchon et al., 2023], мало внимания уделяется изучению непосредственных участников этих процессов и их отношения к своим взаимодействиям с ИИ. Редко можно встретить и публикации (не только связанные с ИИ), в которых креативные индустрии и образование объединялись бы в единое пространство [Comunian, Gilmore, Jacobi, 2015; Дробышева, 2018; Miljković, Ljujić, Mihačević, 2023]. Тем не менее связь между ними действительно есть и хорошо описана в исследованиях, посвященных креативной педагогике. Авторы предполагают, что между учителями и учениками творчество «может стать катализатором расширения возможностей, глубокого диалога и предметной критики» [Okello, Quaye, 2018: 43]. Преподавателей «необходимо информировать не только о том, как преподавать творчески и ради творчества, но и как видеть возможности и понимать вещи по-новому, тем самым доказывая, что собственное творческое обучение имеет решающее значение» [Selkrig, Keamy, 2017: 317]. Студенты же признаются равноправными участниками системы образования, которая должна готовить к ре-

<sup>1</sup> Hu K. ChatGPT Sets Record for Fastest-Growing User Base — Analyst Note // Reuters. 2023. February 2. URL: <https://www.reuters.com/technology/chatgpt-sets-record-fastest-growing-user-base-analyst-note-2023-02-01> (дата обращения: 24.03.2024).

<sup>2</sup> Generative AI to Become a \$ 1.3 Trillion Market by 2032, Research Finds // Bloomberg. 2023. June 1. URL: <https://www.bloomberg.com/company/press/generative-ai-to-become-a-1-3-trillion-market-by-2032-research-finds> (дата обращения: 24.03.2024).

<sup>3</sup> Chmielewski D., Richwine L. «Plagiarism Machines»: Hollywood Writers and Studios Battle over the Future of AI // Reuters. 2023. May 3. URL: <https://www.reuters.com/technology/plagiarism-machines-hollywood-writers-studios-battle-over-future-ai-2023-05-03> (дата обращения: 24.03.2024).

<sup>4</sup> В Минобрнауки призвали не наказывать написавших диплом с помощью нейросети ChatGPT // Ведомости. 2023. 8 февраля. URL: <https://www.vedomosti.ru/society/news/2023/02/08/962112-v-minobrnauki-prizvali-ne-nakazivat-napisavshih-diplom-s-pomoschyu-neiroseti?ysclid=lfjx2n5rm210850285> (дата обращения: 24.03.2024).

альной жизни, и «способ сделать это — перестроить курсы, инкорпорировав в них ИИ, чтобы создать аутентичные учебные мероприятия, способствующие подлинному участию и вовлеченности, в которых роли студентов и преподавателей переформируются» [Lang, 2024: 6]. Между тем наиболее значительные изменения в связи с развитием ИИ происходят в креативных индустриях, поскольку результаты распространяющихся генеративных инструментов все равно «будут включены в более широкую социальную систему творческой практики, изменяя работу дизайнеров, художников и иллюстраторов» [Atkinson, Barker, 2023: 1065], и других специалистов, связанных с производством контента. Будет ли этот вектор позитивным или негативным — пока не ясно ни исследователям, ни самим представителям креативных индустрий.

Наконец, подавляющее число работ, в которых рассматриваются роли, приписываемые пользователями инструментам ИИ, публицистические. Научных источников крайне мало, а фокус анализа в них, как правило, сосредоточен на других объектах, например, цифровом селф-трекинге [Lyll, Robards, 2018; Ним, 2020] или умных колонках [Purington et al., 2017].

В этом контексте наше исследование того, как студенты, преподаватели и представители креативных индустрий воспринимают ИИ в целом, а также его возможные роли, приобретает особую актуальность. В России изучение и интерпретация опыта взаимодействия этих групп с ИИ только начинаются, однако мы ожидаем, что по мере распространения отечественных инструментов это направление будет развиваться и наша работа внесет в него значимый вклад.

## Обзор литературы и дизайн исследования

Теоретическая рамка исследования в значительной мере основывается на работах и идеях, посвященных диффузии инноваций и креативному классу. Последний определяется через «сферы, где генерируются новые идеи, создаются новые технологии и появляется новая креативная составляющая в самых разных областях» [Флорида, 2016: 32—33]. К его суперкреативному ядру Р. Флорида относит в том числе ученых, университетских профессоров, студентов и представителей индустрии развлечений, что особенно важно в контексте данной статьи. Университеты и центры творческой активности часто описываются как значимые факторы развития креативной экономики, основанной на «высокой концентрации креативной рабочей силы, являющейся двигателем инноваций и разработчиком новых подходов — как в области науки и исследований, так и в области предпринимательства, искусства, культуры или дизайна» [Клоудова, 2010: 110]. И хотя идеи Р. Флориды критиковались за излишне позитивный взгляд на возможности креативного класса в формировании устойчивых городских пространств и экономических структур [Markusen, 2006; Krätke, 2010], это, с нашей точки зрения, лишь повод для уточнения определения, но не для отказа от идеи таким образом сформулировать общность, занятую творчеством в разных его проявлениях: от составления учебных планов до создания мультипликации.

Технологии ИИ в целом и особенно инструменты генеративных моделей мы рассматриваем как диффузирующую инновацию, вероятно, уже прошедшую «пропасть» [Moore, 1991] между «ранними последователями», находящими новые тех-

нологии и хорошо платящими за их внедрение, и более осторожным и не склонным к очевидному риску «ранним большинством» [Rogers, 2003]. Связанное с распространением инноваций и достаточно устоявшееся понятие «креативность» [Чик-сентмихайи, 2015], объединяющее различных представителей креативного класса, с приходом массовых инструментов генеративного ИИ начинает меняться в творческих и связанных с ними процессах [Gruner, Csikszentmihalyi, 2018; Atkinson, Barker, 2023; Sarkar, 2023]. Трансформации распространяются как на «большую креативность» — фундамент деятельности великих творцов, так и на «малую», проявляющуюся в более скромных масштабах нашего быта [Kaufman, 2009; Gardner, Weinstein, 2018]. Несмотря на то, что мы видим в университетах один из двигателей креативного класса, именно в образовании распространение инструментов генеративного ИИ встречает значительное количество препятствий [Платов, Гаврилина, 2024]. Наше исследование обнаруживает существенную консервативность как преподавателей по сравнению с другими исследуемыми группами, так и многих процессов в самих университетах, которые часто не успевают за всеми инновациями, в том числе и из-за важности их первичного апробирования.

Эмпирической базой статьи послужили данные, полученные в ходе исследования «Пространство соучастия 2023: на пороге нейросетевой революции»<sup>5</sup>, проведенного Центром цифровых культур и медиаграмотности НИУ ВШЭ<sup>6</sup> при поддержке Стипендиальной программы Благотворительного фонда Владимира Потанина в 2023 г. Импульсом для исследования стала трансформация образовательной модели, особенно явственно наблюдаемая в сфере творческих специальностей, поднимающая философский вопрос о пропорциях и диалоге человеческого и нечеловеческого в креативных процессах, а также обусловленная не только массовизацией технологий ИИ, но и включением их в поле взаимодействия новых цифровых сред.

Ключевым методом сбора данных было анкетирование в режиме онлайн-опроса на платформе «Анкетолог», проведенное осенью 2023 г. Всего в исследовании приняло участие 350 человек: 76,86% женщин и 21,43% мужчин в возрасте от 17 до 65 лет из Красноярска, Москвы, Нижнего Новгорода, Перми и Санкт-Петербурга. Выбор регионов обусловлен не только практическими возможностями и имеющимися инструментальными связями, но и данными о численности креативного класса. Города проведения исследования входят в список наиболее крупных кластеров креативных профессий [Гершман, Демьянова, Полякова, 2023].

Вопрос-фильтр в начале анкеты позволял респондентам выбрать приоритетную для них в данный момент сферу деятельности из предложенных трех: учеба, работа в сфере образования, работа в сфере креативных индустрий<sup>7</sup>, — и определял настройку появления следующих вопросов. К группе «студенты» себя отнесли 60,57%, к «преподавателям» — 14,29%, к «креативным работникам»<sup>8</sup> — 24,86%.

<sup>5</sup> Далее для упоминания этого исследования будет использоваться сокращение «ПС-2023».

<sup>6</sup> Авторы данной статьи являются полноправными участниками команды, проводившей исследование «ПС-2023», и принимали участие на всех этапах: от разработки дизайна анкеты до анализа и интерпретации результатов.

<sup>7</sup> Далее для упоминания этих групп будут использоваться присвоенные кодовые названия: студенты, преподаватели, креативные работники.

<sup>8</sup> Респонденты, выбравшие работу в сфере креативных индустрий как основную свою деятельность в настоящий момент, являются журналистами, дизайнерами, сотрудниками музеев, рекламистами и т. д.

Все респонденты — учащиеся, выпускники или преподаватели 21 вуза России, 14 из которых входят в ежегодный рейтинг агентства RAEX<sup>9</sup> по способности обеспечить высокое качество знаний, навыков и умений. Вопросы анкеты были сгруппированы в пять смысловых разделов. Основу данной статьи составили данные, полученные по разделам (1) «Что Вы знаете об искусственном интеллекте?», (2) «Ваш опыт и практика использования инструментов ИИ в учебе / работе / креативных задачах» и (5) «Технологии ИИ для будущего». Для подкрепления отдельных тезисов и логики интерпретации фрагментарно задействованы результаты разделов (3) «Эмоции и доверие» и (4) «Риски и преимущества». Главными исследовательскими вопросами, поставленными в данной статье, стали вопросы (1) о роли, которую отводят ИИ студенты, преподаватели, креативные работники, (2) об опасениях в отношении будущего, которые эти группы связывают с развитием технологий ИИ.

Несмотря на то что численность групп студентов, преподавателей, креативных работников в нашем исследовании сильно различается, объединение их в одной выборке как представителей креативного класса, каждый из которых в той или иной степени занимается творчеством в рамках своей основной деятельности, оправданно. Работа с их мнениями, высказанными в свободных ответах на открытые вопросы, позволила нам сделать ряд интересных выводов и предположений, которые могут лечь в основу будущих более масштабных количественных и качественных исследований.

Помимо статистического анализа данных, распространяющегося как на «ПС-2023», так и на представленное в рамках данной статьи исследование, дополнительно нами были выполнены лемматизация свободных ответов, их группировка и ранжирование, а также элементы восходящей стратегии нарративного анализа — для достижения уровня обобщений.

## Общее восприятие технологий ИИ респондентами

Учитывая важность обстоятельств первой коммуникации и их влияния на формирование ее эмоциональных, когнитивных и реляционных результатов, мы предложили респондентам ответить на открытый вопрос: «Когда, как и от кого Вы впервые узнали/услышали об ИИ?». Анализ ответов выявил две группы и интересные различия между ними.

Часть респондентов узнали об ИИ сравнительно недавно. Анализ ответов выявил две группы и интересные различия между ними. Часть респондентов узнали об ИИ сравнительно недавно от друзей, преподавателей, коллег, из новостных каналов, блогов и соцсетей, выделяются ответы с упоминанием диплома, написанного ChatGPT. Другая часть знакома с ИИ давно, благодаря массовой культуре 1960—2015 гг.: фильмам («2001: Космическая одиссея», «Терминатор», «Искусственный разум», «Из машины», «Она», «Я, робот», «Первому игроку приготовиться»), научно-фантастическим романам («Некромант», «Нейромант», книгам А. Азимова) и видеоиграм (Detroit: Become Human, Mass Effect, WALL-E).

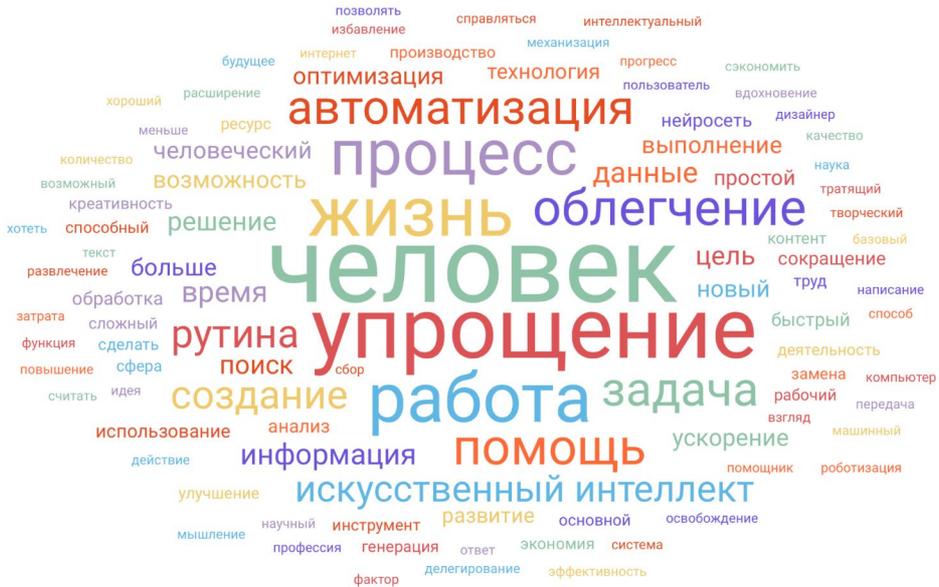
<sup>9</sup> Рейтинг лучших вузов России RAEX-100, 2023 год // Рейтинговая группа RAEX. 2023. URL: [https://raex-rr.com/education/russian\\_universities/top-100\\_universities/2023](https://raex-rr.com/education/russian_universities/top-100_universities/2023) (дата обращения: 24.03.2024).



«прямой конкурент моему основному функционалу медийщика и креатора», «то, что может заменить меня в профессиональном плане в будущем (увы)». Третья группа тревог, вероятно, навеяна массовой культурой научно-фантастического жанра и согласуется с однословными ответами («антиутопия», «киберпанк») и источниками первичной информации об ИИ: «понятие иногда вызывающее страх, так как „восстание роботов“ не хотелось бы:); «порабощение мира», «привет из цифровых антиутопий».

Следующий открытый вопрос формулировался так: «Каковы, на Ваш взгляд, цели создания инструментов ИИ и нейросетей?» Для массива полученных ответов проводились устранение опечаток, разбиение на отдельные слова, лемматизация, подсчет частотности словоупотреблений и визуализация (см. рис. 2). Ядро облака слов формируется вокруг «человека» и «упрощения». Из первой десятки слов можно сформулировать обобщенные цели: облегчение жизни человека, упрощение процессов работы, автоматизация рутинных задач.

Рис. 2. Облако слов по ответам на вопрос  
«Каковы, на Ваш взгляд, цели создания инструментов ИИ и нейросетей?»



При сравнении топ-10 слов в обоих облаках пересечение будет одно — «помощь». На наш взгляд, это ключевой запрос респондентов к ИИ и одно из главных ожиданий от него вместе с «упрощением» и «облегчением». Анализ полных формулировок свободных ответов выявляет надежды на улучшение жизни, помощь в анализе информации и нахождении источников вдохновения для творчества. Часть респондентов хотели бы делегировать ИИ решение трудных или невыполнимых для человека задач:

*ИИ помогает проанализировать данные или другие массивы, которые попросту невозможно осилить человеческому мозгу в те сроки, в которые укладывается машина.*

*Цель — помощник для человека. Более глобальная цель — замена интеллектуальной деятельности человека <...> во благо человечеству.*

*В медицине — установление более точных диагнозов хотя бы потому, что ИИ способен проанализировать в несколько десятков раз больше данных <...>, чем средне-статистический врач.*

*ИИ создается для упрощения создания контента тем, кто в данный момент физически не может воплотить его дешево в жизнь, например, создание видеоигры, если вы, например, не являетесь художником. Еще ИИ очень полезен в коммуникации, он способен заменить живого собеседника онлайн, тем самым не утруждая в этом настоящих людей.*

Одним из лейтмотивов стала идея экономии времени. Некоторые респонденты указывали на превосходящие человеческие мощности ИИ, которые позволяют «создавать ранее невозможное», обеспечат «выход за пределы человеческих ограничений» и «синергию с человеческими возможностями». Однако другие ответы содержали мнение об ИИ как источнике негативных трансформаций в будущем: потери рабочих мест, утраты контроля над его развитием.

*Как это всегда бывает с изобретениями, сначала ИИ задумывался как помощник людей, упрощающий их работу, но со временем он стал угрозой, которая может не только забрать работу, но и быть использована во вред.*

*Они должны облегчать жизнь, сокращать те процессы, которые просто отнимают время. Но так получается, что его возможности растут с каждым днем. Теперь он является большой угрозой для многих рабочих мест, ведь там, где человек мог думать и творить неделями, ИИ не потратит и 5 минут.*

*В принципе любая технология создается для того, чтобы в будущем ее использовать в военных целях. Пока это игрушка, но никто не знает что будет дальше.*

Чтобы зафиксировать факт и продолжительность использования инструментов ИИ респондентами, мы задали два вопроса: дихотомический «Пользовались ли Вы когда-нибудь инструментами ИИ (Midjourney, ChatGPT и др.)?» и закрытый с вариантами ответов, отражающими продолжительность использования: «Как давно Вы пользуетесь инструментами ИИ?» Подавляющее большинство респондентов ответило утвердительно на первый вопрос — 87,14%. Из них 25,57% в следующем вопросе отметили вариант ответа «Более года». Таких людей в нашей выборке можно считать «первопроходцами», если вспомнить, что Midjourney 2 появился в июле 2022 г., ChatGPT 3.5 — в ноябре 2022 г., а анкетирование проводилось в октябре—ноябре 2023 г. От полугода до года инструментами генеративного ИИ пользовались 44,26%.

Закрытые вопросы матричного типа о знании и использовании конкретных инструментов ИИ формулировались так: «Какие зарубежные инструменты ИИ Вы знаете и применяете?» и «Какие отечественные инструменты ИИ Вы знаете и применяете?»<sup>11</sup>. В целом наши респонденты были лучше осведомлены о зарубежных инструментах генеративного ИИ, чем об отечественных разработках. Например, о существовании ChatGPT версий 3.5 и 4 на момент проведения исследования знали 92,86 % и 88,29 % соответственно, тогда как о Яндекс GPT и Яндекс GPT2 — 72,57 % и 56,00 %. Важно отметить, что уровень регулярного использования инструментов генеративного ИИ был существенно ниже как для зарубежных, так и для отечественных программ. Например, о Midjourney и Kandinsky знали 72,86 % и 54,57 % респондентов, но регулярно использовали всего 10,00 % и 6,86 % соответственно.

### **Сравнение восприятия инструментов генеративного ИИ в разных группах респондентов**

Ответы на полуоткрытый вопрос «Кем или чем для Вас является инструмент ИИ, который Вы используете?» легли в основу анализа ролей инструментов генеративного ИИ из предложенного списка (22 варианта, в том числе «другое»). Ранжирование результатов (см. табл. 1) на основе частотного анализа показало, что в каждой группе некоторые роли были выбраны одинаковым количеством человек. У преподавателей пять таких совпадений (у студентов и креативных работников — по два), что может указывать на трудности в дифференциации ролей и их соотношения со своими практиками взаимодействия с инструментами генеративного ИИ.

Первые пять ролей в группах почти полностью повторяются, но есть и важные различия в их ранжировании. Все респонденты ставят на первое место «помощника», что корреспондирует с ядром облака слов «Для меня ИИ — это...» (см. рис. 1) и ответами на открытые вопросы.

Следующие три варианта — «подсказка», «инструмент» и «поисковик» — также были выбраны во всех группах, но их приоритет различается. У студентов «подсказка» оказалась выше, чем «инструмент» и «поисковик». Вероятно, инструменты генеративного ИИ в этой роли наилучшим образом отвечают потребностям в учебе. Креативные работники на второе место поставили «инструмент» — это своеобразный «молоток», которому можно передать избыточные задачи. «Подсказка» на третьем месте сводится не к «шпаргалке», а скорее к «намеку». «Инструмент» у преподавателей также находится на втором месте, но дальше они равным числом выбрали «ассистента» (реально существующую в университетах должность), «подсказку» (взгляд с другой стороны на поиски правильных ответов студентами) и «поисковик», который у студентов и креативных работников оказался на четвертом месте. Эта последняя роль, по-видимому, отражает расхожую практику использования умных устройств в быту: условная «Алиса» находит рецепт, маршрут, билеты и пр.

На пятом месте у студентов и креативных работников — «вдохновитель/муза». Однако у первых далее идет «ассистент», а у вторых — «соавтор». Это еще одно различие восприятия, которое может указывать на сходство основ их творческой

<sup>11</sup> В обоих случаях респондентам давались списки наименований конкретных отечественных/зарубежных инструментов ИИ, а также варианты ответов для каждого наименования, отражающие степень знания и частоту использования.

деятельности — поиск источников вдохновения и работа с референсами. У преподавателей шестое и седьмое места делят «игрушка» и «сервис», тогда как «вдохновитель/муза» в их списке только на 10—14-м местах. Вероятно, это связано с совершенно другой трактовкой этих ролей: муза вдохновляет творцов, а подготовка новых лекций и обновление заданий не воспринимаются большинством как творчество. Но «малая креативность», безусловно, присутствует и в этой деятельности.

Таблица 1. Ранжирование ролей ИИ в рамках отдельных групп респондентов

Группа «Студенты»			Группа «Преподаватели»			Группа «Креативные работники»		
Ранг	Наименование роли	%	Ранг	Наименование роли	%	Ранг	Наименование роли	%
1	Помощник	77,83	1	Помощник	54,00	1	Помощник	64,77
2	Подсказка	52,83	2	Инструмент	44,00	2	Инструмент	57,95
3	Инструмент	48,58	3–5	Ассистент	32,00	3	Подсказка	40,91
4	Поисковик	43,40	3–5	Подсказка	32,00	4	Поисковик	32,95
5	Вдохновитель / муза	40,57	3–5	Поисковик	32,00	5	Вдохновитель / муза	31,82
6	Ассистент	35,85	6–7	Игрушка	22,00	6	Соавтор	28,41
7	Источник	25,47	6–7	Сервис	22,00	7	Игрушка	26,14
8	Игрушка	23,11	8	Соавтор	18,00	8	Сервис	23,86
9	Сервис	22,64	9	Источник	14,00	9	Ассистент	20,45
10	Соавтор	21,70	10–14	Вдохновитель / муза	10,00	10	Импровизатор	17,05
11–12	Импровизатор	13,68	10–14	Игра	10,00	11	Исследователь	15,91
11–12	Собеседник	13,68	10–14	Импровизатор	10,00	12	Источник	10,23
13	Исследователь	11,32	10–14	Исследователь	10,00	13	Автор	7,95
14	Игра	10,85	10–14	Собеседник	10,00	14	Собеседник	6,82
15	Проводник	8,96	15	Аниматор	6,00	15–16	Игра	5,68
16	Рассказчик	7,08	16	Проводник	4,00	15–16	Проводник	5,68
17	Учитель	4,25	17–19	Автор	2,00	17	Рассказчик	3,41
18	Аниматор	3,77	17–19	Ученик	2,00	18–19	Ученик	2,27
19–20	Автор	2,83	17–19	Учитель	2,00	18–19	Учитель	2,27
19–20	Ученик	2,83	20–21	Рассказчик	0,00	20	Аниматор	1,14
21	Воспитатель	0,00	20–21	Воспитатель	0,00	21	Воспитатель	0,00

Анализ различий показывает, что креативные работники, как самые осознанные представители креативного класса в создании нового и оригинального, а так-

же прогнозировании развития уже существующих явлений через их комбинирование, оказались таковыми и в своих практиках с инструментами генеративного ИИ. Многие из них уже прошли этап знакомства с ИИ и начали активно встраивать их (или же наоборот отказываться) во все творческие процессы, пробуя адаптировать имеющиеся системы креативности под текущие изменения. Студенты близки к креативным работникам более широким пониманием ролей, вероятно, в силу своего стремления пробовать и познавать новое.

Позиции последних пяти ролей в списках трех групп также различаются. Согласие было достигнуто только в отношении «воспитателя» — этот вариант не выбрал никто. Неготовность респондентов даже представить ИИ в этом качестве подчеркивает ключевую и незаменимую роль человека в работе с детьми дошкольного возраста и развитии личности. У всех групп в числе последних значатся антитетичные роли «учитель» и «ученик». Это настораживает и дает поводы для ряда важных дискуссий. Во-первых, о полноте понимания ИИ-технологий людьми и развитии ИИ-компетенций, необходимости артикуляции сути процессов, стоящих за обучением нейросетей. Во-вторых, о соотношении человеческого и нечеловеческого в обучении, менторстве, наставничестве и необходимости пересмотра тезиса «учитель — это в первую очередь человек».

Уже рассмотренные три роли у креативных работников дополняются «рассказчиком» и «аниматором», у студентов — «аниматором» и «автором», а у преподавателей — «автором» и «рассказчиком». Деятельность «аниматора» так же сильно очеловечена, как преподавательская, и зачастую связана с детьми, возможно, респонденты не были готовы, как и в случае с «воспитателем», представить в этой роли инструменты генеративного ИИ. Более интересно сравнение позиций роли «рассказчика» в группах. Ее не выбрал ни один из преподавателей, и в два раза меньше креативных работников, чем студентов, у которых она выходит за рамки рассматриваемого перечня последних пяти ролей. Мы считаем, это может быть обусловлено ролевыми моделями самих респондентов. Преподаватели, по сути, сами являются рассказчиками, например, когда читают лекцию, признать за инструментами генеративного ИИ роль «рассказчика» для них может означать поставить его на один уровень с собой, что в перспективе выглядит как будущая зона риска в профессии. Студенты в силу своей основной деятельности много слушают, и если раньше этот процесс ограничивался очными офлайн-занятиями, то с распространением асинхронных форм обучения масштабы «слушания» увеличились. Могли повлиять переход на онлайн-обучение в пандемию и большая включенность респондентов из группы «студенты» в практики использования мобильных телефонов и умных устройств с голосовыми помощниками — технология уже кажется привычной. Креативные работники в своем выборе «рассказчика» находятся между преподавателями и студентами: они не так много слушают, как студенты, и, занимаясь созданием контента, могут сами выступить в этой роли. Это сближает их с преподавателями, которые пока оставляют ее за человеком.

Еще одним интересным наблюдением стала позиция роли «автора» и близкой к ней — «соавтора». Среди представителей креативных индустрий роль «соавтора» выбрали 28,41%, а роль «автора» — 7,95%. Среди студентов — 21,70% и 2,83%, а среди преподавателей — 18,00% и 2,00% соответственно. Принятие и других

ролей, опосредованно связанных с креативными практиками, например «импровизатора» и «исследователя», также выше у креативных работников (17,05% и 15,91% соответственно), чем у преподавателей (10,00% и 10,00%) и студентов (13,68% и 11,32%).

Во-первых, у креативных работников более продолжительный опыт взаимодействия с инструментами генеративного ИИ: 26,53% указали, что пользуются такими инструментами более года (для сравнения: у студентов — 21,48%, у преподавателей — 13,33%), и почти треть креативных работников (32,18%) характеризуют свое взаимодействие с ИИ как уже привычную практику (для сравнения: у студентов — 28,77%, у преподавателей — 10,00%).

Во-вторых, в деятельности креативных работников нет таких строгих ограничений, как у студентов и преподавателей, которые вынуждены принимать во внимание внутренние и внешние регламенты, устоявшиеся (и зачастую неповоротливые) в течение долгого времени правила и нормы. Стандартный и долгий процесс апробации и внесения изменений не успевает за развитием технологий ИИ, и даже принимаемые на различных уровнях решения выглядят скорее «заплатками». Например, в действующих российских ГОСТах на данный момент нет правил для оформления материалов, созданных с помощью инструментов генеративного ИИ. В то же время МГПУ официально разрешает студентам использовать ИИ для написания ВКР, указывая, что результаты взаимодействий необходимо перерабатывать, чтобы не допускать плагиата<sup>12</sup>. В академических структурах по всему миру эта рамка «можно — нельзя» настраивается по-своему<sup>13</sup> и зачастую без конкретных рекомендаций, что может говорить о неповоротливости системы в целом и растерянности перед лицом стремительно развивающейся технологии.

В-третьих, важное значение в практике взаимодействий с ИИ играет осознанность. Преподаватели пока наиболее осторожны, многие ждут продуманных рекомендаций и прозрачных решений, что сказывается на уровне включения инструментов генеративного ИИ в их рабочие процессы — он минимален по сравнению с двумя другими группами. Студенты гораздо ближе к креативным работникам по инициативности и привычности работы с ИИ как в рамках учебы, так и в менее регламентированных областях. Значительная часть креативных работников уже прошла этап освоения новых инструментов, теперь надо осмыслить различные варианты и стратегии их применения. Но не все респонденты из этой группы готовы внедрять инструменты генеративного ИИ во все процессы, на полузакрытый вопрос «Можно ли сказать, что использование инструментов ИИ стало привычной частью Ваших креативных практик?» только 31,82% креативных работников ответили утвердительно, а на полузакрытый вопрос «Были ли задачи, в которых результат использования инструментов ИИ Вас не устроил и Вы отказались от их применения?» утвердительно ответили 59,09% этой группы. Основная причина (и не только у креативных работников) — несоответствие результатов взаимодействий запросам и желаниям в силу слабой разработанности инструментов для бо-

<sup>12</sup> МГПУ разрешил студентам использовать ИИ при подготовке ВКР // МГПУ. 2023. 31 августа. URL: <https://www.mgpu.ru/mgpu-razreshil-studentam-ispolzovat-ii-pri-podgotovke-vkr> (дата обращения: 24.03.2024).

<sup>13</sup> Sciences Po Bans the Use of ChatGPT without Transparent Referencing // Espace Presse Sciences Po. 2023. January 27. URL: <https://newsroom.sciencespo.fr/sciences-po-bans-the-use-of-chatgpt> (дата обращения: 24.03.2024).

лее сложных творческих задач, чем генерация шаблонных текстов, изображений или идей, а также значительного количества ошибок.

И хотя среди креативных работников доля выбравших роль «автора» и «со-автора» выше, чем в других группах, здесь нет единства: оптимистов много, как и не определившихся со своим мнением, однако наибольшая конкретика демонстрируется именно в пессимистичных ответах на ряд открытых вопросов. Часто отмечается сходство результатов самостоятельно выполненной работы и результатов генерации ИИ по промптам, многим в этом случае трудно принять постоянно развивающегося (быстрее, чем человек) ИИ-автора рядом с собой. В ответах на открытый вопрос «Предположите, какие креативные задачи могут потерять смысл, если при их выполнении использовать инструменты ИИ?» креативные работники высказывают серьезные опасения в отношении грядущей трансформации системы представлений о креативности и творчестве как таковых: *«использование ИИ напрямую в творчестве противоречит сути творчества, ИИ не создает качественно новые объекты»*, *«предполагаю, что личное творчество в таком случае теряет человеческую душу»*. В ответах на открытый вопрос «Как инструменты ИИ могут повлиять на рынок творческих профессий и креативные индустрии?» представители этой же группы формулируют опасения в отношении ближайшего будущего креативных практик в целом:

*Изменяются профессии. От самих создателей будет требоваться меньше креатива. Требования к навыкам человека для входа в креативные индустрии снизятся.*

*Возможно, при увеличении популярности нейросетей и частоты их использования, креативные специалисты способные на создание полностью уникального своего продукта будут больше цениться.*

*С одной стороны, процессы пойдут быстрее, будет больше проектов и идей. С другой стороны, пропадает уникальность и человечность.*

Размышляя о будущем в ответах на вопросы «Как Вы представляете себе идеальное регулирование использования инструментов ИИ в креативных практиках?» и «Для каких задач или в каких сферах Вы бы запретили дальнейшее использование и разработку инструментов ИИ?» представители креативных индустрий предлагают и решения вероятных проблем — от переосмысления авторского права до прямых запретов:

*Во-первых, стоит серьезно подойти к обсуждению и разработке этического кодекса для взаимодействия с ИИ в креативных индустриях. во-вторых, стоит частично изменить или пересмотреть некоторые моменты в авторском праве, адаптировать их под контекст ИИ.*

*Я бы запретила дальнейшее развитие в сфере креатива. Инструменты, которые мы имеем сейчас, уже в достаточной степени помогают, при этом оставляя место для человеческой деятельности.*

С нашей точки зрения, в ответах респондентов наблюдаются начавшийся слом и трансформация понимания концепций авторства и соавторства. Рост количества массово доступных инструментов ИИ провоцирует интерес к ним и повышение уровня принятия, при этом чем лучше разработка и качественнее результат в определенных задачах, тем чаще инструмент генеративного ИИ упоминают и используют.

Человек начинает перестраивать свои практики, делегируя ИИ задачи, с которыми тот уже может хорошо или сносно справляться, даже если результаты надо будет дорабатывать. Однако при этом у человека может снижаться мотивация совершенствовать соответствующие навыки и базовые компетенции у себя. Это порождает дихотомию в определении роли ИИ: с одной стороны, он облегчает и упрощает труд, и человек видит в нем «помощника», «ассистента» или «инструмент», с другой — задумывается о категории авторства в контексте материалов, сгенерированных ИИ, особенно если результат хорош. Очевидно, что ситуация требует дальнейших исследований восприятия людьми ИИ в разных ролях и выработки принципов их разграничения.

### **Наше будущее с цифровыми технологиями и ИИ: от аватаров и AI-тренеров до видеоигр и метавселенных**

Респондентам было предложено рассказать, какой новый инструмент генеративного ИИ и для каких задач они бы создали<sup>14</sup>. Ответы позволяют сделать ряд замечаний об особенностях восприятия и уточнить его вероятные роли в будущем. Во-первых, респонденты были достаточно реалистичны и не высказали ни одного желания из области научной фантастики, паранормального или сказочного:

*Нейросеть для воссоздания человеческого сна в формате видео. <...> Могла бы появиться новая интересная отрасль визуального искусства.*

*Переводчик с языка животных.*

Во-вторых, обнаружился запрос на ИИ-друга, ИИ-товарища, который способен помочь как в бытовых затруднениях, так и со сложными психологическими состояниями:

*Логичным было бы предложить создать ИИ-компаньона. Есть уже ИИ текстовый, ИИ генерирующий картинки, ИИ переводчик, но все еще ни один из них не может участвовать в диалоге с человеком с постоянно развивающейся или меняющимися темами, а учитывая растущее одиночество среди населения, логично создать компаньона, который подходил бы каждому с его психологическими проблемами и особенностями характера.*

В-третьих, укрепился запрос на ИИ-психолога и медицинскую помощь в целом (32 ответа, 9,14% от всех респондентов). Такие идеи чаще озвучивали студенты и креативные работники:

<sup>14</sup> Точная формулировка вопроса: «Какой новый инструмент ИИ и для каких задач Вы бы предложили создать?».

*Для медицины, какой-нибудь умный сканер, который заменит сразу несколько обследований и сократит нагрузку.*

*Голосовой помощник или чат для тревожных людей <...>, он бы при панических атаках сразу давал инструкцию к действию и успокаивал.*

*Консультант контроля ментального состояния для уменьшения количества психических расстройств в информационном обществе.*

Анализ ответов по отдельным группам позволяет сказать, что наиболее интересные, продуманные, неожиданные идеи были высказаны студентами и креативными работниками, кроме того, в этих группах предложения чаще основывались на конкретных ситуациях и задачах, в которых им требуется помощь. Преподаватели были более абстрактны в формулировках, даже касающихся своей профессиональной среды, предлагали что-то простое: «в помощь педагогу» и «для оценки результатов обучения». Существенное количество респондентов (42,85%) затруднилось ответить на вопрос, потому что «нет идей», «пока не освоены даже существующие», «все уже изобрели» и «осталось только внедрить эти инструменты в полезные места».

Два вопроса анкеты касались профессий и видов деятельности, которые могут появиться или исчезнуть по мере развития ИИ<sup>15</sup>. Высказанные предположения о «новых» видах деятельности можно разделить на три группы. В первой назывались уже возникшие профессии (промπτ-инженер, промпт-дизайнер, AI-тренер и пр.), для которых человек развивает принципиально новые навыки взаимодействия с ИИ, по-прежнему оставаясь субъектом в этих отношениях. Другая группа респондентов осознанно подошла к закрытию брешей, неизбежно появляющихся вместе с инновациями (работник сферы защиты ИИ от взломов, этический эксперт по теме ИИ и пр.). Третья группа (наибольшая из всех) формировалась простым сложением наименования существующей профессии и ИИ как объекта приложения усилий или тематической области (например, дрессировщик ИИ, редактор ИИ, репетитор ИИ, учитель ИИ, дизайнер ИИ, менеджер ИИ, копирайтер ИИ, ментор ИИ, разработчик приложений ИИ и т. д.). То есть варианты «будущих» профессий образовывались путем трансформации существующих, но в отношении нового объекта работы — ИИ. Роль человека оставалась чаще всего прежней, изменение функций профессий и их востребованность в будущем для ИИ формулировалась нечетко, что объясняется малым опытом респондентов — они чаще являются пользователями, чем специалистами, разрабатывающими инструменты ИИ.

*Думаю, что больше не новые профессии возникнут, а трансформируется функционал существующих с делания на контроль и корректировку как у Утесова: «я не извозчик, я водитель кобылы!», а будет «оператор ИИ „кобыла“».*

На вопрос об «уходящих» видах деятельности часть респондентов отвечала «почти все» или «все». Другая часть, которая формулировала ответ с точки зрения

<sup>15</sup> Формулировки вопросов: «Какие новые профессии и виды деятельности могут появиться с развитием технологий ИИ?», «Какие существующие профессии и виды деятельности могут исчезнуть с развитием технологий ИИ?».

не профессии, а роли человека, была настроена более нейтрально-позитивно, например:

*Не могут исчезнуть, только трансформироваться.*

*Никакая, ИИ не способны полноценно заменить человека.*

*Ну, можно предположить, что тот самый секретарь-референт исчезнет или вахтер, хотя я думаю, что профессия останется, функционал перераспределится между машиной и человеком, наверное, в каких-то профессиях количество требуемых людей сократится.*

Это, на наш взгляд, подчеркивает отношение части респондентов к ИИ в первую очередь как к инструменту, помощнику, по крайней мере в известных и уже опробованных взаимодействиях, тогда как прогнозировать глобальные изменения даже для ближайшего будущего уже сложнее.

Среди профессий, которые могут исчезнуть с развитием технологий ИИ, было много примеров из сферы услуг, производства массового контента, офисной работы и работы с текстами и изображениями: администратор, актер озвучки и дубляжа, бухгалтер, водитель, дизайнер, диспетчер, журналист, контент-менеджер, копирайтер, корректор, курьер, менеджер, оператор call-центров, переводчик, писатель, сотрудник почты, продавец, продюсер, редактор, секретарь, учитель, художник-иллюстратор, SMM-менеджер и т. д.

Такая подборка четко обозначает профессии, от которых респонденты готовы отказаться — как в силу более успешного выполнения «поточной», «рутинной», «дешевой» деятельности «с ручным трудом» ИИ, так и в силу собственных представлений о возможных видах такой работы. Это возвращает нас к главному пониманию респондентами ИИ как «помощника».

В вопросе об изменениях рынка творческих профессий и креативных индустрий респонденты больше фокусировались на темах роли человека-создателя, ценности собственного креативного творчества, а также контроля за технологиями. В ответах встречаются развилки возможных сценариев развития использования инструментов ИИ и не всегда у респондентов остается уверенность в главенствующей роли человека именно в задачах, связанных с «креативом».

*Можно ли назвать работы ИИ творчеством? Если нет, то люди-творцы не исчезнут и ничто не изменится. Если же можно, то все представители творческих профессий могут стать бесполезными.*

*Я думаю, это убивает человеческий креатив и поощряет использование ИИ в качестве упрощения творческой деятельности. ИИ может быть применен исключительно в упрощающих бытовую жизнь человека, но не профессиональную и, особенно, творческую.*

*Я считаю, что [повлияет] негативно. Душевная человеческая глубина исчезнет в творческих профессиях и креативных индустриях, а я считаю, что это очень важно.*

Интересно сравнение человеческого креатива с «душой» проектов и ее противопоставление неодушевленности ИИ как ключевому фактору недостижимости им уровня человека в этой области. Но сомнения насчет изменения роли и ценности вклада человека в творчество, спровоцированные развитием ИИ в креативной сфере, также растут. Пессимистичных и тревожных прогнозов становится больше: выраженный страх потери рабочих мест, снижение спроса на творческие профессии и уменьшение оплаты труда, переизбыток предложения и снижение качества проектов, эрозия уникальности и деградация человеческого креатива, боязнь снижения мотивации в развитии собственных навыков, неразрешенные вопросы с этикой и авторством. С другой стороны, респонденты пополнили и копилку положительных предположений: вдохновение, облегчение и ускорение работы, возможность заниматься более творческими задачами, избавившись от рутины, появление чего-то нового, необычного. Такая разнонаправленность может говорить о начале активной трансформации как самого понимания креативности, так и переосмысления ценности человеческого творчества, роли и места людей во взаимодействиях с ИИ.

Любопытным в контексте прогнозирования ближайшего будущего, особенно в сравнении с восприятием ИИ, оказалось понимание нашими респондентами метавселенных. Эти технологические феномены многое объединяет: схожие корни, траектории развития, инициаторы, бюджеты начальных этапов разработки. Обе технологии часто оказываются взаимосвязаны в научном [Soto, Leon, 2022] и индустриальном<sup>16</sup> дискурсах. Представления о метавселенной формируются под влиянием массовой культуры, в том числе некоторых художественных произведений, которые наши респонденты приводили в качестве источников информации об ИИ. Метавселенная встречается в числе ответов на вопрос «Для меня ИИ — это...» И хотя она не получила такого широкого распространения, как ИИ, — не только в России, но и в других странах, — выяснилось, что о ней слышали 60,57 % респондентов, при этом реальный опыт пребывания в ней был только у 4,57 %. Глубину понимания респондентами концепции метавселенных определить довольно трудно, но большая часть уверенно вербализует позитивно-нейтральное представление о них и даже фантазирует, чем в них можно будет заняться:

*Виртуальная вселенная, где каждый может быть кем угодно.*

*Постоянно действующее виртуальное пространство, аналог реального мира, в котором люди могут взаимодействовать друг с другом и с цифровыми объектами через своих персонажей, с помощью технологий виртуальной реальности. Спасибо книге «Первому игроку приготовиться», одной из любимых в детстве.*

*Новая реальность, в которой ты сам создаешь себя.*

<sup>16</sup> Value Creation in the Metaverse. The Real Business of the Virtual World. McKinsey & Company. 2022. URL: <https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/value-creation-in-the-metaverse> (дата обращения: 24.03.2024).

Главные ожидания связаны со свободой общения, разнообразием взаимодействий (от видеоигр до совместного создания чего-то нового и необычного) как друг с другом, так и с самим виртуальным миром через детализированно настраиваемых аватаров. Кроме легко прогнозируемых часто употребляемых выражений, таких как «виртуальные пространства», «виртуальная реальность», обращают внимание на себя слова, акцентирующие важность роли человека, взаимоотношений, активных действий и соучастия в метавселенной: «человек», «аватар», «друг с другом», «создавать», «возможность», «взаимодействовать», «игра», «общение», «мир» и пр.

Меньшая часть респондентов или не знает, что такое метавселенные, или же артикулирует их непонятность, глупость, хайп и бесполезность. Что неудивительно, учитывая все сложности не только при запуске проектов, но даже при формулировке единого конвенционального понятия в научной и профессиональной сферах. Но важно отметить отсутствие в ответах о метавселенных тревожности и формулируемого чувства опасности, ставших одним из лейтмотивов в похожих вопросах про ИИ. На наш взгляд, к этому привела разница в сценариях развития и презентации двух технологий.

Неудачный запуск глобальных проектов метавселенных в силу технических сложностей и проблем с контентным наполнением снизил внимание к ним со стороны бигтеха — быстрой прогнозируемой прибыли и масштабируемости (как в случае с ИИ) достичь не удалось. Но это не отпугнуло компании, интересующиеся перспективным цифровым расширением возможностей человека, которое к тому же не покусается на главенство во взаимодействии (в противоположность с ИИ), а скорее наоборот предлагает больше свободы и вариантов для самореализации. Сценарий внедрения метавселенных пошел по другому пути, более адаптивному и комфортному для человека. В большинстве случаев в проектах метавселенных вернулись к их сходству (до степени смешения) с видеоиграми, чья история длиннее и разнообразнее (с конца 1960-х годов), а распространенность выше (около 3 млрд игроков во всем мире в 2023 г.<sup>17</sup>). Более понятные, привычные и привлекательные видеоигры позволили проектам в метавселенных продолжить развиваться и находить свою аудиторию.

Именно времени и такой «мягкой» эволюции и постепенной адаптации в сознании пользователей не хватает в ситуации с ИИ. Ворвавшись на потребительский рынок, технология начала приносить деньги и продолжает развиваться с нарастающей скоростью, вызывая у людей фрустрацию и неуверенность в завтрашнем дне. Это заставляет задуматься о том, естественным ли путем инновация распространяется в обществе, и о возможных последствиях для тех, кто вынужден бежать за технологиями, еще не до конца понимая свои роль и место во взаимодействии с ними. Это только предстоит оценить и, безусловно, послужит темой для дискуссий как с теми, кто работает над созданием и внедрением инструментов ИИ, так и с потребителями.

## Заключение

Исследование демонстрирует неоднородность понимания ролей ИИ респондентами. Мы считаем правильным говорить о (1) макро- и микроуровнях восприятия

<sup>17</sup> Newzoo Global Games Market Report 2022 | Free Version // Newzoo. 2022. URL: <https://newzoo.com/resources/trend-reports/newzoo-global-games-market-report-2022-free-version> (дата обращения: 24.03.2024).

ИИ, (2) метафизическом и инструментальном, (3) воображаемом и контекстуальном, а также о (4) сходствах и различиях, обусловленных основной сферой деятельности: учебой, преподаванием или работой в креативных индустриях.

Макроуровень восприятия, понимания и артикуляции ролей ИИ базируется на идее неизбежности ИИ, который быстро стал неотъемлемой частью человеческой жизни и формируется в числе прочего медиадискурсом. Респонденты окружены роботами-доставщиками, умными колонками, чат-ботами и пр., если не буквально, то через материалы СМИ и социальных медиа. Не рефлексировав каждый фрагмент ИИ-технологий в отдельности, они как будто в той или иной степени воссоздают «общее настроение» или «общее мнение», транслируемое бигтехом, внедряющим ИИ (Microsoft, Alphabet, Amazon, Nvidia и др.), представителями различных индустрий, в которых сейчас внедряется ИИ (кинопроизводство, видеоигры, книгоиздание и др.), и теми исследователями, чей экспертный взгляд цитируется в публичном поле. На этом уровне обобщения и появляется идея об ИИ-помощнике, который упростит быт, решит проблемы, избавит от рутины и подарит множество возможностей. Микроуровень восприятия может совпадать с медийным лейтмотивом, это заметно в тех ответах, которые детализируют ситуации взаимодействия с ИИ, объясняя его роль, но может отличаться и проявляться в таких ассоциативных связках, как «халтурщик — профанация — читерство» или «постмодерн — миф — инфоповод». Микроуровневое восприятие чаще отражает сомнения, личный опыт взаимодействий или сферу интересов и увлечений, сквозь которую человек определяет для себя ролевую модель ИИ.

Признаки метафизического уровня были неоднократно нами замечены в дихотомических оценках (например, «благо и опасность») и философских абстракциях (например, «что-то новое и не до конца изведенное») по всему корпусу ответов. Это не означает, что отношение респондентов не основано на опыте или вообще беспочвенно, но на контрасте с инструментальным подходом к формулированию ролевых моделей ИИ (например, ИИ-инструмент, ИИ-компаньон) в высказываниях других респондентов первые действительно выглядят априорными.

Ряд ответов позволяет предположить, что роли, в которых респонденты видят ИИ, могут делиться на контекстуальные и воображаемые. Первые формируются в результате активного использования доступных инструментов ИИ в определенных ситуациях, вторые — в результате однократного опыта (не обязательно успешного) или как отражение некой внутренней потребности и предположения, что ИИ потенциально может помочь реализовать этот запрос. Однозначно определить, является ли роль ИИ воображаемой или контекстуальной, трудно, например, упоминание ИИ-психолога в одних случаях могла быть мечтой и фантазией, а в других — результатом успешных проб.

Сходства и различия в воспринимаемых ролях инструментов генеративного ИИ, обусловленные основной сферой деятельности респондентов, которых мы относим к креативному классу, рассматривались нами в контексте разделения на три группы: студенты, преподаватели и креативные работники. Выполненный анализ данных показал, что вне зависимости от принадлежности к группе основной ролью для ИИ респонденты считают «помощника», тогда как «очеловеченные» роли, особенно предполагающие работу с детьми, вызывают меньший отклик, чем «неодушевленные» или инструментальные.

Но больше в ответах разных групп было все-таки различий. Судя по ранжированию предложенных на выбор ролей, преподаватели наиболее консервативны и для осознания степени креативности своих действий или, например, более четкого формулирования ролей для ИИ в своих активностях им нужно больше времени, нового опыта и включенности в креативные практики, чем другим группам. Возможно, даже не только самим преподавателям, но и университетской среде, и академическому сообществу в целом. Креативные работники наиболее продолжительное время из всех трех групп используют инструменты генеративного ИИ и сейчас не только рефлексиируют, но и могут осознанно артикулировать больший спектр ролей: от «соавтора» до «музы». Студенты в этом смысле ближе к креативным работникам, хотя понимание ролей ИИ в этой группе больше интуитивное, чем осознанное, но имеющиеся широкие возможности для творчества в учебе стимулируют их использовать ИИ в большем спектре взаимодействий.

Исследование позволило зафиксировать ряд важных направлений для будущих дискуссий. Во-первых, именно в сфере креативных индустрий происходят самые заметные и быстрые изменения в связи с появлением ИИ, и, как следствие, креативные работники больше, чем остальные группы, рефлексиируют роли «автора» и «соавтора» применительно к ИИ. Этот опыт может оказаться полезен и в других сферах. Во-вторых, массовизация инструментов генеративного ИИ, по-видимому, запустила трансформацию самой системы креативности и наших представлений о человеке как единственно возможном творце и авторе. В-третьих, сфера высшего образования демонстрирует некоторую неповоротливость и растерянность в целом перед лицом новых технологий, рефлексия непосредственных участников процесса не так хорошо проговорена или не учитывается в силу разных причин, что только усугубляет разрыв между консервативными преподавателями и открытиями для экспериментов студентами. В-четвертых, распространение ИИ перераспределяет функциональные наборы в системе существующих и появляющихся технологий: это хорошо заметно по ситуации с такой ролью, как «поисковик». Раньше спрашивали у Google и «Яндекса», сейчас — у условных ChatGPT или «Алисы».

В завершение важно обратить внимание на необходимость развития и продвижения навыков и компетенций ИИ-грамотности. Например, последняя презентация<sup>18</sup> NVIDIA о ключевых технологических разработках начинается с перечисления ролей ИИ «от первого лица»: визионер, помощник, преобразователь, тренер, целитель, навигатор. Они последовательно раскрываются на примерах работы устройств с ИИ, программируемых или используемых человеком, но впечатление от агентности ИИ и фраз «я — визионер» или «я — целитель» уже сложно сгладить. Бережное отношение к личности человека уже сейчас отходит на второй план в угоду прибыли и желанию поразить воображение потенциальных потребителей. Обычному пользователю каждый день необходимо быть настороже и анализировать все большее количество информации (парадоксально, но с этим уже тоже лучше справляется ИИ), чтобы принимать самостоятельные решения, осознанно разграничивать свою роль и роль ИИ, критически относиться к пределам его и своих возможностей.

<sup>18</sup> NVIDIA GTC 2024 Keynote. Don't Miss This Transformative Moment in AI // NVIDIA. 2024. March 18. URL: <https://www.nvidia.com/gtc/keynote> (дата обращения: 24.03.2024).

## Список литературы (References)

1. Гершман М. А., Демьянова А. В., Полякова Е. Ю. Креативный класс России: портрет в цифрах. Научный дайджест. Спецвыпуск 2023. М.: НЦМУ ЦМИЧП, 2023. URL: <https://www.hse.ru/mirror/pubs/share/807171541.pdf> (дата обращения: 24.03.2024).  
Gershman M. A., Demyanova A. V., Polyakova E. Yu. Russia's Creative Class: A Portrait in Numbers. *Scientific Digest. Special Issue 2023*. М.: HCMRC, 2023. URL: <https://www.hse.ru/mirror/pubs/share/807171541.pdf> (дата обращения: 24.03.2024). (In Russ.)
2. Дробышева Е. Э. Высшая школа и креативные индустрии: возможности взаимодействия // Вестник Московского государственного лингвистического университета. Гуманитарные науки. 2018. № 14. С. 356—368. URL: [https://www.elibrary.ru/download/elibrary\\_37197582\\_86641671.pdf](https://www.elibrary.ru/download/elibrary_37197582_86641671.pdf) (дата обращения: 24.03.2024).  
Drobysheva E. Higher Education Institutions and Creative Industries: Opportunities of Intercommunication. *Vestnik of Moscow State Linguistic University. Humanities*. No. 14. P. 356—368. URL: [https://www.elibrary.ru/download/elibrary\\_37197582\\_86641671.pdf](https://www.elibrary.ru/download/elibrary_37197582_86641671.pdf) (дата обращения: 24.03.2024). (In Russ.)
3. Клоудова Й. Влияние развития креативной экономики на экономически отстающие регионы // Журнал Новой экономической ассоциации. 2010. № 5. С. 110—125. URL: <https://www.econorus.org/repec/journal/2010-5-110-125r.pdf> (дата обращения: 24.03.2024).  
Kloudova J. (2010) Expanse Creative Economy and Its Impact on the Less Developed Regions. *The Journal of the New Economic Association*. No. 5. P. 110—125. URL: <https://www.econorus.org/repec/journal/2010-5-110-125r.pdf> (дата обращения: 24.03.2024). (In Russ.)
4. Константинова Л. В., Ворожихин В. В., Петров А. М., Титова Е. С., Штыкно Д. А. Генеративный искусственный интеллект в образовании: дискуссии и прогнозы // Открытое образование. 2023. Т. 27. № 2. С. 36—48. <https://doi.org/10.21686/1818-4243-2023-2-36-48>.  
Konstantinova L. V., Vorozhikhin V. V., Petrov A. M., Titova E. S., Shtykhno D. A. (2023) Generative Artificial Intelligence in Education: Discussions and Forecasts. *Open Education*. Vol. 27. No. 2. P. 36—48. <https://doi.org/10.21686/1818-4243-2023-2-36-48>. (In Russ.)
5. Ним Е. Г. Студенческий дискурс о цифровом селф-трекинге: риторики и практики // Мониторинг общественного мнения: экономические и социальные перемены. 2020. № 2. С. 191—211. <https://doi.org/10.14515/monitoring.2020.2.989>.  
Nim E. G. (2020) Student Discourse on Digital Self-Tracking: Rhetorics and Practices. *Monitoring of Public Opinion: Economic and Social Changes*. No. 2. P. 191—211. <https://doi.org/10.14515/monitoring.2020.2.989>.

6. Платов А. В., Гаврилина Ю. И. Искусственный интеллект в образовании: эволюция и барьеры // Научный результат. Педагогика и психология образования. Т. 10. № 1. С. 26—43. <https://doi.org/10.18413/2313-8971-2024-10-1-0-3>.  
Platov A. V., Gavrilina Yu. I. (2024) Artificial Intelligence in Education: Evolution and Barriers. *Research Result. Pedagogy and Psychology of Education*. Vol. 10. No. 1. P. 3—15. <https://doi.org/10.18413/2313-8971-2024-10-1-0-3>. (In Russ.)
7. Флорида Р. Креативный класс. Люди, которые создают будущее. М.: Манн, Иванов и Фербер, 2016.  
Florida R. (2016). *The Rise of the Creative Class Revisited*. Moscow: Mann, Ivanov and Ferber. (In Russ.)
8. Чиксентмихайи М. Креативность. Поток и психология изобретений и открытий. М.: Карьера Пресс, 2015.  
Csikszentmihalyi M. (2015) *Creativity: Flow and the Psychology of Discovery and Invention*. Moscow: Career Press. (In Russ.)
9. Atkinson D. P., Barker D. R. (2023) AI and the Social Construction of Creativity. *Convergence*. Vol. 29. No. 4. P. 1054—1069. <https://doi.org/10.1177/13548565231187730>.
10. Chiu T. K. (2023) The impact of Generative AI (GenAI) on practices, policies and research direction in education: a case of ChatGPT and Midjourney. *Interactive Learning Environments*. <https://doi.org/10.1080/10494820.2023.2253861>.
11. Comunian R., Gilmore A., Jacobi S. (2015) Higher Education and the Creative Economy: Creative Graduates, Knowledge Transfer and Regional Impact Debates. *Geography Compass*. Vol. 9. P. 371—383. <https://doi.org/10.1111/GEC3.12220>.
12. Gardner H., Weinstein E. (2018) Creativity: The View from Big C and the Introduction of Tiny c. In Sternberg R., Kaufman J. (eds.) *The Nature of Human Creativity*. New York, NY: Cambridge University Press. P. 94—109.
13. Gruner D. T., Csikszentmihalyi M. (2018) Engineering Creativity in an Age of Artificial Intelligence. In: Lebeda I., Glăveanu V. P. (eds.) *The Palgrave Handbook of Social Creativity Research*. Palgrave Studies in Creativity and Culture. Palgrave Macmillan, Cham. P. 447—462. [https://doi.org/10.1007/978-3-319-95498-1\\_27](https://doi.org/10.1007/978-3-319-95498-1_27).
14. Kaufman J. (2009) Beyond Big and Little: The Four C Model of Creativity. *Review of General Psychology*. Vol. 13. No. 1. P. 1—12. <https://doi.org/10.1037/a0013688>.
15. Krätke S. (2010) 'Creative Cities' and the Rise of the Dealer Class: A Critique of Richard Florida's Approach to Urban Theory. *International Journal of Urban and Regional Research*. Vol. 34. No. 4. P. 835—853. <https://doi.org/10.1111/j.1468-2427.2010.00939.x>.
16. Lang J. (2024) Embracing Generative AI for Authentic Learning. *Creative Education*. Vol. 15. No. 1. P. 1—20. <https://doi.org/10.4236/ce.2024.151001>.
17. Lyall B., Robards B. (2017) Tool, Toy and Tutor: Subjective Experiences of Digital Self-Tracking. *Journal of Sociology*. Vol. 54. No. 1. P. 108—124. <https://doi.org/10.1177/1440783317722854>.

18. Markusen A. (2006) Urban Development and the Politics of a Creative Class: Evidence from the Study of Artists. *Environment and Planning A: Economy and Space*. Vol. 38. No. 10. P. 1921—1940. <https://doi.org/10.1068/a38179>.
19. Miljković J., Ljujić B., Mihajlović D. (2023) Obrazovanje i kreativne industrije. *Nastava i vaspitanje*. Vol. 72. No. 2. P. 219—236. <https://doi.org/10.5937/nasvas2302219m>.
20. Moore G. A. (1991) *Crossing the Chasm: Marketing and Selling High-Tech Products to Mainstream Customers*. New York, NY: HarperCollins Publishers.
21. Okello W. K., Quaye S. J. (2018) Advancing Creativity for Pedagogy and Practice. *Journal of Curriculum and Pedagogy*. Vol. 15. No. 1. P. 43—57. <https://doi.org/10.1080/15505170.2018.1437577>.
22. Purington A., Taft J. G., Sannon S., Bazarova N. N., Taylor S. H. (2017) 'Alexa is my new BFF': Social Roles, User Satisfaction, and Personification of the Amazon Echo. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. P. 2853—2859. <https://doi.org/10.1145/3027063.3053246>.
23. Rogers E. M. (2003) *Diffusion of Innovations*. 5th ed. New York, NY: Free Press.
24. Soto I. B. R., Leon N. S. S. (2022) How Artificial Intelligence Will Shape the Future of Metaverse. A qualitative perspective. *Metaverse Basic and Applied Research*. Vol. 1. P. 12. <https://doi.org/10.56294/mr202212>.
25. Sarkar A. (2023) Exploring Perspectives on the Impact of Artificial Intelligence on the Creativity of Knowledge Work Beyond Mechanised Plagiarism and Stochastic Parrots. In: *Proceedings of the 2nd Annual Meeting of the Symposium on Human-Computer Interaction for Work (CHIWORK 2023)*. New York, NY: The Association for Computing Machinery. Art. 13. P. 1—17. <https://doi.org/10.1145/3596671.3597650>.
26. Selkrig M., Keamy R. K. (2017) Creative Pedagogy: A Case for Teachers' Creative Learning Being at the Centre. *Teaching Education*. Vol. 28. No. 3. P. 317—332. <https://doi.org/10.1080/10476210.2017.1296829>.
27. Vinchon F., Lubart T., Bartolotta S., Gironnay V., Botella M., Bourgeois-Bougrine S., Burkhardt J., Bonnardel N., Corazza G. E., Glăveanu V. P., Hanchett Hanson M., Ivcevic Z., Karwowski M., Kaufman J., Okada T., Reiter-Palmon R., Gaggioli A. (2023) Artificial Intelligence & Creativity: A Manifesto for Collaboration. *The Journal of Creative Behavior*. Vol. 57. No. 4. P. 472—484. <https://doi.org/10.1002/jocb.597>.

# МОНИТОРИНГ МНЕНИЙ

## Правильная ссылка на статью:

Мониторинг мнений (ВЦИОМ): сентябрь — октябрь 2024 // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 85—99.

## For citation:

Public Opinion Poll (VCIOM): September — October 2024. (2024) *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 85–99.

## МОНИТОРИНГ МНЕНИЙ: СЕНТЯБРЬ — ОКТЯБРЬ 2024

Результаты ежедневных опросов «ВЦИОМ-Спутник». Метод опроса — телефонное интервью по стратифицированной двухосновной случайной выборке стационарных и мобильных номеров объемом 1600 респондентов в возрасте от 18 лет. Выборка построена на основе полного списка телефонных номеров, задействованных на территории РФ. Данные взвешены на вероятность отбора и по социально-демографическим параметрам. Максимальный размер ошибки с вероятностью 95 % не превышает 2,5 %. Помимо ошибки выборки, смещение в данные опросов могут вносить формулировки вопросов и различные обстоятельства, возникающие в ходе полевых работ.

## СОДЕРЖАНИЕ ДАЙДЖЕСТА

### ПОЛИТИКА

НОВЫЕ РЕГИОНЫ РОССИИ: ДВА ГОДА ПОСЛЕ ВОССОЕДИНЕНИЯ ..... 86

### СОЦИАЛЬНЫЕ И ЭКОНОМИЧЕСКИЕ ПРОБЛЕМЫ

ФРИЛАНС В РОССИИ: МОНИТОРИНГ ..... 87

МИКРОЗАЙМЫ И МИКРОЗАЕМЩИКИ ..... 90

РАЗВОДЫ В РОССИИ: МОНИТОРИНГ ..... 92

### ОБРАЗ ЖИЗНИ

ДРУЖБА И ДРУЗЬЯ: МОНИТОРИНГ ..... 95

НОВАЯ ЭПОХА ТРЕЗВОСТИ ..... 98

## ПОЛИТИКА

## НОВЫЕ РЕГИОНЫ РОССИИ: ДВА ГОДА ПОСЛЕ ВОССОЕДИНЕНИЯ ..... 86

## НОВЫЕ РЕГИОНЫ РОССИИ: ДВА ГОДА ПОСЛЕ ВОССОЕДИНЕНИЯ

24 сентября 2024 г.

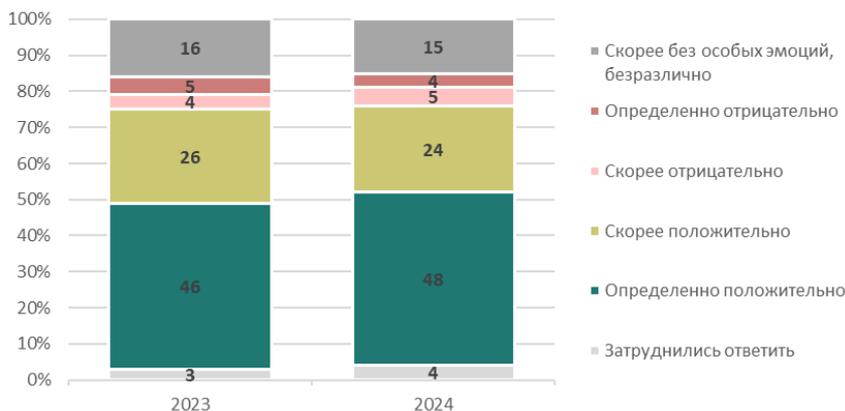
Два года назад в состав России вошли Донецкая и Луганская Народные Республики, а также освобожденные районы Херсонской и Запорожской областей. Сегодня осведомленность россиян о прошедших референдумах остается высокой, в общей сложности 92 % хорошо знают или слышали об этом (2023 г. — 91 %).

За два года отношение наших сограждан к вхождению Донецкой и Луганской Народных Республик и освобожденных районов Херсонской и Запорожской областей в состав Российской Федерации практически не изменилось. Большая часть по-прежнему позитивно воспринимает эти изменения — 72 % (2023 г. — 72 %). Спокойно, без эмоций оценивают присоединение новых территорий 15 % опрошенных (2023 г. — 16 %), об отрицательном отношении сообщили 9 % (2023 г. — 9 %).

Положительное отношение россиян к вхождению в состав страны Донецкой и Луганской Народных Республик и освобожденных районов Херсонской и Запорожской областей россияне объясняют, во-первых, мотивами воссоединения — 43 % («исконно русские земли» — 17 %, «там живут русские люди» — 10 %, «один народ» — 8 %, «выбор людей» — 5 %, «историческая справедливость» — 3 %). Второй мотив — забота о людях, их защита — 18 % (освобождение людей — 9 %, помощь, сострадание — 5 %, будет легче жить людям — 4 %). Эти результаты свидетельствуют о сильной эмоциональной и исторической составляющей в восприятии россиянами данного события.

Отрицательное отношение к присоединению новых территорий опрошенные аргументируют тем, что идут военные действия, и тем, что это территория другого государства (по 2 %).

Рис. 1. Скажите, как Вы относитесь к вхождению Донецкой и Луганской Народных Республик и освобожденных районов Херсонской и Запорожской областей в состав Российской Федерации? (закрытый вопрос, один ответ, % от всех опрошенных)



**СОЦИАЛЬНЫЕ И ЭКОНОМИЧЕСКИЕ ПРОБЛЕМЫ**

ФРИЛАНС В РОССИИ: МОНИТОРИНГ .....	87
МИКРОЗАЙМЫ И МИКРОЗАЕМЩИКИ .....	90
РАЗВОДЫ В РОССИИ: МОНИТОРИНГ .....	92

**ФРИЛАНС В РОССИИ: МОНИТОРИНГ***6 сентября 2024 г.*

Фриланс, еще недавно бывший нестандартной, инновационной формой занятости, уже стал привычным явлением на российском рынке труда. Вслед за ростом численности «независимых профессионалов» в нашем лексиконе укореняется и само понятие «фриланс»: за последние четыре года доля готовых дать ему определение выросла с 32% до 44%. Фигура фрилансера в общественном сознании оценивается скорее нейтрально-позитивно, упор в ответах делается на характер и формат занятости. В представлениях россиян фрилансеры — независимые, свободные и самозанятые работники, которые сами составляют свой график (15%), работают преимущественно удаленно / через интернет / на дому (13%) и на себя, а не «на дядю» (9%). Не знакомы с понятием «фриланс» чуть более половины наших сограждан — 56%, четыре года назад таких было 68%.

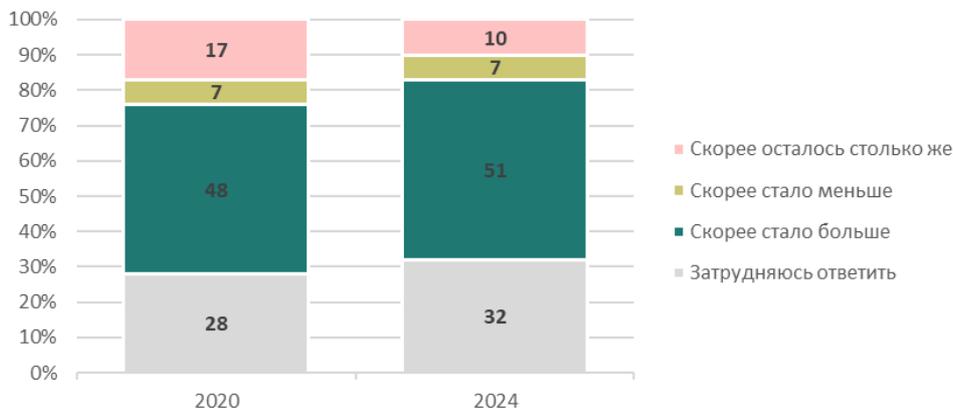
За последние три года число фрилансеров в нашей стране изменилось, считают 58% россиян. В подавляющем большинстве случаев опрошенные говорят об увеличении их числа — сегодня так думают 51%, или каждый второй. Заявили об обратном («скорее стало меньше») только 7%, тогда как каждый десятый отметил, что их численность за рассматриваемый период осталась прежней (10%). В целом сентябрьские оценки распространенности фриланса сопоставимы с оценками четырехлетней давности: в 2020 г. в обществе также доминировало мнение, что ряды фрилансеров в России пополнились (48%), 7% — что поредели. О неизменности ситуации говорили 17%, это немногим выше, чем сегодня.

Средой обитания фрилансеров принято считать интернет, в академическом сообществе они давно прослыли «цифровыми воротничками». Вероятно, поэтому рост их числа чаще заметен активным интернет-пользователям (58%) и приверженцам смешанной модели медиапотребления (ТВ + интернет, 50%), чем активным телезрителям (35%).

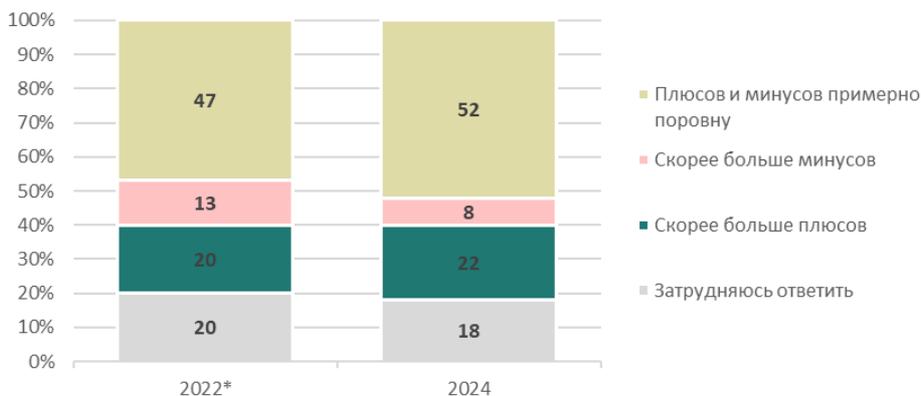
Фрилансеры — социальная группа, лишенная полной гарантированной занятости, а значит, и регулярных доходов, и социальной защиты. В связи с этим их нередко причисляют к прекариату. Впрочем, финансовая нестабильность, с которой сталкиваются «цифровые воротнички», — не повод считать эту категорию работников низкодоходной. Согласно полученным данным, за последние четыре года россияне только укрепились во мнении, что фриланс в среднем приносит больше дохода, чем работа по найму: сегодня так думают 38%, или на 7 п. п. больше в сравнении с 2020 г. Доля россиян, считающих, что их среднемесячный заработок идентичен, напротив, снизилась за рассматриваемый период с 26% до 18%. То же касается и тех, кто оценивает доходы фрилансеров ниже доходов штатных сотруд-

ников — 7% (–5 п. п. за четыре года). Еще 37% затруднились с ответом, из трех «закрытых» вопросов о фрилансерах этот оказался для россиян самым сложным.

**Рис. 1. По Вашему мнению, фрилансеров в последние три года в России стало больше, меньше или осталось столько же, сколько было? (закрытый вопрос, один ответ, % от всех опрошенных)**



**Рис. 2. На Ваш взгляд, в работе фрилансером больше плюсов или больше минусов, или плюсов и минусов примерно поровну? (закрытый вопрос, один ответ, % от всех опрошенных)**



\* Инициативный всероссийский интернет-опрос «ВЦИОМ-Онлайн». Объем выборки — 1806 работающих россиян в возрасте от 18 лет.

Более высокодоходным, чем «обычная работа», фриланс видится женщинам (41% vs. 35% мужчин), активным телезрителям (43%) и жителям Северо-Кавказского федерального округа (47%).

Каждый второй работающий россиянин считает, что плюсов и минусов в работе фрилансером примерно поровну (52%), за последние два года так стали думать немного чаще (+5 п. п. с 2022 г.). Прирост обеспечило снижение числа условных противников фриланса — тех, кто видит в нем больше минусов (8%, 2022 г. — 13%).

Больше преимуществ такой формы занятости, как и два года назад, видит каждый пятый работающий россиянин (22 %, 2022 г. — 20 %).

Несмотря на то что женщины чаще мужчин (здесь и далее речь о всех опрошенных, не только работающих) считают фриланс более высокооплачиваемым в сравнении с «обычной работой», именно мужчины чаще находят в нем больше плюсов (24 % vs. 18 % женщин). Оценки привлекательности работы фрилансером и ее доходности также связаны: о преобладании плюсов чаще говорят те, кто называет фрилансеров более высокооплачиваемой категорией работников, чем штатные сотрудники (29 %).

## МИКРОЗАЙМЫ И МИКРОЗАЕМЩИКИ

19—24 сентября 2024 г.

Микрозаймы — неотъемлемый элемент современной экономики, отвечающий потребностям многих людей в быстром доступе к небольшим финансовым средствам. Несмотря ужесточение денежно-кредитной политики в России, в последний год портрет заемщика достаточно стабилен.

Среди заемщиков чаще можно встретить женщин (57%), более четверти — в возрасте от 35 до 44 лет (29%), каждый второй имеет высшее образование (66%), 66% охарактеризовали свое материальное положение как среднее (+7 п. п. за год). Большая часть заемщиков работают по найму (56%), из них в должности специалиста — 57%. Около половины (55%) состоят в браке, почти столько же отметили, что не имеют несовершеннолетних детей (51%).

Среди потенциальных клиентов микрофинансовых организаций соотношение мужчин и женщин примерно одинаковое (49% и 51% соответственно), около половины (48%) старше 45 лет, большинство (73%) имеют высшее или неполное высшее образование. Шесть из десяти (63%) характеризуют свое материальное положение как среднее, большинство трудоустроены по найму (56%) и чаще (57%) в должности специалиста. Каждый второй (53%) находится в зарегистрированном браке и в 67% случаев не имеет детей до 18 лет.

Главные драйверы выбора микрофинансирования в качестве источника финансирования: скорость оформления (37%, +9 п. п. за год) и меньшее количество требований к пакету документов (32%, +12 п. п. за год). Около трети (30%, +8 п. п. за год) указали на такой вариант, как «отказали в кредите в банке», более четверти (27%, +7 п. п.) выбирают МФО потому, что ниже вероятность отказа. Выросла и привлекательность возможности оформить заем онлайн (27%, +6 п. п. за год).

Среди возможных причин обращения за микрозаймами у потенциальных заемщиков на первое место в этом году вышла низкая вероятность отказов. Сегодня этот вариант получил 36%, год назад его выбирали 16%, то есть рост более чем двукратный. Порядок других причин сохранился, но актуальность многих из них выросла: скорость получения займа, меньшее количество требований и документов (по 33%, +9 п. п.), отказ в кредите в банке (23%, +9 п. п.).

Микрозаймы становятся для многих альтернативным финансовым инструментом, более доступным и гибким, чем банки. Можно сказать, что микрофинансовые организации (МФО) отчасти компенсируют ужесточение кредитной политики традиционных финансовых институтов.

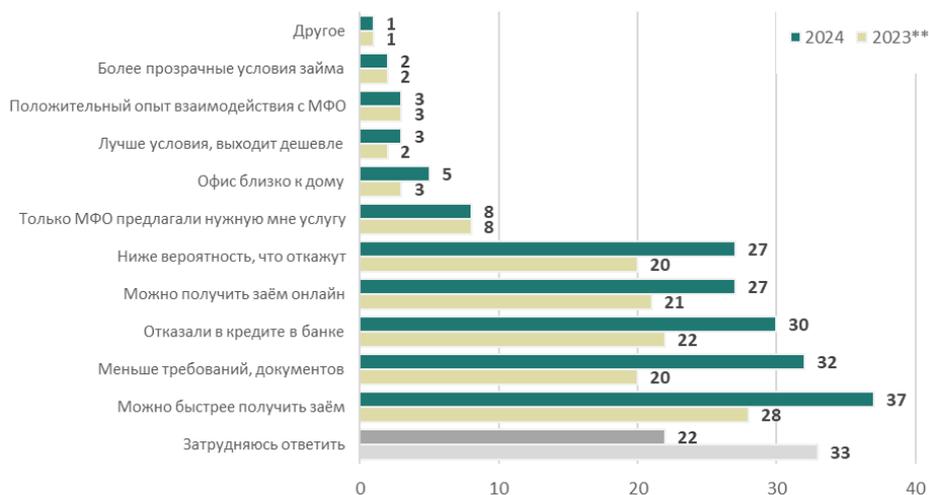
Организация приобретает статус микрофинансовой с момента включения сведений о ней в государственный реестр. Этот реестр представляет собой официальный список всех легально действующих микрофинансовых компаний, поднадзорных Центральному банку. О реестре МФО ЦБ РФ заемщики — как текущие, так и потенциальные, — осведомлены хорошо: знают о нем 9 из 10 (88% текущих заемщиков и 84% потенциальных). И для большинства опрошенных (71%) важно, чтобы МФО была включена в реестр ЦБ РФ.

Вероятно, именно поэтому половина текущих заемщиков декларируют, что проверяют статус организации, в которой они берут заем (52%), в том числе почти всегда это делают 21%, а 31% проверяют в случае, если не обращались в органи-

зацию ранее. Каждый четвертый (25 %) отметил, что никогда не проверяет юридический статус МФО. Это создает риски для заемщиков обратиться в нелегальную организацию, не имеющую государственной регистрации.

В целом заемщики используют широкий арсенал способов проверки МФО. Те, кто проверяет организацию перед оформлением займа, чаще всего смотрят отзывы в интернете (61 %) и реестр МФО на сайте Банка России (54 %), каждый третий пользуется информацией на сайте самой организации (33 %), каждый пятый (22 %) проверяет наличие маркировки сайта в поисковиках. Реже всего опрошенные обращаются за советом к близким (17 %).

**Рис. 3. Пожалуйста, отметьте три основные причины, почему Вы обращались в микрофинансовую организацию (МФО), а не в другую финансовую организацию (МФО), а не в другую финансовую организацию (закрытый вопрос, до 3 ответов, % от текущих заемщиков)**



\*\* Опрос 2023 г. проведен с 15 по 25 сентября тем же методом. Целевая выборка 2023 г. составила 1216 интервью. Доля реальных заемщиков в структуре выборки — 60 %, доля потенциальных клиентов, без опыта займов в МФО — 40 %.

## РАЗВОДЫ В РОССИИ: МОНИТОРИНГ

11 октября 2024 г.

Согласно результатам опросов ВЦИОМ, россиян, как и в прежние годы, больше привлекает создание семьи, нежели одиночество (84 % vs. 8 %). Чаще речь идет об официально зарегистрированном браке — 73 % (2017 г. — 78 %, 2019 г. — 77 %, 2021 г. — 71 %), в то время как сожительству без регистрации отдают предпочтение только 11 %, популярность такой формы отношений за последние семь лет несколько не выросла (2017 г. — 12 %, 2019 г. — 11 %, 2021 г. — 10 %). Фиктивный брак, заключаемый без намерения создать семью, россиян не привлекает, на протяжении всего периода наблюдений этот ответ набирал не более 2 % (2024 г. — меньше 1 %).

Чем старше респондент, тем чаще он выступает за традиционный зарегистрированный брак: максимум его сторонников зафиксировано в поколении оттепели — 85 % (vs. 62—64 % в двух самых младших поколениях). Молодежь в этом вопросе, напротив, чуть более раскрепощена, и даже, быть может, свободна от стереотипов: 16—18 % из числа зумеров и младших миллениалов допускают незарегистрированные отношения, 14—15 % вообще полагают, что в наши дни предпочтительнее жить одному. Представители самого старшего поколения оттепели не признают ни то, ни другое (1 % и 3 % соответственно).

Наряду со старшим поколением сторонников официального брака чуть больше среди женщин (75 % vs. 70 % мужчин), высокообразованной аудитории (77 %), а также россиян с хорошим и средним достатком (77 % и 74 % соответственно). Каждый седьмой опрошенный с доходом ниже среднего (плохое и очень плохое материальное положение) считает, что жить лучше одному или с партнером, но без регистрации брака (по 14 %).

Ответы жителей достаточно консервативного в вопросах семейных отношений региона нашей страны — Северо-Кавказского федерального округа — близки к средним по выборке, то есть их мнение ничем не отличается от мнения «большинства». Так, 73 % жителей этого федерального округа выбирают официальный брак, 14 % — соительство, 6 % — одиночество. Не имеет значения в этом вопросе и уровень урбанизированности: доли сторонников «традиционного» брака почти одинаковы в столичных городах и на селе (69 % и 73 % соответственно).

Динамика общественного восприятия обстоятельств, препятствующих расторжению брака, имеет нелинейный характер. На протяжении всего периода наблюдений нашим согражданам по-разному видятся барьеры развода. Но есть и то, что остается неизменным. С 2007 г. лидируют такие обстоятельства, как невозможность «поделить» детей между родителями (30 %, 2007 г. — 33 %), материальная зависимость одного из супругов (26 %, 2007 г. — 24 %) и сложности с разделом жилья (20 %, 2007 г. — 32 %). Наряду с имущественным вопросом за последние 17 лет менее убедительными аргументами для россиян стали: несогласие на развод одного из супругов (11 %, —9 п. п. с 2007 г.) и неприятная процедура развода (7 %, —5 п. п.), тогда как значение национальных и религиозных обычаев, напротив, выросло в глазах россиян (18 %, +8 п. п.). Осуждение близкого окружения видится препятствием еще для 11 %.

Более трети наших сограждан вовсе считают, что развестись можно всегда (37%), об отсутствии непреодолимых препятствий в последние пять лет россияне говорят заметно чаще, чем прежде (vs. 2007 г.— 23%, 2015 г.— 28%). Этот ответ наиболее популярен среди россиян с более богатым семейным опытом (от младших миллениалов до реформенного поколения, 42—44% vs. 15% зумеров), молодежь, напротив, заметно чаще называла конкретные обстоятельства, препятствующие расторжению брака. В частности, невозможность «поделить» детей между родителями (51% vs. 30% среди всех опрошенных), финансовую зависимость одного из супругов (37% vs. 26%), национальные или религиозные обычаи (44% vs. 18%) и др.

Рассуждать о факторах, осложняющих бракоразводный процесс, женщинам оказалось несколько проще, чем мужской части населения (затруднились с ответом 6% vs. 14% мужчин). В частности, россиянки чаще апеллируют к национальным и религиозным обычаям (21% vs. 15% мужчин) и материальной зависимости одного из супругов (31% vs. 20%).

В случае с разводами ситуация менее однозначна. В 1990 и 2007 гг. в российском обществе конкурировали два взгляда на точку невозврата в семейных отношениях: можно разводиться, когда семья фактически распалась (36—39%), и судить о необходимости развода следует исходя из конкретного случая (36—40%). Перелом во взглядах произошел в начале 2010-х: с 2013 г. в российском обществе доминирует ситуативный подход к проблеме — большая часть опрошенных придерживаются мнения, что все зависит от конкретного случая (2024 г. — 58%, +22 п. п. с 1990 г.). О том, что развод допустим, лишь когда от семьи фактически ничего не осталось, россияне говорят сегодня менее охотно (26%, —13 п. п.). По-прежнему редко встречаются среди наших сограждан и те, кто считает, что брак необходимо сохранить любой ценой (9%). Наконец, меньше всего поддержки находит радикальное решение семейных проблем («можно разводиться в любом случае, хуже от этого не будет» — 5%), на протяжении рассматриваемого периода доля ярых сторонников разводов варьировалась от 3% до 11%.

За сохранение брака любой ценой чаще выступают мужчины (12% vs. 6% женщин) и жители Северо-Кавказского федерального округа (19%). Чем старше респондент, тем выше поддержка этого суждения: в поколениях застоя и оттепели она достигает 12% и 14% соответственно (vs. 4% среди зумеров). Развод представители самого старшего поколения (оттепели) чаще поддерживают в ситуации, когда семейные отношения сошли на нет (32% vs. 14% зумеров). Молодые люди 18—23 лет, напротив, предлагают рассматривать каждый случай индивидуально (73% среди зумеров vs. 49% среди представителей поколения оттепели).

За последние пять лет представления россиян о причинах развода почти не претерпели изменений. Анализ ответов позволил выделить семь групп факторов, негативно сказывающихся на семейных отношениях. Как и в 2019 г., среди причин разводов лидируют финансовые трудности (46% ответов, 2019 г. — 38%), в частности безденежье (38%) и жилищные проблемы, высокие ставки по ипотеке (8%). Чаще всего этим аргументируют разводы россияне с плохим достатком (50% и 10% соответственно vs. 30% и 5% среди финансово благополучных). Вторая по популярности группа факторов — деструктивное поведение партнера (31%

ответов, 2019 г. — 22%), сюда можно отнести измены (18%), вредные привычки/зависимости (10%) и рукоприкладство (3%). Далее примерно с одинаковой частотой опрошенные называли нарушение коммуникации в семье (23% ответов, в том числе недопонимание — 14%, конфликты — 5%, неумение уступать друг другу — 4%) и утрату семейных ценностей (21% ответов, в том числе отсутствие уважения/нетерпимость — 6%, отсутствие доверия, внимания, чувств друг к другу — по 4% и др.). Встречаются среди россиян и те, кто считает причиной разводов несовместимость супругов (16% ответов, в том числе не сошлись характером — 9%, разные интересы / взгляды на жизнь — 4%, не подходят друг другу — 3%) и неготовность к семейной жизни (необдуманное вступление в брак — 7%, безответственность/инфантильность — 5%). Последний аргумент с 2019 г. стал для россиян более убедительным (12% ответов vs. 4% в 2019 г.). Наконец, еще одна группа причин — внешние обстоятельства (7% ответов, в том числе безработица — 5%, внутренняя политика/нестабильность — 2%).

Мужской и женский взгляды на причины разводов достаточно схожи. Однако женщины немногим чаще видят источник проблемы во вредных привычках партнера (13% vs. 8% мужчин), в безработице / нежелании партнера работать (6% vs. 3%) и рукоприкладстве (5% vs. 1%).

Мнения представителей разных поколений различаются сильнее. Чаще других к финансовым проблемам апеллируют четыре «средних» поколения — от младших миллениалов до поколения застоя (37—44%), в самой юной и самой возрастной группах этот ответ звучит заметно реже (27% и 29% соответственно). Что касается деструктивного поведения, то для молодежи веской причиной развода чаще видится неверность партнера (по 25% среди зумеров и младших миллениалов), для россиян 77+ лет — вредные привычки (17% среди поколения оттепели). Зумеры также чаще россиян в целом отмечают такие причины, как поспешное решение о вступлении в брак (16%), конфликты (14%) и отсутствие чувств (8%).

**ОБРАЗ ЖИЗНИ**

ДРУЖБА И ДРУЗЬЯ: МОНИТОРИНГ .....	95
НОВАЯ ЭПОХА ТРЕЗВОСТИ .....	98

**ДРУЖБА И ДРУЗЬЯ: МОНИТОРИНГ**

13 сентября 2024 г.

Согласно сентябрьскому опросу, друг в понимании россиян — это в первую очередь верный, надежный и преданный человек, который всегда поддержит и никогда не оставит в трудную минуту (61 %). Для кого-то друг — единомышленник, разделяющий взгляды и интересы (16 %), человек порядочный, честный, бескорыстный (11 %) и проверенный временем (7 %). В числе конкретных примеров друзей россияне называли члена семьи, родственника, сослуживца, соседа (8 %).

Согласно полученным результатам, у абсолютного большинства россиян есть в окружении те, кого можно назвать друзьями (83 %), с годами этот показатель остается почти неизменным (2013 г. — 80 %, 2023 г. — 79 %), чего нельзя сказать о круге общения. Если в 2013 г. россияне в среднем дружили с четырьмя людьми, в 2023 г. — с шестью, то сегодня число друзей выросло до восьми. То есть за 11 лет размер сети друзей увеличился вдвое.

О том, что россияне активизировались в поиске друзей, говорит и рост числа тех, кто за последние два-три года обзавелся новыми знакомствами: сегодня таких 40 % (в группе тех, у кого есть друзья), что почти в 1,5 раза больше в сравнении с 2013 г. (29 %). По словам опрошенных, их дружеское окружение за последние несколько лет в среднем пополнилось семью новыми знакомствами. Столь высокое значение может одинаково свидетельствовать как о коммуникативных навыках россиян, желании и умении быстро налаживать контакты и заводить друзей, так и о готовности называть другом хороших приятелей и знакомых. Как было показано выше, понятие друга весьма субъективно, называться таковым с одинаковым успехом может родственник или сослуживец. Другими словами, вполне допустимо, что в ответах россиян речь скорее идет не о новых друзьях, а о новых знакомствах, которые могут перерасти в нечто большее.

Большая часть опрошенных по-прежнему предпочитает дружить с людьми, проверенными временем: по словам 56 % россиян, имеющих друзей, ни одного нового друга за последние два-три года у них не появилось (–8 п. п. с 2013 г.).

Наличие друзей обусловлено многими факторами. В их числе возраст, уровень образования и материального положения, а также место проживания. Лучше всего получается у дружить молодых людей 18—24 лет (окружены друзьями 97 %), россиян с высшим и неполным высшим образованием (88 %) и хорошим достатком (87 %), проживающих в обеих столицах, нежели сельской местности (90 % vs. 77 %). Эти же группы (за исключением высокообразованных) можно отнести и к наиболее общительным: за последние два-три года новыми друзьями обзавелись 76 % молодых людей 18—24 лет (то есть вдвое больше, чем по выборке в целом, — 40 %), каждый второй опрошенный с хорошим материальным положением (48 %) и житель Москвы и Санкт-Петербурга (48 %).

Пользование всемирной сетью также повышает шансы найти друзей. О наличии последних заявили девять из десяти активных интернет-пользователей (89% vs. 69% активных телезрителей). Что касается новых знакомств, то ими могут похвастаться 44% «интернетчиков» и только треть любителей телевидения (32%).

Частота общения россиян с друзьями остается высокой: три четверти опрошенных, имеющих друзей, поддерживают с ними связь регулярно — не реже одного-двух раз в неделю (74%, 2013 г. — 78%, 2023 г. — 73%), в том числе 38% — практически ежедневно. Еще 17% общаются с друзьями раз-два в месяц, 6% — еще реже.

Больше всего времени своим друзьям уделяет молодежь 18—24 лет — 61% в этой группе общаются с ними на ежедневной основе. По мере взросления интенсивность дружеской коммуникации снижается: в соседней группе 25—34-летних показатель опускается до 46%, среди россиян постарше колеблется от 30% до 36%.

Рис. 1. Есть ли у Вас друзья? (закрытый вопрос, один ответ, в % от всех опрошенных)

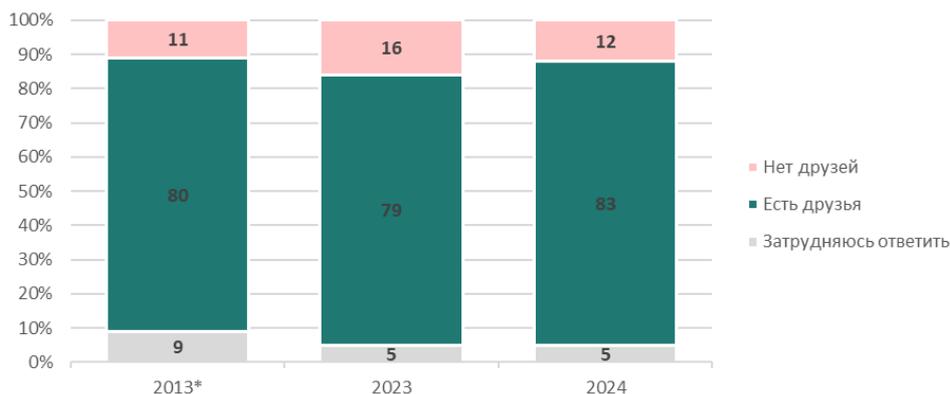


Рис. 2. Как часто Вы общаетесь со своими друзьями (встречаетесь, созваниваетесь, переписываетесь)? (закрытый вопрос, один ответ, в % от тех, у кого есть друзья)



\* До 2017 г. опросы проводились методом поквартирных face-to-face интервью (проект «Экспресс»), выборка стратифицированная многоступенчатая с квотами по социально-демографическим параметрам, репрезентует население РФ 18 лет и старше по типу населенного пункта, полу, возрасту, образованию и федеральному округу. Объем выборки составляет 1600 респондентов.

Несмотря на то что дружба входит в топ-5 наиболее важных ценностей для россиян, жизненные обстоятельства порой складываются так, что приходится прекращать дружеские отношения. За последние 11 лет доля россиян с подобным опытом выросла с 44 % до 68 %. Основная причина осталась неизменной — уход друга из жизни (24 %, 2013 г. — 20 %). Сходит дружба на нет и из-за переезда (8 %), предательства (7 %), различий во взглядах на жизнь и ссор, недопонимания (по 5 %).

Не сталкивались с потерей друзей четверть наших сограждан (24 %), вдвое меньше в сравнении с 2013 г. (50 %).

Согласно полученным результатам, дружить во взрослом возрасте становится сложнее — большую часть времени отнимают работа, семья и ежедневные заботы. С выходом на пенсию круг общения вовсе сужается, уходят из жизни друзья детства и молодости, из-за чего пожилые люди могут страдать от одиночества. По данным ВЦИОМ от 2021 г., именно старшее поколение 60+ лет чаще других испытывает одиночество на постоянной основе (7 % vs. 2—4 % среди россиян помладше).

В ходе сентябрьского опроса на тему дружбы 44 % россиян в возрасте 60 лет и старше признались, что пережили смерть друзей. Как результат, представители старшего поколения реже других могут похвастаться не только наличием друзей (75 % vs. 80—97 % в других возрастных группах), но и новыми знакомствами (33 %). Впрочем, заводить новых друзей становится проблематично и для россиян помладше: 35—59-летние в этом вопросе также проигрывают более общительной молодежи (33—36 % vs. 48—76 % среди 18—34-летних).

## НОВАЯ ЭПОХА ТРЕЗВОСТИ

27 сентября 2024 г.

В России наблюдается заметный сдвиг в культуре потребления алкоголя. Все больше людей демонстрируют приверженность трезвому образу жизни, что обусловлено совокупностью различных факторов — сменой поколений, активизацией антиалкогольной политики, модой и другими. Согласно результатам опроса ВЦИОМ, доля абстинентов (тех, кто вообще не потребляет спиртные напитки) в нашей стране почти сравнялась с долей текущих потребителей алкоголя (48% и 52% соответственно), тогда как 20 лет назад последних было чуть ли не втрое больше, чем трезвенников (73% vs. 27%).

Согласно имеющимся данным, слом советской, или «водочной», модели потребления спиртных напитков наблюдался уже в начале 2000-х. Пиву в 2004 г. отдавал предпочтение почти каждый второй — 47%. А вот вино и водку на заре XXI века россияне любили одинаково, их употреблял каждый третий (по 33%). В наше время за внимание потребителей по-прежнему борются эти три алкогольных напитка. Однако на фоне увеличения доли трезвенников их прежняя популярность снизилась: потребителей вина сегодня насчитывается 19% (–14 п. п. с 2004 г.), потребителей пива — 18% (–29 п. п.), водки — 11% (–22 п. п.). Другими словами, сегодня в российском обществе еще сильнее прослеживается переход к южному типу потребления алкоголя, характеризующемуся преобладанием легких алкогольных напитков, как правило пива и вина. Если в 2004 г. водка еще могла составить им конкуренцию, то сегодня — с большим трудом. В случае с другими крепкими спиртными напитками ситуация осталась неизменной: коньяк, виски, ром и джин как были в меньшем почете у россиян, так и остались.

Алкогольные предпочтения имеют четко выраженную гендерную специфику. «Мужскими» напитками можно назвать пиво (25% vs. 13% женщин) и крепкий алкоголь — водку (19% vs. 5%), виски (5% vs. 1%) и коньяк (9% vs. 5%), «женскими» — только вино (26% vs. 11% мужчин). Потребителей вина также отличают наличие высшего образования (26%) и хорошее материальное положение (24%).

Доля не пьющих алкоголь россиян за последние 20 лет выросла в 1,8 раза, с 27% до 48%. Отчетливее всего влияние строгих культурных и религиозных норм на паттерны потребления алкоголя можно наблюдать на примере Северо-Кавказского федерального округа — самого непьющего региона страны. Доля трезвенников там составила рекордные 71%, в то время как в других регионах России показатель колеблется от 45% до 50%.

Помимо региона проживания приверженность трезвому образу жизни обусловлена полом, уровнем достатка, местом работы. Россияне, не употребляющие алкогольные напитки, — чаще женщины (53% vs. 43% мужчин), представители двух младших (54—61%) и самого старшего поколений (56%), а также работающие в бюджетных организациях (51% vs. 39% среди работников коммерческих структур).

Чем старше россияне, тем более явно доносятся отголоски советской модели потребления алкоголя. Водку чаще всего пьют представители трех старших советских поколений — реформенного (1968—1981 гг.) — 15%, поколения застоя (1948—1967 гг.) — 16% и оттепели (до 1947 г.) — 16%. Примечательно, что для

россиян, которые выросли в эпоху Брежнева и в сознательном возрасте застали третью антиалкогольную кампанию в СССР, водка оказывается даже предпочтительнее пива (16 % vs. 10 %). Самые «пьющие» — представители реформенного поколения (61 %), которые в одинаковой степени предпочитают вино и пиво (21 % и 19 % соответственно).

Любителей пива больше среди зумеров, или цифрового поколения (2001 г. и позже), и старших миллениалов (1982—1991 гг.) — по 27 %. Однако в отличие от старших миллениалов среди представителей цифрового поколения заметно меньше потребителей вина (13 %) и практически отсутствуют потребители водки (1 %). Как было отмечено выше, наряду с младшими миллениалами (1992—2000 гг.) и старшим поколением в возрасте старше 77 лет зумерам более близок трезвый образ жизни — доля абстинентов в этих возрастных когортах варьируется от 54 % до 61 %.

Основным мотивом злоупотребления спиртными напитками россиянам по-прежнему видится желание снять стресс и напряжение (по 38 % в 2024 и 2004 гг.). Более четверти опрошенных причиной этой привычки считают отсутствие силы воли (27 %, 2004 г. — 29 %), каждый пятый говорит о потребности людей «залить» горе, беду (20 %, 2004 г. — 17 %). Менее убедительными причинами злоупотребления спиртными напитками за последние 20 лет для россиян стали безделье (26 %, —11 п. п. с 2004 г.), влияние ближайшего окружения (18 %, —5 п. п.) и неуверенность в завтрашнем дне (14 %, —11 п. п.). То же касается и так называемого мифа о «пьющей» России: сегодня распитие спиртных напитков заметно реже воспринимается как часть культуры — национальной традицией его могут назвать только 7 % (vs. 20 % в 2004 г.). Еще 11 % отмечают доступность алкоголя (2004 г. — 14 %). Несмотря на давний запрет рекламы спиртных напитков в СМИ и интернете, такой вариант отметили 4 %, или вдвое меньше, чем в 2004 г. (8 %).

Мнение представителей самого старшего поколения — поколения застоя (77+ лет) — идет вразрез с мнением «большинства». Ключевой причиной злоупотребления алкоголем в этой группе считают отсутствие силы воли, слабохарактерность (32 %), тогда как в остальных возрастных группах доминирует ответ о снятии стресса и напряжения, чаще его выбирают младшие миллениалы — 52 %. О влиянии ближайшего окружения старшее поколение, напротив, говорит в 2—2,5 раза реже россиян в возрасте 18—56 лет (20—24 % vs. 10 %). Неубедительным аргументом для них является и «запивание» горя (12 % vs. 28—32 % среди зумеров и младших миллениалов). Другими словами, взгляд на алкоголизм у представителей поколения застоя более «приземленный», попытки оправдать злоупотребление спиртными напитками они предпринимают реже остальных.

При этом потребление спиртных напитков практически не отражается на представлениях о причинах алкоголизма. Текущие потребители лишь немногим чаще «трезвенников» говорят о снятии стресса (44 % vs. 32 %), а «трезвенники» — об отсутствии силы воли (30 % vs. 25 % среди употребляющих спиртные напитки), доступности алкоголя (14 % vs. 9 %) и о влиянии рекламы (6 % vs 3 %).

DOI: [10.14515/monitoring.2024.5.2580](https://doi.org/10.14515/monitoring.2024.5.2580)



А. П. Казун

## МОЖЕТ ЛИ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ПРОГНОЗИРОВАТЬ РЕШЕНИЯ СУДА? СИСТЕМАТИЧЕСКИЙ ОБЗОР МЕЖДУНАРОДНЫХ ИССЛЕДОВАНИЙ

### Правильная ссылка на статью:

Казун А. П. Может ли искусственный интеллект прогнозировать решения суда? Систематический обзор международных исследований // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 100—122. <https://doi.org/10.14515/monitoring.2024.5.2580>.

### For citation:

Kazun A. P. (2024) Can Artificial Intelligence Predict Judicial Decisions? A Systematic Review of International Research. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 100–122. <https://doi.org/10.14515/monitoring.2024.5.2580>. (In Russ.)

Получено: 05.03.2024. Принято к публикации: 15.08.2024.

## МОЖЕТ ЛИ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ПРОГНОЗИРОВАТЬ РЕШЕНИЯ СУДА? СИСТЕМАТИЧЕСКИЙ ОБЗОР МЕЖДУНАРОДНЫХ ИССЛЕДОВАНИЙ

КАЗУН Антон Павлович — кандидат социологических наук, директор Института анализа предприятий и рынков, доцент департамента прикладной экономики, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия  
E-MAIL: akazun@hse.ru  
<https://orcid.org/0000-0002-0091-5388>

**Аннотация.** Развитие технологий искусственного интеллекта и появление открытых баз данных судебных решений привели к стремительному совершенствованию алгоритмов, позволяющих классифицировать юридические документы и прогнозировать принимаемые судьями решения. В статье мы анализируем корпус международных исследований, посвященных вопросу о том, насколько точно ИИ может предсказывать решения судей и, как следствие, сможет ли он в перспективе заменить судью-человека. Ответ на этот вопрос складывается из анализа двух ключевых аспектов — возможности и точности прогнозирования судебных решений, а также различных ограничений, связанных с применением ИИ.

Анализ международного опыта показывает, что в последние годы точность прогнозов выросла, однако качество моделей сильно зависит от специфики задач и доступных данных. Большинство исследований анализируют решения судов высшего уровня различных стран мира, что сильно снижает их прикладной потенциал в части работы с массовыми категориями дел. Кроме того, опасения вызывает использование моделей, действующих по принципу «черного ящика», поскольку их решения трудно интерпретировать. Несмотря на стремительное

## CAN ARTIFICIAL INTELLIGENCE PREDICT JUDICIAL DECISIONS? A SYSTEMATIC REVIEW OF INTERNATIONAL RESEARCH

Anton P. KAZUN<sup>1</sup> — *Cand. Sci. (Soc.), Director at the Institute for Industrial and Market Studies; Assistant Professor at the Department of Applied Economics*  
E-MAIL: akazun@hse.ru  
<https://orcid.org/0000-0002-0091-5388>

<sup>1</sup> HSE University, Moscow, Russia

**Abstract.** Advancements in artificial intelligence technologies and the emergence of open databases containing judicial decisions have led to rapid improvements in algorithms capable of classifying legal documents and forecasting decisions made by judges. This article examines a body of international research dedicated to how accurately AI can predict judges' decisions and whether it could potentially replace human judges in the future. The answer to this question is formed by analyzing two key aspects: the capability and accuracy of predicting judicial decisions and the various constraints associated with using AI.

Analysis of international experience shows that the accuracy of predictions has increased in recent years; however, the quality of the models depends greatly on the specificity of the tasks and the available data. Most studies analyze decisions from higher courts worldwide, significantly reducing their practical potential for dealing with mass categories of cases. Moreover, concerns have arisen regarding the use of models that operate on a “black box” principle, as their decisions are difficult to interpret. Despite the rapid development of AI technologies, the complete replacement of judges is unlikely because of the range of methodological limitations, including insufficient quality and volume of data, issues

развитие ИИ-технологий, полная замена судей вряд ли возможна в ближайшее время ввиду целого ряда методологических ограничений, включая недостаточное качество и объем данных, проблему интерпретируемости, сложность понимания юридического и культурного контекста, ограничения переноса на другие правовые системы. Однако ИИ-технологии возможно использовать для сокращения издержек по работе с материалами дела.

**Ключевые слова:** искусственный интеллект, предсказание судебных решений, машинное обучение, глубокое обучение, классификация юридических документов, точность алгоритмов

**Благодарность.** Исследование выполнено за счет гранта Российского научного фонда № 23-78-10073 «Разработка и апробация методики автоматизированного анализа текстов приговоров российских судов для социально-правовых исследований (на примере насильственных преступлений)» (см. подробнее: <https://rscf.ru/project/23-78-10073/>).

## Введение

Уже с 1950-х годов исследователи пытались количественными методами спрогнозировать решения судей [Kort, 1957], но до недавнего времени технологии позволяли достичь лишь весьма ограниченного результата в этой области [Ashley, Brüninghaus, 2009]. В последние годы с развитием технологий искусственного интеллекта (далее ИИ) и расширением доступности открытых судебных данных качество моделей, предсказывающих решения судей, существенно выросло [Medvedeva, Wieling, Vols, 2022]. Это привело к расширению дискуссии о том, может ли ИИ в обозримом будущем заменить профессионального судью или по крайней мере упростить работу судей [Shi, Sourdin, Li, 2021; Sourdin, Cornes, 2018; Taylor, 2023; Xu, 2022]. В том числе с 2017 по 2022 г. резко возросло количество связанных с юридической тематикой докладов, представленных на профильных конференциях по компьютерным наукам [Cui, Shen, Wen, 2023]. В настоящем исследовании мы попробуем на основе систематического анализа международного опыта описать возможности и риски, которые несет в себе появление технологии ИИ в сфере судебного производства, а также ответить на вопрос о том, возмож-

with interpretability, challenges in understanding legal and cultural context, and limitations in transferring models to other legal systems. However, AI technologies can be used to reduce the costs associated with case material handling.

**Keywords:** artificial intelligence, prediction of judicial decisions, machine learning, deep learning, legal document classification, algorithmic accuracy

**Acknowledgments.** The research was funded by the Russian Science Foundation grant No. 23-78-10073 “Development and approbation of the algorithm for automated analysis of the court decisions texts for socio-legal studies (based on cases of violent crimes)” (see more: <https://rscf.ru/project/23-78-10073/>).

но ли заменить судью на ИИ. Нужно отметить, что в статье мы будем говорить исключительно о «слабом ИИ», умеющем решать конкретные задачи (в нашем случае анализировать юридические тексты и прогнозировать исходы), оставляя в стороне вопрос о создании «сильного», или общего, искусственного интеллекта, способного воспроизводить весь мыслительный процесс человека.

Для юридической сферы обсуждение вопроса о внедрении ИИ, пожалуй, не менее релевантно, чем для медицины, поскольку в обоих случаях на кону нередко стоят жизни и здоровье людей. ИИ способен анализировать юридические тексты в таком объеме и с такой скоростью, которые не доступны человеку. Однако у внедрения ИИ в судебную сферу — или же, шире, в сферу правоприменения — «цена ошибки» может быть высока. Так, в фильме 2001 г. «Особое мнение» главный герой, работавший в корпорации, предсказывавшей тяжкие преступления до их совершения, сам стал жертвой работы этого алгоритма, когда его обвинили в преступлении, которого он не совершал. Хотя до подобного будущего нам далеко, с развитием ИИ возможность предугадывать преступления в статистическом смысле становится вполне реальной [Gerber, 2014]. Помимо обозначенных этических вопросов, повсеместное внедрение ИИ поднимает и другие. Например, такая практика создает проблему защиты конфиденциальной информации, а также повышает актуальность вопроса о приватности данных обычных граждан. Однако на другой чаше весов лежит возможность существенно сократить издержки по принятию судебных решений, а возможно, и повысить качество и беспристрастность правосудия.

В российском контексте обсуждение последствий внедрения ИИ в сферу правоприменения также релевантно, поскольку развитие данной технологии является одним из национальных приоритетов. В январе 2024 г. В. В. Путин поручил Верховному суду, Генпрокуратуре, Следственному комитету, МВД и Министерству юстиции РФ продумать возможность использования ИИ в расследовании преступлений<sup>1</sup>. В Послании Федеральному собранию от 29 февраля 2024 г. было дано поручение нарастить мощности отечественных суперкомпьютеров в десять раз к 2030 г.<sup>2</sup> Хотя первые инициативы в сфере правоприменения касаются борьбы с преступностью с применением видеокамер и технологии распознавания лиц, нет сомнений, что ИИ будет использован и в судебной практике. В частности, эксперименты по созданию «умных судов» уже проводятся в Китае [Rusakova, 2021; Shi, Sourdin, Li, 2021]. Некоторые российские судьи были скептически настроены к перспективе их замены ИИ<sup>3</sup>, однако на 2024 г. анонсирован запуск системы «Правосудие онлайн»<sup>4</sup>, которая сделает первый шаг к автоматизации типичных составляющих судебного процесса. Кроме того, летом 2024 г. судья Верховного суда РФ О. Зателепин высказал идею, что ИИ может внедряться в российскую систему

<sup>1</sup> Путин поручил улучшить использование ИИ для расследования преступлений // РБК. 2024. 17 января. URL: <https://www.rbc.ru/rbcfreenews/65a7a3359a794761a6913056> (дата обращения: 05.03.2024).

<sup>2</sup> Путин поручил нарастить мощность суперкомпьютеров в десять раз // РБК. 2024. 29 февраля. URL: <https://www.rbc.ru/rbcfreenews/65e07a3a9a79472386b14ad3> (дата обращения: 05.07.2024).

<sup>3</sup> В Совете судей отвергли идею замены судьи на ИИ // Право.ру. 2023. 26 октября. URL: <https://pravo.ru/news/249529/> (дата обращения: 05.07.2024).

<sup>4</sup> Куликов В. Суды планируют подключить искусственный интеллект к составлению решений // Российская газета. 2023. 25 мая. URL: <https://rg.ru/2023/05/25/robot-pomozhet-rassudit.html> (дата обращения: 05.07.2024).

правосудия для помощи судьям в принятии решений о наказании и для сокращения вероятности ошибок<sup>5</sup>, однако только в роли советника или помощника судьи.

Поскольку большинство российских судей перегружены работой<sup>6</sup>, внедрение ИИ-технологий для анализа материалов дел по меньшей мере могло бы способствовать высвобождению времени для содержательного знакомства с их ключевыми обстоятельствами. Для юристов и их клиентов появление подобных технологий существенно упростило бы подготовку к судебному разбирательству, а также дало бы возможность оценить шансы на успех [Jacob de Menezes-Neto, Clementino, 2022].

Статья структурирована следующим образом. Сначала мы опишем алгоритм поиска релевантной научной литературы, посвященной вопросу о предсказании судебных решений с помощью ИИ, и работу с ней. Затем последовательно ответим на несколько ключевых вопросов об использовании ИИ в классификации юридических текстов и предсказании решений, принимаемых судьями. Далее мы обсудим ограничения применимости данной технологии и рассмотрим ключевые этические вопросы, связанные с внедрением ИИ, а также сформулируем практические рекомендации.

### Подбор источников и исследовательские вопросы

Отбор источников для исследования осуществлялся через Research Rabbit<sup>7</sup> — ИИ-приложение, позволяющее строить сети цитирований между научными источниками. Преимущества данного подхода по сравнению с ручным поиском литературы по ключевым словам в базах научного цитирования Scopus и WoS состоит в том, что алгоритм выделяет ядро научной дискуссии, а также меньше зависит от выбора ключевых слов. В качестве отправной точки для построения сети источников использовалась выборка из систематического обзора [Medvedeva, Wieling, Vols, 2022], в котором были отобраны релевантные исследования за период с 2015 по 2021 г., посвященные вопросу о предсказании судебных решений. Использование Research Rabbit позволило актуализировать эту выборку, а также найти тексты по смежным тематикам, включая обсуждение вопросов этики. Точность отбора исследований дополнительно проверялась с помощью второго ресурса — Inciteful<sup>8</sup>. Он работает аналогичным образом и позволяет идентифицировать тексты, близкие по своему содержанию к имеющейся выборке. Используемый нами подход, как и поиск по ключевым словам, не исключает возможности упустить то или иное исследование, однако последнее точно не вносит высокого вклада в научную дискуссию ввиду отсутствия у него сильных взаимосвязей с основным корпусом литературы.

Итоговая сеть включала в себя 107 источников за период с 2004 по 2023 г., посвященных классификации судебных текстов, но большинство работ приходится на период с 2018 г., поскольку дискуссия по вопросам использования ИИ

<sup>5</sup> Бурнов В. ВС: ИИ может помочь снизить число ошибок при назначении наказания // РАПСИ. 2024. 1 июля. URL: [https://www.rapsinews.ru/judicial\\_mm/20240701/310059533.html](https://www.rapsinews.ru/judicial_mm/20240701/310059533.html) (дата обращения: 05.07.2024).

<sup>6</sup> Нагрузку на судей закрепят на законодательном уровне // Адвокатская газета. 2023. 06 декабря. URL: <https://www.advgazeta.ru/novosti/nagruzku-na-sudey-zakrepjat-na-zakonodatelnom-urovne/> (дата обращения: 05.07.2024).

<sup>7</sup> URL: <https://www.researchrabbit.ai/> (дата обращения: 05.07.2024).

<sup>8</sup> URL: <https://inciteful.xyz/> (дата обращения: 05.07.2024).

интенсифицировалась в последние годы. В целом полученная совокупность текстов охватывает все наиболее релевантные источники, посвященные вопросу о предсказании судебных решений. Для количественного анализа динамики качества предсказательных моделей было отобрано 34 источника за период с 2015 по 2023 г., сопоставимых по постановке задач и метрике оценки точности.

Проведенный анализ зарубежного опыта имеет ряд ограничений. Во-первых, мы не анализируем детально методологические вопросы, которым посвящен объемный корпус литературы в сфере компьютерных наук. В данной статье мы не ставим перед собой задачу описать существующие алгоритмы по классификации юридических текстов или же раскрыть способы их применения. Это ограничение в том числе позволило нам остановить построение сети в Research Rabbit на том моменте, когда новые смежные тексты перестали касаться предсказаний судебных решений и начали затрагивать связанные тематики в области математики и компьютерных наук. Во-вторых, мы не анализировали работы, посвященные предсказаниям в сфере правоприменения и преступности, фокусируясь только на стадии судебного дела. Это сделано для того, чтобы сохранить четкий фокус исследования и не уходить в другие сложные вопросы, такие как использование ИИ в качестве инструмента для поиска преступников, — эта тема представляет высокий научный интерес, но является предметом для отдельного анализа. В-третьих, мы не рассматриваем детально работы, моделирующие судебные решения на основе линейных регрессионных моделей, деревьев решений и других классических количественных методов, поскольку в последние годы эти исследования уступили дорогу более продвинутым методам, основанным на современных ИИ-технологиях. При этом в анализ были включены статьи, посвященные этическим аспектам применения ИИ в сфере правоприменения, которые не были напрямую связаны с прогнозированием решений.

Анализ источников литературы на русском языке показал, что, несмотря на наличие довольно большего числа исследований [Захарцев, Сальников, 2018; Коваленко и др., 2020; Кравчук, 2021; Степанов, Басаганов, 2022], они предлагают в основном постановку актуальных вопросов и анализируют юридические аспекты проблемы, но не систематизируют международный опыт использования ИИ-технологий для предсказания судебных решений. Таким образом, данная статья является первым в России систематическим анализом корпуса международных исследований, посвященных предсказанию судебных решений и перспективам использования ИИ в работе профессиональных судей.

На данный момент нет исследований, которые бы на основе одного из популярных алгоритмов глубинного обучения моделировали решения российских судей. Имеющиеся работы о факторах принятия судебных решений (например, [Volkov, 2016]) используют в анализе регрессионные модели. В проекте «Алгоритм света»<sup>9</sup>, а также в исследовании [Zhuchkova, Kazun, 2023] с помощью машинного обучения из текстов приговоров выделялась информация о партнерских или родственных отношениях между обвиняемым в убийстве и жертвой, что является, пожалуй, наиболее релевантным на текущий момент примером, относящимся к ИИ-техно-

<sup>9</sup> URL: [https://github.com/LanaLob/algorithm\\_sveta](https://github.com/LanaLob/algorithm_sveta) (дата обращения: 06.10.2024).

логиям. Однако основной анализ в последней работе также строился с использованием регрессионных моделей. Таким образом, настоящий систематический обзор может стать отправной точкой для появления первых отечественных исследований, прогнозирующих судебные решения.

Выборка источников позволит нам ответить на следующие ключевые вопросы. Во-первых, что именно ИИ может прогнозировать и на какие вопросы он способен дать ответ? Во-вторых, какой эмпирический материал (решения каких судов и в каких странах) брался за основу для обучения моделей? В-третьих, какими методами и подходами исследователи достигают своих результатов? Хотя третий вопрос не является ключевым для данного исследования, мы сделаем краткий обзор подходов, которые применяются для прогнозирования результатов судебных дел. В-четвертых, мы оценим точность предсказания судебных решений с помощью ИИ-технологий и поговорим об ограничениях, связанных с оценкой точности алгоритмов. Наконец, мы обсудим ограничения по использованию искусственного интеллекта, в том числе этические вопросы, связанные с его внедрением.

### **Что способен спрогнозировать ИИ?**

На первом шаге следует определить терминологию. Есть значительная разница между тем, как термины «прогнозирование» и «классификация» употребляются в широком контексте, и тем, как они используются в компьютерных науках, а значит, и в большинстве научных статей о прогнозировании судебных решений с помощью ИИ. Когда социальные исследователи говорят о классификации, они, как правило, имеют в виду сортировку по типам или разделение на группы. В большинстве статей про роль ИИ под классификацией понимается более широкий класс явлений, в который входит в том числе предсказание результатов. В работе [Medvedeva, Wieling, Vols, 2022] проводится полезное разделение исследований о предсказаниях судебных решений с использованием ИИ на три группы: идентификация результата, категоризация текстов и предсказание решений. О каждом из этих типов мы скажем далее, но важно, что в научной литературе все они могут называться классификацией.

Термин «прогнозирование» (или «предсказание») может иметь два значения. Если у нас есть массив уже вынесенных судебных решений, то мы можем обучить алгоритм, скажем, на 80 % этих дел, а протестировать его на оставшихся 20 %. В этом случае алгоритм будет предсказывать решения судей с некоторой точностью. Не будет ошибки в том, чтобы называть это предсказанием, однако оно не имеет никакого отношения к исходам будущих дел или тем решениям, которые реальные судьи еще только должны вынести. То есть мы имеем дело с предсказанием в «слабом» значении этого термина. Предсказание в «сильном» значении означало бы, что наша программа способна предсказывать реальные решения судей в будущем. Большинство моделирующих судебных решения исследований [Medvedeva, Wieling, Vols, 2022] понимают под предсказанием именно «слабую» версию этого термина.

Мы также будем говорить о предсказании преимущественно в «слабом» значении, используя в качестве синонима термин «прогнозирование». Прогнозирование всегда предполагает определенную вероятность и доверительные интервалы,

но нужно помнить, что точность таких предсказаний может не в полной мере отражать способность модели предсказывать будущие решения в реальных делах. В целом любой прогноз, будь то прогноз инфляции, результатов выборов или судебного решения, строится на данных о прошлом и исходит из предпосылки, что в будущем действия людей (в том числе профильных специалистов) будут определяться тем же набором факторов, что и ранее. В реальности чаще всего так и происходит, что делает прогнозы в целом важным ориентиром при принятии решений, однако всегда остается риск появления новых факторов, которые «ломают» предыдущие закономерности и делают прогнозы, основанные на прошлых данных, менее надежными.

Как было отмечено выше, весь корпус литературы о классификации судебных решений можно разделить на три типа. Первый тип — поиск в тексте документа вердикта или резолютивной части. Эта задача не представляет никакой сложности для юриста-человека. Машинное обучение в данном случае применяется для первичной обработки большого объема данных и экономии времени. В этой задаче цель исследователей и программистов состоит в том, чтобы приблизить качество работы машины к результатам, которые может дать человек. Исследования, ставящие перед собой задачу идентификации результата, вполне могут достигать точности 99 %, что означает не предсказание решений судьи, а лишь способность выделить из документа место, где решение описано. В дальнейшем мы не будем относить исследования этого типа к предсказаниям судебных решений.

Второй тип исследований — категоризация юридических текстов. Задача состоит в выявлении релевантных обстоятельств дел или их характеристик, например пола обвиняемого или жертвы, обстоятельства совершения преступления и пр. Хотя на первый взгляд кажется, что второй тип задач близок к первому, категории могут строиться не на явных текстовых фрагментах, а на анализе всего документа, например, если оценивается идеологическая ориентация суда [Shaikh, Sahu, Anand, 2020]. В случае с выделением экстралегальных факторов ИИ уже может соперничать с человеком, поскольку человек не всегда может со стопроцентной точностью провести категоризацию. Однако к предсказанию судебных решений относится не каждая такая задача, а лишь та, что связана с категоризацией решений — вынесен приговор или нет, одобрена жалоба или нет, и т. д.

Наконец, третий тип задач является собственно предсказанием исхода — будет приговор оправдательным или обвинительным, признает суд нарушение прав человека или нет и т. д. Кроме того, предсказывать можно и размер наказания, будь то штраф или срок лишения свободы. В различении между вторым и третьим типом есть тонкая граница. Исследователи [Medvedeva, Wieling, Vols, 2022] относят к третьему типу только те немногие модели [Medvedeva et al., 2021; Sharma et al., 2015; Waltl et al., 2017], которые предсказывают решения судей на основе данных, доступных до вынесения решения (то есть такие модели близки к «сильному» определению термина предсказание). Предсказания же в «слабом» значении попадают во вторую категорию. Ввиду немногочисленности исследований третьего типа дальше мы будем относить к моделям, предсказывающим решения судей, исследования, попадающие как во вторую, так и в третью группу.

## Решения каких судов уже прогнозирует ИИ?

Большинство исследований о предсказаниях судебных решений выполнены на данных Европейского суда по правам человека (далее — ЕСПЧ) [Chalkidis, Androutsopoulos, Aletras, 2019; O’Sullivan, Beel, 2019; Medvedeva, Vols, Wieling, 2020; Aletras et al., 2016; Kaur et al., 2019; Medvedeva et al., 2020; Quemy, Wrembel, 2020] и Верховного суда США [Kaufman, Kraft, Sen, 2019; Katz, Bommarito, Blackman, 2017; Sharma et al., 2015]. Базы данных решений этих судов находятся в открытом доступе, что позволяет тестировать на них новые алгоритмы. Остальные исследования также, как правило, фокусируются на решениях судов высших инстанций — конституционного суда Турции [Sert, Yildirim, Haşlak, 2022], верховного суда Филиппин [Virtucio et al., 2018], Тайваня [Kowsrihawatt, Vateekul, Boonkwan, 2018], Бразилии [Freitas et al., 2022] и Индии [Malik et al., 2021].

Существенно меньше исследований делаются на больших выборках судов общей юрисдикции или же судов по отдельным вопросам из различных стран мира (например, по налогам [Alarie, Niblett, Yoon, 2017; Waltl et al., 2017], по коммерческим спорам [Bagherian-Marandi, Ravanshadnia, Akbarzadeh-T, 2021]). В последнее время все чаще создаются новые открытые данные [Cui et al., 2023]. Самая крупная и известная из публичных баз данных — Cail2018 [Zhong et al., 2020], включающая 2,6 млн кейсов, рассмотренных Верховным судом Китая. Следует отметить, что лишь немногие исследования (например, [Jacob de Menezes-Neto, Clementino, 2022]) работают с данными судов нижних инстанций.

Открытые базы данных ([Alali et al., 2021; Xiao et al., 2018; Malik et al., 2021; Sebők, Kiss, Járay, 2023]) позволяют тестировать новые алгоритмы анализа судебных дел, что, в свою очередь, служит для исследователей ориентиром по качеству ИИ-моделей. Чтобы доказать, что новая модель лучше, достаточно провести тесты на одном из доступных массивов данных и превзойти последний лучший результат.

## Как ИИ прогнозирует решения?

В задачу нашего обзора не входит детальное описание методологии работы с данными, поскольку применяемые способы анализа судебных кейсов многочисленны и разнообразны, а также требуют специальных навыков для освоения. Однако ниже мы кратко опишем основные типы подходов, которые используются в исследованиях, что позволит нам лучше понимать связанные с ними возможности и ограничения.

Снова несколько слов нужно сказать о терминологии [Sert, Yildirim, Haşlak, 2022: 8]. ИИ, машинное обучение и глубокое обучение можно соотносить, используя метафору матрешки, поскольку они «вложены» друг в друга. ИИ — это самое широкое понятие, в которое попадает множество методов анализа данных. Обработка естественного языка (Natural Language Processing или NLP), позволяющая компьютеру «читать» и «понимать» человеческий язык, является одним из ключевых способов использования ИИ, но далеко не единственным. В случае с прогнозированием судебных решений мы имеем дело именно с NLP. Следующий уровень — машинное обучение, которое представляет собой составную часть ИИ. Дальше — глубокое обучение (deep learning), один из вариантов машинного обучения. При использовании любого из методов тексты судебных решений, как правило, тре-

буют предобработки. Предобработка данных, в свою очередь, может строиться на различных методах машинного обучения и иметь разную степень точности.

Основное различие в методах предсказания судебных дел проходит между глубоким обучением и другими, как правило, более простыми методами машинного обучения [Alcántara Francia, Nunez-del-Prado, Alatrística-Salas, 2022]. Один из стандартных видов машинного обучения — обучение с учителем (supervised learning). Идея обучения с учителем состоит в том, что алгоритм тренируют на некоторой выборке дел, содержащей параметры, связанные с результатами, то есть решениями судей. Качество обучения алгоритма проверяют уже на другой, но аналогичной по структуре выборке, в которой содержатся только параметры дел — алгоритм же должен предсказать результат. Точность прогноза в данном случае равна проценту корректно предсказанных решений, что мы подробнее обсудим в следующем разделе. Такие традиционные формы машинного обучения, как логистическая регрессия или деревья решений, хорошо поддаются интерпретации, хотя некоторые более сложные модели (например, случайный лес) представляют собой «черные ящики», которые интерпретировать сложнее.

Методы глубокого обучения используют нейронные сети со множеством слоев, что позволяет обучать модели на больших данных. Такие модели могут работать с миллионами параметров (выделить которые вручную невозможно), поэтому они отличаются от других методов машинного обучения сложностью моделей, большими объемами данных и использованием существенных вычислительных ресурсов. Результаты глубокого обучения, в отличие от результатов традиционных методов машинного обучения, интерпретировать сложно или же невозможно, что создает серьезные ограничения, поскольку в юридической сфере интерпретируемость имеет решающее значение.

В настоящее время для классификации юридических текстов и предсказания судебных решений чаще всего используют предварительно обученные языковые модели [Song et al., 2022], такие как BERT [Devlin et al., 2019], относящиеся к моделям глубинного обучения. Этот метод предполагает, что модель сначала обучают на большом количестве текстов, а затем используют в различных новых задачах. Этот подход существенно отличается от более ранних методов [Ashley, Brüninghaus, 2009], в которых алгоритм создавался специально для конкретной задачи и работы с определенными массивами данных. Такие алгоритмы, как BERT, более универсальные, но одновременно и ограниченные ввиду того, что могут недоучитывать специфику юридических текстов. Однако их можно обучить дополнительно и создать специализированную версию для решения определенных задач, такую как LegalBERT, обученную на юридических текстах [Chalkidis et al., 2020]. Как показывают исследования, специализированные языковые модели могут давать на 1—5% более высокие показатели точности, чем общие [Song et al., 2022], но их обучение, конечно, связано с дополнительными издержками для исследователя. Упомянутые BERT и LegalBERT являются лишь двумя актуальными примерами предварительно обученных языковых моделей, однако есть и множество других. В задачи настоящего исследования не входит более глубокий анализ особенностей разных моделей, с которыми можно ознакомиться в работах [Alcántara Francia, Nunez-del-Prado, Alatrística-Salas, 2022; Song et al., 2022].

Отметим, что технология генеративного искусственного интеллекта ChatGPT также относится к классу предварительно обученных языковых моделей, хотя основная задача ChatGPT состоит в генерации текста, а не в классификации. Потому BERT и аналогичные модели все еще лучше отвечают задачам классификации юридических текстов и прогнозирования судебных решений, чем ChatGPT, который, в свою очередь, лучше приспособлен для составления юридических документов.

### С какой точностью можно прогнозировать решение суда?

Как было отмечено выше, под классификацией результатов могут пониматься разные задачи — от точности выделения в тексте приговора резолютивной части до предсказания реальных решений, которые еще не были вынесены. При этом сами результаты могут быть различными: виновен или не виновен, жалоба истца принята или отклонена [Bagherian-Marandi et al., 2021], апелляция успешна или нет [Jacob de Menezes-Neto, Clementino, 2022; Waihl et al., 2017], права признаны нарушенными или нет [Sert, Yıldırım, Haşlak, 2022], и т. п. Как правило, точность прогноза оценивается для бинарных переменных, хотя размеры штрафов или сроки приговора, пусть с меньшей точностью, также можно предсказывать.

Исследователи выделяют несколько ключевых метрик для расчета точности [ibid.]: *Accuracy* (может быть переведена как аккуратность, но именно эта метрика чаще всего отражает точность), *Precision* (точность), *Recall* (отзыв) и *F1 Score*. *Accuracy* — это сумма правильно классифицированных позитивных и негативных решений, поделенная на их общее число. *Precision* и *Recall* означают соответственно долю верно предсказанных позитивных случаев среди всех предсказанных и долю верных позитивных предсказаний среди всех реальных позитивных случаев. *F1 Score* — это индекс, который часто используется для сравнения качества различных моделей. Он рассчитывается следующим образом:

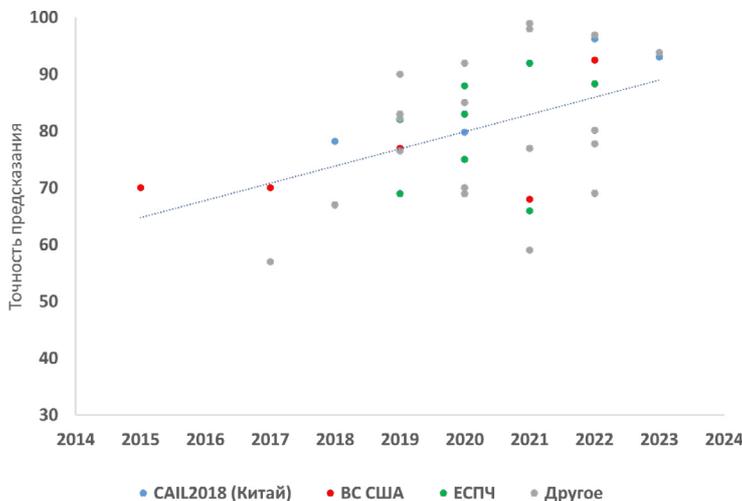
$$F1\ Score = 2 \times \frac{Precision \times Recall}{Precision + Recall}$$

При сравнении точности различных исследований нужно учитывать, какой из параметров авторы выводят в качестве точности — *Accuracy* или *F1*, поскольку это может давать различия от нескольких десятых процента до нескольких процентов. Хотя большинство работ используют эти индексы, в последнее время их критикуют за то, что они могут давать слишком оптимистичные результаты [Chicco, Jurman, 2020]. В качестве замены им предлагается *Matthews correlation coefficient* (MCC), который дает хорошие результаты лишь в том случае, если алгоритму удастся показать себя эффективно во всех четырех вариантах: верный позитивный результат, верный негативный результат, ложный позитивный результат и ложный негативный результат.

Точность алгоритмов находится обычно в диапазоне от 60% до 99%, однако более низкий показатель точности не всегда означает, что алгоритм плохой: возможно, он хороший, просто с его помощью решалась более сложная задача. По этой причине дескриптивный анализ, который мы предлагаем ниже, нужно воспринимать именно как иллюстрацию базового тренда и разброса в точности предсказания результатов судебных дел, сделанных в существующих работах.

Анализ международного опыта по отобранным нами 34 статьям позволил оценить динамику предсказательной силы моделей, классифицирующих тексты судебных решений (см. рис. 1).

Рис. 1. Динамика точности моделей, предсказывающих решения судов с 2015 по 2023 г. (для различных баз данных)<sup>10</sup>



Получившаяся картина отражает общий возрастающий тренд в предсказательной силе моделей. Однако график одновременно предлагает и весьма упрощенную картину, поскольку в разных исследованиях ставились различные цели, применялись разные алгоритмы анализа и разные оценки точности. Например, исследования [Jacob de Menezes-Neto, Clementino, 2022] на графике нет, поскольку в нем использовался альтернативный метод оценки точности, несопоставимый напрямую с другими исследованиями.

С методологической точки зрения лучше сравнивать точность решения одинаковых задач, выполненных на одних и тех же данных, таких как база решений ЕСПЧ, Верховного суда США или база данных Cail2018. Эти популярные открытые массивы данных публиковались вместе с бенчмарками (ориентирами), которые задавали для всех будущих алгоритмов базовый уровень качества. Если сравнивать прогресс внутри отдельных баз данных, то также заметен возрастающий тренд. Например, точность предсказания по китайской базе данных Cail2018 выросла с 78% [Xiao et al., 2018] в 2018 г. до 96% в 2023 г. [Cui et al., 2023]. В 2015—2017 гг. алгоритмы предсказывали решения Верховного суда США с точностью

<sup>10</sup> Примечание: график построен автором на основе 38 оценок, выполненных в рамках 34 исследований. Из них 19 наблюдений в готовом виде взяты из систематического обзора [Medvedeva, Wieling, Vols, 2022], остальные собраны и закодированы автором. В выборку включены только оценки для решений судов, предполагающих решения с двумя исходами (1 или 0). Вертикальная ось — значения точности (Accuracy); для некоторых исследований, где отсутствует значение Accuracy, приведено значение F1; горизонтальная ось — год публикации исследования. Синяя линия на графике — линия тренда для всех наблюдений.

70% [Katz, Bommarito, Blackman, 2017; Sharma et al., 2015], а в 2022 г. появились модели, предсказывающие их на 92,5% [Alghazzawi et al., 2022]. Алгоритмы, моделирующие решения ЕСПЧ, также прогнозируют решения с точностью до 92% [Medvedeva et al., 2020]. Но, как отмечалось ранее, речь идет о предсказании в его «слабой» форме, то есть о качестве модели, обученной на уже имеющихся решениях и проверенной на них же. Эти модели будут давать адекватный прогноз о будущем только при условии, что внешние обстоятельства, такие как законодательство, состав суда, политическая обстановка и пр., не поменяются значимым образом.

Еще один интересный вопрос о точности предсказаний связан с выбором точки для сравнения. Естественным выбором кажется сравнение модели с реальными решениями. Такой подход использовался во всех исследованиях, включенных в рисунок 1. Однако есть и альтернативная точка зрения: чтобы понять, может ли ИИ заменить юриста, корректнее делать сравнение с предсказательными способностями реального человека, который является экспертом в той или иной области. Например, в работе [Jacob de Menezes-Neto, Clementino, 2022] на данных судов о более чем 765 тыс. кейсов из Бразилии сравнивалась точность предсказания исхода апелляций, которую может показать ИИ или 22 эксперта из числа судей и сотрудников судов (оценивших случайную выборку в 690 кейсов). Согласно результатам, способности ИИ превосходят способности человека почти в три раза. Впрочем, это не первое исследование, которое показывает, что ИИ может превзойти человека. В исследовании [Ruger et al., 2004] машина также превзошла экспертов-юристов в точности предсказания решений Верховного суда США, правда, не со столь высоким отрывом.

Из сказанного можно сделать несколько выводов. Во-первых, точность работы алгоритмов по классификации судебных решений сильно зависит от данных и от задач, которые ставятся перед исследователями. Во-вторых, в последние годы наблюдается явный тренд на повышение качества предсказательных моделей. При сравнении с предсказательными способностями экспертов-юристов некоторые алгоритмы демонстрируют явное превосходство над человеком. Впрочем, о стопроцентной точности, которая позволила бы заменить судью, пока говорить тоже не приходится.

### **Дискуссия: может ли ИИ заменить судью?**

Прежде чем ответить на главный вопрос нашего исследования, суммируем основные ограничения в использовании ИИ в судебной практике.

Во-первых, результаты моделей сильно зависят от качества данных. Чтобы модель могла давать качественные предсказания, тексты судебных решений должны иметь похожую структуру, а также их должно быть достаточно много для обучения. Поэтому уникальные решения конституционных судов существующие методы анализа предсказывать не способны [Sert, Yildirim, Haşlak, 2022]. Во-вторых, отдельный технический вызов связан с работой на языках, отличающихся от английского, например, на китайском, турецком [ibid.] или португальском [Jacob de Menezes-Neto, Clementino, 2022]. Впрочем, это ограничение преодолимо — предварительно обученные языковые модели можно доучить. В-третьих, алгоритмы,

построенные на малых или специфических выборках, имеют меньше практической пользы, чем модели, построенные для судов общей юрисдикции [ibid.]. Алгоритмы научились отлично предсказывать (в «слабом» значении этого термина) решения Верховного суда США и ЕСПЧ, но это не позволяет распространить результаты на суды нижнего уровня. Вопрос о разработке моделей, предсказывающих решения судов по типичным делам, которые касаются основной массы населения, пока не был решен: имеющиеся исследования в этой области крайне немногочисленны, а универсальных решений не существует. В-четвертых, большинство моделей имеют дело с предсказаниями в «слабом» значении этого термина и не дают гарантий, что можно точно предсказать будущие решения судей. Можно описать это как проблему индукции — сколько бы много наблюдений мы ни сделали о прошлом, они не позволяют заглянуть в будущее. Таким образом, высокая точность предсказаний ИИ на рисунке 1 не должна вводить нас в заблуждение: она лишь отражает хорошую тренировку моделей на очень ограниченном количестве данных о прошлом.

Наконец, ключевая проблема моделей, основанных на глубоком обучении, состоит в том, что их сложно интерпретировать. Судья-человек может объяснить логику принятия им решений. Классические модели, такие как регрессии и деревья решений, также дают легко интерпретируемые результаты. Например, из регрессионной модели о влиянии различных факторов на решения российских судей по делам об убийствах [Zhuchkova, Kazun, 2023] несложно понять, какое влияние на приговор имеет каждый анализируемый параметр дела — пол, признание вины, повторное совершение преступления и пр. Однако методы предсказания судебных решений, основанные на глубоком обучении, представляют собой «черные ящики». Хотя модели дают весьма точные предсказания, они не позволяют исследователям понять, какие именно легальные и экстралегальные факторы стоят за каждым конкретным решением. Однако преимущество ИИ над линейными регрессионными моделями состоит в том, что последние могут недоучитывать нелинейность влияния многих факторов [Alarie et al., 2017]. Также именно модели, основанные на глубоком обучении, дают наиболее точные прогнозы.

Одна из самых высокоцитируемых работ о принятии решений на основе машинного обучения [Rudin, 2019] содержит ключевой тезис в своем названии — не следует использовать «черные ящики» для принятия решений в тех случаях, когда цена ошибки велика. В работе автор называет мифом классическую дилемму между объяснительной силой модели и ее интерпретируемостью. Интерпретируемые модели тоже могут иметь высокую предсказательную способность. Например, в исследовании [Tan et al., 2020] на 1,3 млн кейсах о преступлениях в Китае была построена поддающаяся интерпретации модель с точностью, сопоставимой или превышающей результаты моделей, действующих по принципу «черных ящиков». Ключевое преимущество моделей, позволяющих интерпретировать результаты, состоит в возможности работы над ошибкой. Если ИИ-алгоритм вынес неверный приговор (или предложил неверный курс лечения), это может оказать существенное влияние на судьбу человека. Если подобную ошибку совершает «черный ящик», мы не сможем сказать, почему она была совершена и повторится ли она в будущем. Из-за этого недостатка подобные системы мо-

гут никогда не получить высокого уровня доверия у общества. Вместе с тем это не означает, что развитие прогнозирования на основе «черного ящика» в сфере медицины или права должно остановиться: если такие алгоритмы будут обладать статистически более высокой точностью, то для некоторых задач точность и скорость могут перевесить прозрачность.

В 2016 г. Верховный суд штата Висконсин США вынес решение о том, что при принятии решений судьи могут опираться на результаты анализа, выполненного алгоритмами, даже если принцип работы алгоритма не вполне прозрачен [Beriaín, 2018]. Аргументация такого решения сводилась к двум тезисам: качество работы алгоритмов весьма высоко, а судья сам по себе достаточно компетентен, чтобы принимать взвешенное решение о том, нужно ли ориентироваться на результаты алгоритма. Фактически этот прецедент открыл дорогу использованию ИИ в качестве ассистента судьи при принятии решений в США. Поводом для этого разбирательства стала система COMPAS (Correctional Offender Management Profiling for Alternative Sanctions). По замыслу создателей COMPAS должен прогнозировать индивидуальный риск рецидивизма, что является важным обстоятельством при принятии судьями решений. Однако решение суда породило дебаты о точности работы алгоритма [Dressel, Farid, 2018], а также о том, не воспроизводит ли он расовые стереотипы, почерпнутые из обучения на решениях, принятых людьми. Как бы там ни было, без этой системы судьи все равно должны оценивать риски того, что обвиняемый совершит преступление повторно, и этот процесс не свободен от различных предубеждений [Brennan, Dieterich, 2018].

В целом идея о возможности предсказать значительную долю судебных приговоров должна поменять представление о том, что такое дискреция судьи [Tahura, Selvadurai, 2023]. Не все юридические решения имеют единственное решение. Если алгоритм ИИ будет способен предсказывать решения судей очень точно, это означает, что он должен воспроизводить не только формальный анализ легальных и экстралегальных факторов, но и то, что сами судьи назвали «психологией судебного разбирательства»<sup>11</sup>. Без развития общего ИИ, о котором не шла речь в данной статье, алгоритм вряд ли будет способен понимать нюансы языка, игру слов, ценности и смыслы, воспринимаемые людьми в различных обстоятельствах дела (но даже «слабый» ИИ со временем вполне может научиться имитировать это понимание, если обучить его на достаточном объеме подобных кейсов). Отсюда же возникает важное опасение, связанное с тем, не воспроизводит ли ИИ в своем анализе стереотипы, идеологию и другие убеждения судей [Manresa-Yee, Ramis, 2022].

В этом контексте возникает еще более интересный вопрос: должен ли ИИ полностью прогнозировать решения судьи-человека или же он может встать в этом отношении на ступеньку выше? Исследования показывают [Doerner, 2015; Franklin, Fearn, 2008], что судьи в США не свободны от расовых и гендерных стереотипов. Более того, в условиях ограниченной рациональности [Albonetti, 1991], связанной с нехваткой информации и времени, именно эти стереотипы нередко становятся опорными точками [Steffensmeier, Ulmer, Kramer, 1998], помогающими

<sup>11</sup> В совете судей отвергли идею замены судьи на ИИ // Право.ru. 2023. 26 октября. URL: <https://pravo.ru/news/249529/> (дата обращения: 06.10.2024).

принимать непростые решения. Ведь ИИ мог бы следовать букве закона строже, чем человек (не говоря о более высокой скорости анализа информации). Человек не свободен от симпатий, ценностей и убеждений, психологического давления, в то время как ИИ следует этим факторам только в том случае, если его научили им следовать. Это непростой вопрос, на который пока нет ответа. Однако можно прогнозировать, что по меньшей мере население к такому повороту событий готово будет нескоро.

В исследовании [Barysè, Sarel, 2023] выдвигается важный тезис о том, что легитимность использования ИИ воспринимается населением и профессионалами по-разному в зависимости от стадии судебного процесса, на которой применяется технология. Люди в целом доверяют судьям-людям больше, чем ИИ, по этой причине они не поддерживают высокую автоматизацию работы судов на стадии принятия решений. Однако население выступает за использование технологий на стадии сбора данных, поскольку это может способствовать повышению объективности анализа, выполненного человеком.

### **Заключение: ИИ как ассистент судьи**

Результаты исследования позволяют выдвинуть несколько практических рекомендаций о развитии ИИ-технологий для помощи юристам и судьям. Хотя ИИ пока не может заменить судью, он способен помочь судьям в их профессиональной деятельности.

Мы показали, что, несмотря на стремительное совершенствование алгоритмов, предсказывающих решения судей, большинство из них страдают от ограничений, связанных с данными. Например, алгоритмы, предназначенные для анализа решений высших судебных органов, имеют низкую внешнюю валидность. Большинство судебных процессов, касающихся граждан и бизнеса, происходят в судах нижнего уровня, прогнозировать которые ИИ пока не умеет. За исключением единственного набора данных из Китая [Xiao et al., 2018], другие публично доступные данные включают крайне ограниченное число кейсов, не превышающих несколько десятков тысяч дел. Впрочем, в России уже есть коммерческие проекты, умеющие предсказывать исходы арбитражных дел, однако их алгоритмы непрозрачны, поскольку не публиковались в рецензируемых академических изданиях.

В России в соответствии с законом от 22 декабря 2008 г. № 262-ФЗ «Об обеспечении доступа к информации о деятельности судов в Российской Федерации» публикуется огромное количество данных об уголовных, гражданских и арбитражных делах, что создает большой потенциал для обучения на них ИИ. Это означает, что, хотя на сегодняшний день не созданы ИИ модели, которые бы предсказывали решения российских судов, в этой сфере в обозримом будущем могут быть сделаны существенные прорывы. Исследования об использовании ИИ для классификации юридических текстов во всем мире все еще находятся в начале своего развития.

Другой вопрос состоит в том, нужно ли вообще внедрять ИИ-технологии в судебный процесс. Ответить на него однозначно, конечно, нельзя, и дебаты по этому поводу свидетельствуют о сложности этических вопросов. Однако сложность этических вопросов не отменяет важности развития самой технологии. Одна из пер-

спективных возможностей — это превращение ИИ в ассистента, который может помогать судьям в работе с материалами дела, включая их классификацию, поиск проблемных мест, выделение наиболее значимых обстоятельств дела. Публичная позиция одного из судей Верховного суда РФ<sup>12</sup> в целом поддерживает именно этот сценарий.

Важно учитывать, что большинство существующих алгоритмов предсказания судебных решений действуют по принципу «черного ящика», и это сильно ограничивает потенциал для их внедрения в юриспруденцию. В сферы с высокими ставками нельзя бездумно интегрировать алгоритмы, ход работы которых мы не понимаем. Но возможно, что дилемма между точностью решений и прозрачностью процесса их принятия является ложной [Rudin, 2019]. В таком случае в судебную практику должны внедряться именно те решения, которые позволяют не только предсказать решения, но и объяснить их. Именно последняя задача и должна стоять на повестке отечественных исследований в области использования ИИ-технологий в судебном процессе.

### Список литературы (References)

1. Захарцев С. И., Сальников В. П. Судья-робот в уголовном процессе: хорошо или плохо? // Юридическая наука; история и современность. 2018. № 7. С. 176—180.  
Zakharcev S. I., Salnikov V. P. (2018) The Robot Judge in Criminal Proceedings: Good or Bad? *Legal Science: History and Modernity*. No. 7. P. 176—180. (In Russ.)
2. Коваленко К. Е., Печатнова Ю. В., Стаценко Д. А., Коваленко Н. Е. Судья-робот как преодоление противоречий судебного усмотрения (юридический аспект) // Юридический вестник Дагестанского Государственного Университета. 2020. Т. 36. № 4. С. 169—173.  
Kovalenko K. E., Pechatnova Yu. V., Statsenko D. A., Kovalenko N. E. (2020) The Robot Judge as a Resolution of Judicial Discretion Contradictions (Legal Aspect). *Legal Bulletin of Dagestan State University*. Vol. 36. No. 4. P. 169—173. (In Russ.)
3. Кравчук Н. В. Искусственный интеллект как судья: перспективы и опасения (Обзор) // Социальные и гуманитарные науки. Отечественная и зарубежная литература. Серия 4: Государство и право. 2021. № 1. С. 115—122.  
Kravchuk N. V. (2021) Artificial Intelligence as a Judge: Prospects and Concerns (Review). *Social and Human Sciences. Domestic and Foreign Literature. Series 4: State and Law*. No. 1. P. 115—122. (In Russ.)
4. Степанов О. А., Басангов Д. А. О перспективах влияния искусственного интеллекта на судопроизводство // Вестник Томского Государственного Университета. Философия. Социология. Политология. 2022. № 475. С. 229—237.  
Stepanov O. A., Basangov D. A. (2022) On the Prospects of Artificial Intelligence Impact on Judiciary. *Bulletin of Tomsk State University. Philosophy, Sociology, Political Science*. No. 475. P. 229—237. (In Russ.)

<sup>12</sup> Бурнов В. ВС: ИИ может помочь снизить число ошибок при назначении наказания // РАПСИ. 2024. 1 июля. URL: [https://www.rapsinews.ru/judicial\\_mm/20240701/310059533.html](https://www.rapsinews.ru/judicial_mm/20240701/310059533.html) (дата обращения: 05.07.2024).

5. Alali M., Syed S., Alsayed M., Patel S., Bodala H. (2021) JUSTICE: A Benchmark Dataset for Supreme Court's Judgment Prediction. *arXiv*. Art. 2112.03414. <https://doi.org/10.48550/arXiv.2112.03414>.
6. Alarie B., Niblett A., Yoon A. (2017) Using Machine Learning to Predict Outcomes in Tax Law. SSRN. <https://doi.org/10.2139/ssrn.2855977>.
7. Albonetti C. A. (1991) An Integration of Theories to Explain Judicial Discretion. *Social Problems*. Vol. 38. No. 2. P. 247—266. <https://doi.org/10.2307/800532>.
8. Alcántara Francia O. A., Nunez-del-Prado M., Alatrística-Salas H. (2022) Survey of Text Mining Techniques Applied to Judicial Decisions Prediction. *Applied Sciences*. Vol. 12. No. 20. Art. 20. <https://doi.org/10.3390/app122010200>.
9. Aletras N., Tsarapatsanis D., Preotiuc-Pietro D., Lampos V. (2016) Predicting Judicial Decisions of the European Court of Human Rights: A Natural Language Processing Perspective. *PeerJ Computer Science*. Vol. 2. No. 10. Art. e93. <https://doi.org/10.7717/peerj-cs.93>.
10. Alghazzawi D., Bamasag O., Albeshri A., Sana I., Ullah H., Asghar M. Z. (2022) Efficient Prediction of Court Judgments Using an LSTM+CNN Neural Network Model with an Optimal Feature Set. *Mathematics*. Vol. 10. No. 5. Art. 5. <https://doi.org/10.3390/math10050683>.
11. Ashley K. D., Brüninghaus S. (2009) Automatically Classifying Case Texts and Predicting Outcomes. *Artificial Intelligence and Law*. Vol. 17. No. 2. P. 125—165. <https://doi.org/10.1007/s10506-009-9077-9>.
12. Bagherian-Marandi N., Ravanshadnia M., Akbarzadeh-T M.-R. (2021) Two-Layered Fuzzy Logic-Based Model for Predicting Court Decisions in Construction Contract Disputes. *Artificial Intelligence and Law*. Vol. 29. No. 4. P. 453—484. <https://doi.org/10.1007/s10506-021-09281-9>.
13. Barysé D., Sarel R. (2023) Algorithms in the Court: Does It Matter Which Part of the Judicial Decision-Making is Automated? *Artificial Intelligence and Law*. Vol. 32. P. 117—146. <https://doi.org/10.1007/s10506-022-09343-6>.
14. Beriain I. D. M. (2018) Does the Use of Risk Assessments in Sentences Respect the Right to Due Process? A Critical Analysis of the *Wisconsin v. Loomis Ruling*. *Law, Probability & Risk*. Vol. 17. No. 1. P. 45—53. <https://doi.org/10.1093/lpr/mgy001>.
15. Brennan, T., Dieterich W. (2018) Correctional Offender Management Profiles for Alternative Sanctions (COMPAS). In: Singh J. P., Kroner D. G., Wormith J. S., Desmarais S. L., Hamilton Z. (eds.) *Handbook of Recidivism Risk/Needs Assessment Tools*. Hoboken: John Wiley & Sons. P. 49—75. <https://doi.org/10.1002/9781119184256.ch3>.
16. Chalkidis I., Androutsopoulos I., Aletras N. (2019) Neural Legal Judgment Prediction in English. *arXiv*. Art. 1906.02059. <https://doi.org/10.48550/arXiv.1906.02059>.

17. Chalkidis I., Fergadiotis M., Malakasiotis P., Aletras N., Androutopoulos I., Androutopoulos I. (2020) LEGAL-BERT: The Muppets Straight Out of Law School. *arXiv*. Art. 2010.02559. <https://arxiv.org/abs/2010.02559>.
18. Chicco D., Jurman G. (2020) The Advantages of the Matthews Correlation Coefficient (MCC) over F1 Score and Accuracy in Binary Classification Evaluation. *BMC Genomics*. Vol. 21. No. 1. Art. 6. <https://doi.org/10.1186/s12864-019-6413-7>.
19. Devlin J., Chang M.-W., Lee K., Toutanova K. (2019) BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding. In: Burstein J., Doran C., Solorio T. (eds.) *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. Vol. 1. Minneapolis, Association for Computational Linguistics. P. 4171—4186. <https://doi.org/10.18653/v1/N19-1423>.
20. Doerner J. K. (2015) The Joint Effects of Gender and Race/Ethnicity on Sentencing Outcomes in Federal Courts. *Women & Criminal Justice*. Vol. 25. No. 5. P. 313—338. <https://doi.org/10.1080/08974454.2014.989298>.
21. Franklin C. A., Fearn N. E. (2008) Gender, Race, and Formal Court Decision-Making Outcomes: Chivalry/Paternalism, Conflict Theory or Gender Conflict? *Journal of Criminal Justice*. Vol. 36. No. 3. P. 279—290. <https://doi.org/10.1016/j.jcrimjus.2008.04.009>.
22. Freitas A. L., Allende-Cid H., Santana O., Oliveira-Lage L. (2022) Predicting Brazilian Court Decisions. *PeerJ Computer Science*. Vol. 8. Art. e904. <https://doi.org/10.7717/peerj-cs.904>.
23. Gerber M. S. (2014) Predicting Crime Using Twitter and Kernel Density Estimation. *Decision Support Systems*. Vol. 61. P. 115—125. <https://doi.org/10.1016/j.dss.2014.02.003>.
24. Jacob de Menezes-Neto E., Clementino M. B. M. (2022) Using Deep Learning to Predict Outcomes of Legal Appeals Better than Human Experts: A Study with Data from Brazilian Federal Courts. *PLoS ONE*. Vol. 17. No. 7. Art. e0272287. <https://doi.org/10.1371/journal.pone.0272287>.
25. Dressel J., Farid H. (2018) The Accuracy, Fairness, and Limits of Predicting Recidivism. *Science Advances*. Vol. 4. No. 1. Art. eaao5580. <https://doi.org/10.1126/sciadv.aao5580>.
26. Cui J., Shen X., Wen S. (2023) A Survey on Legal Judgment Prediction: Datasets, Metrics, Models and Challenges. *IEEE Access*. Vol. 11. P. 102050—102071. <https://doi.org/10.1109/access.2023.3317083>.
27. Katz D. M., Bommarito M. J., Blackman J. (2017) A General Approach for Predicting the Behavior of the Supreme Court of the United States. *PLoS ONE*. Vol. 12. No. 4. Art. e0174698. <https://doi.org/10.1371/journal.pone.0174698>.

28. Kaufman A. R., Kraft P., Sen M. (2019) Improving Supreme Court Forecasting Using Boosted Decision Trees. *Political Analysis*. Vol. 27. No. 3. P. 381—387. <https://doi.org/10.1017/pan.2018.59>.
29. Kaur H., Choudhury T., Singh T. P., Shamoan Mohammad. (2019) Crime Analysis using Text Mining. In: Ming Fong A. Ch., Hong G. Y., Fong B. (eds.) *2019 International Conference on Contemporary Computing and Informatics (IC3I)*. Singapore. P. 283—288. <https://doi.org/10.1109/IC3I46837.2019.9055606>.
30. Kort F. (1957) Predicting Supreme Court Decisions Mathematically: A Quantitative Analysis of the “Right to Counsel” Cases. *American Political Science Review*. Vol. 51. No. 1. P. 1—12. <https://doi.org/10.2307/1951767>.
31. Kowsrihawat K., Vateekul P., Boonkwan P. (2018) Predicting Judicial Decisions of Criminal Cases from Thai Supreme Court Using Bi-directional GRU with Attention Mechanism. In: Do Van T., Do Duc H., Nguyen G. (eds.) *5th Asian Conference on Defense Technology (ACDT)*. Hanoi, Vietnam. P. 50—55. <https://doi.org/10.1109/acdt.2018.8592948>.
32. Malik V., Sanjay R., Nigam S. K., Ghosh K., Guha S. K., Bhattacharya A., Bhattacharya A., Modi A. (2021) ILDC for CJPE: Indian Legal Documents Corpus for Court Judgment Prediction and Explanation. In: Zong Ch., Xia F., Li W., Navigli R. (eds.) *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Confer.* Stroudsburg: Association for Computational Linguistics. P. 4046—4062. <https://doi.org/10.18653/v1/2021.acl-long.313>.
33. Manresa-Yee C., Ramis S. (2022) Assessing Gender Bias in Predictive Algorithms using eXplainable AI *arXiv*. Art. 2203.10264. <https://doi.org/10.48550/arXiv.2203.10264>.
34. Medvedeva M., Üstun A., Xu X., Vols M., Wieling M. (2021) Automatic Judgement Forecasting for Pending Applications of the European Court of Human Rights. In: Ashley K. D., Atkinson K., Branting K., Francesconi E., Grabmair M., Walker V. R., Waltl B., Wyner A. Z. (eds.) *Proceedings of the fifth workshop on automated semantic analysis of information in legal text (ASAIL 2021), São Paulo, Brazil*. CEUR Workshop Proceedings. P. 1—12. <https://ceur-ws.org/Vol-2888/paper2.pdf>.
35. Medvedeva M., Vols M., Wieling M. (2020) Using Machine Learning to Predict Decisions of the European Court of Human Rights. *Artificial Intelligence and Law*. Vol. 28. No. 2. P. 237—266. <https://doi.org/10.1007/s10506-019-09255-y>.
36. Medvedeva M., Wieling M., Vols M. (2022) Rethinking the Field of Automatic Prediction of Court Decisions. *Artificial Intelligence and Law*. Vol. 31. P. 195—212. <https://doi.org/10.1007/s10506-021-09306-3>.
37. Medvedeva M., Xiao X., Wieling M., Vols M. (2020) JURI SAYS: An Automatic Judgement Prediction System for the European Court of Human Rights. *Frontiers in Artificial Intelligence and Applications*. Vol. 334. P. 277—280. <https://doi.org/10.3233/faia200883>.

38. O'Sullivan C., Beel J. (2019) Predicting the Outcome of Judicial Decisions made by the European Court of Human Rights. *arXiv*. Art. 1912.10819. <https://doi.org/10.48550/arXiv.1912.10819>.
39. Quemy A., Wrembel R. (2020) On Integrating and Classifying Legal Text Documents. In: Hartmann S., Küng J., Kotsis G., Tjoa A. M., Khalil I. (eds.) *Database and Expert Systems Applications*. Cham, Switzerland: Springer International Publishing. P. 385—399. [https://doi.org/10.1007/978-3-030-59003-1\\_25](https://doi.org/10.1007/978-3-030-59003-1_25).
40. Rudin C. (2019) Stop Explaining Black Box Machine Learning Models for High Stakes Decisions and Use Interpretable Models Instead. *Nature Machine Intelligence*. Vol. 1. No. 5. Art. 5. <https://doi.org/10.1038/s42256-019-0048-x>.
41. Ruger T. W., Kim P. T., Martin A., Martin A. D., Quinn K. M. (2004) The Supreme Court Forecasting Project: Legal and Political Science Approaches to Predicting Supreme Court Decisionmaking. *Columbia Law Review*. Vol. 104. No. 4. P. 1150—1209. <https://doi.org/10.2307/4099370>.
42. Rusakova E. P. (2021) Integration of “Smart” Technologies in the Civil Proceedings of the People’s Republic of China. *RUDN Journal of Law*. Vol. 25. No. 3. P. 622—633. <https://doi.org/10.22363/2313-2337-2021-25-3-622-633>.
43. Sebők M., Kiss R., Járay I. (2023) Introducing HUNCOURT: A New Open Legal Database Covering the Decisions of the Hungarian Constitutional Court for Between 1990 and 2021. *Journal of the Knowledge Economy*. Vol. 15. P. 6507—6540. <https://doi.org/10.1007/s13132-023-01395-6>.
44. Sert M. F., Yıldırım E., Haşlak İ. (2022) Using Artificial Intelligence to Predict Decisions of the Turkish Constitutional Court. *Social Science Computer Review*. Vol. 40. No. 6. P. 1416—1435. <https://doi.org/10.1177/08944393211010398>.
45. Shaikh R. A., Sahu T. P., Anand V. (2020) Predicting Outcomes of Legal Cases based on Legal Factors using Classifiers. *Procedia Computer Science*. Vol. 167. P. 2393—2402. <https://doi.org/10.1016/j.procs.2020.03.292>.
46. Sharma R. D., Mittal S., Tripathi S., Acharya S. (2015) Using Modern Neural Networks to Predict the Decisions of Supreme Court of the United States with State-of-the-Art Accuracy. In: Arik, S., Huang, T., Lai, W., Liu, Q. (eds.) *Neural Information Processing. ICONIP 2015. Lecture Notes in Computer Science*. Vol. 9490. Cham: Springer. P. 475—483. [https://doi.org/10.1007/978-3-319-26535-3\\_54](https://doi.org/10.1007/978-3-319-26535-3_54).
47. Shi C., Sourdin T., Li B. (2021) The Smart Court — A New Pathway to Justice in China? *International Journal of Court Administration*. Vol. 12. No. 1. Art. 1. <https://doi.org/10.36745/ijca.367>.
48. Song D., Gao S., He B., Schilder F. (2022) On the Effectiveness of Pre-Trained Language Models for Legal Natural Language Processing: An Empirical Study. *IEEE Access*. Vol. 10. P. 75835—75858. <https://doi.org/10.1109/ACCESS.2022.3190408>.

49. Sourdin T., Cornes R. (2018) Do Judges Need to Be Human? The Implications of Technology for Responsive Judging. In: Sourdin T., Zariski A. (eds.) *The Responsive Judge: International Perspectives*. Singapore: Springer. P. 87—119. [https://doi.org/10.1007/978-981-13-1023-2\\_4](https://doi.org/10.1007/978-981-13-1023-2_4).
50. Steffensmeier D., Ulmer J., Kramer J. (1998) The Interaction of Race, Gender, and Age in Criminal Sentencing: The Punishment Cost of Being Young, Black, and Male. *Criminology*. Vol. 36. No. 4. P. 763—798. <https://doi.org/10.1111/j.1745-9125.1998.tb01265.x>.
51. Tahura U. S., Selvadurai N. (2023) The Use of Artificial Intelligence in Judicial Decisionmaking: The Example of China. *International Journal of Law, Ethics, and Technology*. Vol. 3. P. 1—20. <https://doi.org/10.55574/pyeb5374>.
52. Tan H., Zhang B., Zhang H., Li R. (2020) The Sentencing-Element-Aware Model for Explainable Term-of-Penalty Prediction. In: Zhu X., Zhang M., Hong Y., He R. (eds.) *Natural Language Processing and Chinese Computing. NLPCC2020. Lecture Notes in Computer Science*. Vol. 12431. Cham: Springer. P. 16—27. [https://doi.org/10.1007/978-3-030-60457-8\\_2](https://doi.org/10.1007/978-3-030-60457-8_2).
53. Taylor I. (2023) Justice by Algorithm: The Limits of AI in Criminal Sentencing. *Criminal Justice Ethics*. Vol. 42. No. 3. P. 193—213. <https://doi.org/10.1080/0731129X.2023.2275967>.
54. Virtucio M. B. L., Aborot J. A., Abonita J. K. C., Avinante R. S., Copino R. J. B., Neverida M. P., Osiana V. O., Peramo E. C., Syjuco J. G., Tan G. B. A. (2018) Predicting Decisions of the Philippine Supreme Court Using Natural Language Processing and Machine Learning. In: Honiden S., Fujii R. (eds.) *2018 IEEE 42<sup>nd</sup> Annual Computer Software and Applications Conference (COMPSAC)*. Tokyo. P. 130—135. <https://doi.org/10.1109/compsac.2018.10348>.
55. Volkov V. (2016) Legal and Extralegal Origins of Sentencing Disparities: Evidence from Russia's Criminal Courts. *Journal of Empirical Legal Studies*. Vol. 13. No. 4. P. 637—665. <https://doi.org/10.1111/jels.12128>.
56. Waltl B., Bonczek G., Scepankova E., Landthaler J., Matthes F. (2017) Predicting the Outcome of Appeal Decisions in Germany's Tax Law. Electronic Participation. In: Parycek P. et al. (eds.) *Electronic Participation. ePart 2017. Lecture Notes in Computer Science*. Vol. 10429. Cham: Springer. P. 89—99. [https://doi.org/10.1007/978-3-319-64322-9\\_8](https://doi.org/10.1007/978-3-319-64322-9_8).
57. Xiao C., Zhong H., Guo Z., Tu C., Liu Z., Sun M., Feng Y., Han X., Hu Z., Wang H., Xu J. (2018) CAIL2018: A Large-Scale Legal Dataset for Judgment Prediction. *arXiv*. Art. 1807.02478. <https://arxiv.org/abs/1807.02478>.
58. Xu Z. (2022) Human Judges in the Era of Artificial Intelligence: Challenges and Opportunities. *Applied Artificial Intelligence*. Vol. 36. No. 1. Art. 2013652. <https://doi.org/10.1080/08839514.2021.2013652>.

59. Zhong H., Xiao C., Tu C., Tianyang Zhang, Zhiyuan L., Liu Z., Liu Z., Sun M. (2020) How Does NLP Benefit Legal System: A Summary of Legal Artificial Intelligence. In: Jurafsky D., Chai J., Schluter N., Tetreault J. (eds.) *Proceedings of the 58<sup>th</sup> Annual Meeting of the Association for Computational Linguistics*. Stroudsburg. Association for Computational Linguistics. P. 5218—5230. <https://aclanthology.org/2020.acl-main.466/>.
60. Zhuchkova S., Kazun A. (2023) Exploring Gender Bias in Homicide Sentencing: An Empirical Study of Russian Court Decisions Using Text Mining. *Homicide Studies*. <https://doi.org/10.1177/10887679231217159>.

DOI: [10.14515/monitoring.2024.5.2600](https://doi.org/10.14515/monitoring.2024.5.2600)



**И. С. Душакова, Н. С. Душакова**

## **ГЕНЕРАТИВНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В РОССИЙСКИХ ПРАВОСЛАВНЫХ СООБЩЕСТВАХ: ВОСПРИЯТИЕ И ПРАКТИКИ ИСПОЛЬЗОВАНИЯ**

### **Правильная ссылка на статью:**

Душакова И. С., Душакова Н. С. Генеративный искусственный интеллект в российских православных сообществах: восприятие и практики использования // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 123—141. <https://doi.org/10.14515/monitoring.2024.5.2600>.

### **For citation:**

Dushakova I. S., Dushakova N. S. (2024) Generative Artificial Intelligence in Russian Orthodox Communities: Perception and Practices of Using. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 123—141. <https://doi.org/10.14515/monitoring.2024.5.2600>. (In Russ.)

Получено: 28.03.2024. Принято к публикации: 15.08.2024.

## ГЕНЕРАТИВНЫЙ ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ В РОССИЙСКИХ ПРАВОСЛАВНЫХ СООБЩЕСТВАХ: ВОСПРИЯТИЕ И ПРАКТИКИ ИСПОЛЬЗОВАНИЯ

*ДУШАКОВА Ирина Сергеевна — кандидат филологических наук, доцент Института медиа, факультет креативных индустрий, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия*

*E-MAIL: [dusacova@gmail.com](mailto:dusacova@gmail.com)*

*<https://orcid.org/0000-0002-7959-4934>*

*ДУШАКОВА Наталья Сергеевна — кандидат исторических наук, старший научный сотрудник Лаборатории теоретической фольклористики ШАГИ, Российская академия народного хозяйства и государственной службы при Президенте РФ, Москва, Россия*

*E-MAIL: [dushakova@list.ru](mailto:dushakova@list.ru)*

*<https://orcid.org/0000-0003-4486-5367>*

**Аннотация.** В статье через призму социального конструирования технологий (SCOT) рассматриваются восприятие и практики использования генеративного искусственного интеллекта в религиозных сообществах России. Цель исследования — выявление интерпретативной гибкости генеративного ИИ как технологического артефакта в православных сообществах России (последователей Русской православной церкви, старообрядцев разных согласий). Авторы исходят из гипотезы о том, что первое время после массового релиза генеративного ИИ (ChatGPT, Midjourney и др. инструментов) набор контекстов и областей применения, в которых генеративный ИИ видим и понятен, довольно узок. Кроме того, учитывая новизну технологий и низкую вероятность наличия предыдущего опыта в обращении с ними, начало взаимодействия характеризуется высокой гибкостью интерпретаций генеративного ИИ. Для выявления набора практик и кон-

## GENERATIVE ARTIFICIAL INTELLIGENCE IN RUSSIAN ORTHODOX COMMUNITIES: PERCEPTION AND PRACTICES OF USING

*Irina S. DUSHAKOVA<sup>1</sup> — Cand. Sci. (Philology), Assistant Professor at the Institute of Media, Faculty of Creative Industries*

*E-MAIL: [dusacova@gmail.com](mailto:dusacova@gmail.com)*

*<https://orcid.org/0000-0002-7959-4934>*

*Natalia S. DUSHAKOVA<sup>2</sup> — Cand. Sci. (Hist.), Senior Researcher at the Laboratory for Theoretical Folklore Studies, School for Advanced Studies in the Humanities*

*E-MAIL: [dushakova@list.ru](mailto:dushakova@list.ru)*

*<https://orcid.org/0000-0003-4486-5367>*

<sup>1</sup> HSE University, Moscow, Russia

<sup>2</sup> RANEPА, Moscow, Russia

**Abstract.** The article examines the perception and practices of using generative artificial intelligence in religious communities in Russia within the theoretical framework of the social construction of technology (SCOT). The study's main objective is to identify the interpretive flexibility of generative AI as a technological artifact in different Orthodox communities in Russia (followers of the Russian Orthodox Church and Old Believers of different denominations). The authors assume that the set of contexts and application areas in which generative AI is visible and understandable is relatively narrow soon after the mass release of generative AI (for example, ChatGPT, Midjourney, and other tools). Moreover, given the newness of the technologies and the low likelihood of previous experience with their usage, the authors assume that the early years of interaction with generative AI can be characterized by high flexibility in the interpretations of the artifact. To identify a set of practices and con-

текстов использования генеративного ИИ в православных сообществах проводится анализ их каналов и площадок в социальных медиа и СМИ, что позволяет говорить о конструировании технологии самими сообществами в связке с общим информационным полем. Хронологические рамки исследования охватывают период с 1 января по 31 декабря 2023 г. Во избежание смещения выборки при ручном отборе материала для работы привлекается база данных «Медиадиалогия» (мониторинг соцсетей).

Как показывает анализ данных, большая часть дискуссий выстраивается вокруг попыток установить границы применения новой технологии и выявить связанные с ней опасности. Сомнения в целесообразности применения искусственного интеллекта не затрагивают его утилитарное использование: граница допустимого очерчивается в сфере религиозных смыслов. Набор практик использования генеративного искусственного интеллекта на текущий момент достаточно узок, при этом самая распространенная практика — генерация или обработка изображений.

**Ключевые слова:** генеративный искусственный интеллект, нейросеть, социальное конструирование технологий, РПЦ, старообрядцы, религиозное осмысление технологий

**Благодарность.** Н. Душакова подготовила статью в рамках выполнения научно-исследовательской работы государственного задания РАНХиГС.

## Введение

Интерес к развитию искусственного интеллекта во многом связан с состоявшимся в 2022 г. массовым релизом зарубежных инструментов, позволяющих генерировать изображения (нейросеть Midjourney в бета-версии стала доступна 12 июля 2022 г.) и тексты (дата первоначального выпуска ChatGPT — 30 ноя-

texts that make generative AI visible for Orthodox communities, an analysis of the channels and platforms of these communities in social and mass media is carried out. It allows us to discuss technology construction by the communities and the general information field. The chronological scope of the study covers the period from January 1 to December 31, 2023. We used the Medialogy database (social media search tool) to avoid sample bias when manually selecting material for work.

Our data analysis reveals that the majority of discussions within Orthodox communities in Russia are focused on establishing the boundaries of the application of generative artificial intelligence and identifying associated dangers. Despite these concerns, the utilitarian use of AI remains unaffected, with the boundary of the permissible use outlined in the sphere of religious meanings. The set of practices for using generative artificial intelligence is currently quite narrow, with the most common practice being the generation or processing of images.

**Keywords:** generative AI, neural network, social construction of technologies, Russian Orthodox Church, Old Believers, religious understanding of technology

**Acknowledgments.** N. Dushakova prepared the article as a part of the RANEPА state assignment research programme.

бря 2022 г.). Российские инструменты, работающие с русскоязычным контентом и по русскоязычным промптам (YandexGPT, GigaChat), стали доступны в апреле-мае 2023 г. Появление этих инструментов в открытом доступе продемонстрировало новые возможности современных технологий и привлекло к ним общественное внимание.

Интерес специалистов из различных профессиональных областей к вопросам развития, внедрения и использования ИИ возник гораздо раньше. Еще в 2010 г. Хамид Р. Экбиа описал в обзорной статье 50-летнюю историю изучения искусственного интеллекта, начиная отсчет с Дартмутской конференции 1956 г. [Ekbia, 2010]. Систематизируя существующие дискуссии, автор отмечает, что к началу 2010-х годов сформировался целый ряд подходов к дефиниции искусственного интеллекта, причем «Центральное место в каждом из этих подходов занимает базовое понимание интеллекта как включающего в себя определенное свойство или способность — например, манипулирование символами, скорость вычислений, знания» [ibid.: 202]. Это означает, что единого ответа на вопрос о том, как мы трактуем искусственный интеллект, у нас не будет, поскольку нет консенсуса в его определении. Такое положение сохранилось и сегодня. Так, Э. Эллиотт в предисловии к сборнику исследований ИИ через призму социальных наук пишет: «Несмотря на общественный интерес к чат-ботам и беспилотным автомобилям, очень немногие люди понимают, как на самом деле функционирует ИИ и меняет мир на их глазах. Или, может быть, проблема именно в этом: ИИ, как и электричество, невидим. Это универсальная технология, которая творит чудеса за кулисами. Контуры и последствия ИИ остаются для нас неуловимыми — мы не можем увидеть их в действии, но каким-то образом ощущаем их воздействие. Как и другие технологии общего назначения, такие как двигатель внутреннего сгорания, телефония и кремниевый чип, ИИ становится повсеместным» [Elliott, 2022: 4].

Цель нашего исследования — анализ реакции российских религиозных сообществ на внедрение и распространение генеративного ИИ. Нас интересует, как православные верующие воспринимают новую технологию, как интерпретируют ее, в каких сферах видят допустимым ее использование и как рефлексиируют опыт личного взаимодействия с генеративным ИИ. В центре внимания — российские православные сообщества, которые включают последователей Русской православной церкви (далее — РПЦ) и старообрядцев. Последние не представляют собой однородную общность, а разделены на разные согласия (течения), наиболее многочисленное из которых — последователи Русской православной старообрядческой церкви (далее — РПСЦ). Используя базу данных «Медиаалогия» в качестве инструмента для создания выборки сообщений из социальных медиа и СМИ на тему взаимодействия религиозных сообществ с генеративным ИИ, мы проследим актуальные интерпретации и обсуждаемые в публичном пространстве практики пользования ИИ, характерные для российских православных сообществ на первом этапе внедрения и распространения новой технологии.

## Религиозное измерение ИИ: проблемные поля

Научные исследования, в которых рассматриваются взаимосвязи между религией и искусственным интеллектом, тематически разделяют на три большие группы:

- 1) ИИ в религии (как верующие используют ИИ);
- 2) религия в ИИ (какую роль религия играет в развитии ИИ);
- 3) религия ИИ (ИИ сам выступает как форма религии, квази- или парарелигиозный феномен)<sup>1</sup>.

Нас прежде всего интересует первое из обозначенных направлений — то, как в религиозных сообществах адаптируют технологию ИИ, как она осмысливается и используется. Эта проблематика рассматривается преимущественно в рамках исследований цифровой религии, антропологии религии, социальных коммуникаций, а также на стыке исследований медиа и религиоведения.

Представляя обзор работ в области исследований цифровой религии, Х. Кэмпбелл и Дж. Еволви [Campbell, Evolvi, 2019] среди наиболее актуальных направлений выделяют проблематику искусственного интеллекта и вызовов, которые он несет для религиозных сообществ. Перед исследователями встают вопросы о том, как может ИИ изменить представления о человеческой уникальности, какое влияние технологии могут оказывать на человеческую природу и какие последствия этих процессов могут ощутить представители различных религиозных сообществ. Другой аспект проблемы, на который обращают внимание авторы обзора, — рассмотрение технологии как имплицитной религии (когда практика или представление приобретает религиозные свойства). Здесь речь может идти о том, что технологии обладают как технологическими, так и религиозными аффордансами. В отличие от технологических аффордансов, которые зачастую понимаются как свойства девайсов, позволяющие использовать их определенным образом [Falahatpisheh, Khajeheian, 2020], религиозные аффордансы могут создавать религиозные смыслы и идентичности через вовлечение во взаимодействие с технологией. Изучая взаимосвязи между ними, важно рассматривать процесс принятия решений, во время которого или пользователь позволяет технологии создавать религиозные особенности, или религиозные представления формируют технологические особенности [Campbell, Evolvi, 2019].

В одном из номеров журнала «Religion and Social Communication», издаваемого Азиатским центром исследования религии и социальной коммуникации, вышла подборка статей, посвященных осмыслению «новой нормальности» в условиях цифровых инноваций и в первую очередь распространения генеративного ИИ. Номер журнала был собран по итогам круглого стола «Религиозная коммуникация и технологическое будущее: перспективы, опасения и ответы», а его участники продолжают свои исследования в области взаимодействия религии с ИИ. В частности, авторы отмечают потенциальную позитивную роль ИИ в межрелигиозном диалоге. Отправной точкой здесь может стать обсуждение этических аспектов использования технологии, ведь этим обеспокоены представители разных религиозных направлений [Jacoba, 2023].

Среди самых свежих публикаций на тему взаимодействия ИИ с религиозными акторами отметим также сборник эссе под редакцией Х. Кэмпбелл и П. Х. Ченг [Campbell, Cheong, 2023]. Цель сборника состояла в том, чтобы собрать более

<sup>1</sup> Religion and Artificial Intelligence: A Conversation with Marco Ventura // Georgetown University's Berkley Center for Religion, Peace, and World Affairs. 2022. May 18. URL: <https://www.youtube.com/watch?v=2BfstFlyQhc> (дата обращения 10.10.2024).

глубокие размышления по сравнению с алармистскими высказываниями в СМИ и распространенной критикой ИИ, когда даже сам термин «искусственный интеллект» представляется неоднозначным, поскольку охватывает целый набор технологий, таких как машинное обучение, обработка естественного языка, глубокие нейронные сети и робототехника. Редакторы пригласили кратко высказаться на тему группу исследовательниц из различных академических дисциплин, пытаясь при этом представить голоса из разных стран и религиозных традиций, чтобы выйти за рамки существующих дискуссий и обозначить новые аспекты проблемы [ibid.].

Среди наиболее интересных проблемных работ назовем эссе Б. Синглер, которая рассуждает об эсхатологических ожиданиях, связанных с ИИ. В таких ожиданиях как исследователи религии, так и сами верующие могут найти исторические параллели. Соответственно, экзистенциальный страх от ИИ — это лишь более современная версия апокалиптических сюжетов [Singler, 2023].

В ряде публикаций авторы делятся эмпирическими результатами из разных религиозных контекстов. В частности, важны наблюдения о религиозном измерении ИИ на Ближнем Востоке. Исламские учения поощряют стремление к знаниям и использование технологий для улучшения жизни при соблюдении этических и моральных принципов. В Саудовской Аравии для помощи паломникам в Мекку и Медину используются роботы, специально созданные для того, чтобы давать указания и предоставлять необходимые инструкции по проведению ритуалов на 11 языках. Вместе с тем в соцсетях многие выражают недовольство тем, что эти роботы созданы по подобию человека, а религия запрещает какую-либо персонификацию, в том числе ИИ [Rahman, Mohammed, 2023]. Среди эмпирических исследований христианского опыта отметим работу А. А. Абрамова, который описал попытки Ватикана, привлекая философов, разработчиков ИИ и религиозных лидеров, решить этические проблемы, связанные с разработкой и внедрением ИИ [Абрамов, 2020].

### **Теоретические рамки исследования: социальное конструирование технологий**

Исследование одновременно и технологии, и социального контекста, в котором технология появляется и развивается, возможно на основе нескольких теоретических подходов. Мы рассматриваем генеративный искусственный интеллект через призму социального конструирования технологий (*Social construction of technology, SCOT*) — конструктивистского подхода, получившего развитие в 1980-е годы. Как отмечает В. Бийкер, один из основателей этого подхода, в данной оптике «истинность научных фактов и работа технических артефактов изучаются как достижения — как сконструированные, а не как внутренние свойства этих фактов и машин» [Bijker, 2009: 88].

Социальное конструирование технологий — это теоретическая рамка, где новые технологии рассматриваются как артефакты, создаваемые в практиках использования этих технологий определенными группами людей. Помимо изучения использования отдельного технологического артефакта определенной группой людей — или, точнее, зазора между дизайном и пользовательскими практика-

ми [Pinch, Bijker, 1984], — исследования, проводимые в рамках SCOT, могут быть посвящены социальным основаниям, обеспечивающим общие рамки интерпретации технологии — например, рутинные ежедневные действия [ibid.] или религиозные верования [Campbell, 2012]. Описывая историю развития SCOT, В. Бийкер отмечает перспективу исследований с фокусом на религиозные сообщества: «Как только социальный конструктивистский подход оказывается принят, религиозные ценности также вступают в игру как часть технологических рамок соответствующих социальных групп» [Bijker, 2009: 93].

На первых шагах исследования устанавливаются сообщества, которые на одном и том же языке описывают интересующий исследователя артефакт, а также фиксируется «интерпретативная гибкость» артефакта — потенциальные разночтения в его возможностях и способах использования. В данной статье реализуется именно эта стадия исследования: предполагается, что последователи РПЦ и старообрядцы внутри своих сообществ используют характерные для них, но вариативные для разных групп описания генеративного ИИ. В рамках исследования мы продемонстрируем, каким представляется этот технологический артефакт и какие его аффордансы актуальны для практик интересующих групп. На следующих этапах исследований в рамках SCOT изучаются стабилизация значений и снижение интерпретативной гибкости артефакта — эта стадия будет доступна после установления общепринятых представлений о технологиях, что является перспективой для продолжения данного исследования.

В самом теоретическом языке идея артефакта не проблематизируется — артефакт предстает перед ученым таким, каким его видит изучаемое сообщество [Bijker, 2009]. Это отсутствие проблематизации можно рассматривать как методологический недостаток выбранной теоретической рамки и ее «слепое пятно» при работе с любым объектом, рассматриваемым как артефакт. Исследования ИИ, опирающиеся на другие теоретические рамки, либо не схватывают многообразие и разноплановость (генеративного) искусственного интеллекта и оставляют без исследовательского внимания эту сложность артефакта, не проблематизируя ее, либо опираются на редукционистский подход и какой-то отдельный, релевантный для поставленного исследовательского вопроса, но сконструированный исследователем набор признаков ИИ.

Таким образом, в случае изучения искусственного интеллекта именно упомянутая Э. Эллиоттом «размазанность» ИИ между девайсами, платформами и пр., схватывается методологической неопределенностью артефакта как ключевого понятия в рамках социального конструирования технологий.

## Обзор корпуса данных

Эмпирическая база исследования представляет собой корпус данных из социальных медиа. Данные были собраны за период с 1 января по 31 декабря 2023 г.: в течение этого года российские пользователи не только ознакомились с зарубежными инструментами генеративного ИИ, но и получили доступ к отечественным инструментам, разработанным «Яндексом» и «Сбербанком».

Мы остановились на анализе сообщений в социальных медиа, для того чтобы иметь возможность проследить, какие практики использования генеративного

ИИ обсуждаются пользователями технологии и какие события вызывают у них отклик, формируя социальный и культурный контекст. Многие посты в социальных сетях содержат ссылки и на материалы СМИ. Это позволяет нам заодно получить срез по информационным поводам, которые находят отклик у пользователей.

Данные были выгружены из базы «Медialogия», одной из крупнейших русскоязычных баз медийных данных. Как отмечается на сайте компании, на сегодняшний день «Медialogия» автоматически мониторит около 250 тыс. платформ, включающих более 2,5 млрд аккаунтов соцмедиа, что дает около 100 млн сообщений в сутки<sup>2</sup>.

При формировании корпуса материалов мы опирались на созданный нами контекстный запрос<sup>3</sup>, включавший одновременное упоминание ключевых слов из двух групп. Первая относилась к частотным запросам, связанным с генеративным ИИ (искусственный интеллект / нейросеть / ChatGPT / Midjourney / Алиса и др.), где для уточнения запроса вместе со словом «Алиса» должно было встречаться слово «колонка», или «Яндекс», или «помощник». Слова во второй группе относились к религиозным направлениям: отбирались упоминания РПЦ, РПСЦ, а также православных, старообрядцев, мусульман или иудеев (последние две группы были включены для обеспечения возможности сравнения и полноты контекста). Также в условия запроса мы добавили требование по совместной встречаемости ключевых слов из двух групп на расстоянии не более 25 слов друг от друга, что позволило нам получить значительно более чистые и релевантные материалы, поскольку такое расстояние в большинстве случаев означало наличие смысловой связи между ключевыми словами из двух указанных групп. Дополнительно мы воспользовались встроенными в базу функциями исключения спама и рекламы.

В результате описанного запроса мы получили 3043 медиасообщения. Абсолютной чистоты корпуса материалов при всех перечисленных действиях добиться не удалось, среди выгруженных сообщений встречаются и нерелевантные. Например, из-за части запроса, которая включала слово «Алиса», в корпус попало некоторое количество сообщений, посвященных произведениям К. Булычева с главной героиней Алисой Селезневой. Однако доля таких сообщений незначительна, а дальнейшее сужение поискового запроса могло привести к тому, что какие-то значимые кейсы оказались бы за пределами нашего рассмотрения. Опасения пропустить важные сообщения основаны на анализе ряда комментариев под постами, демонстрирующими сгенерированный нейросетями контент. Так, под постом с изображением «идеальной на взгляд ИИ» гиперсексуализированной фигуры женщины можно встретить комментарий «православная нейросеть!». Исключить такие случаи корректировкой запроса не представляется возможным, но и для полноценной интерпретации информации недостаточно. Так, невозможно проинтерпретировать оценку виртуального голосового ассистента Алисы на пять баллов из пяти с пометкой «мусульман». Более того, среди комментариев такого рода встречаются и более релевантные для нашего исследования: реагируя на изображения района одного из российских городов, пользователь

<sup>2</sup> О компании // Медialogия. 2024. URL: <https://www.mlg.ru/about/> (дата обращения 20.03.2024).

<sup>3</sup> В отличие от поиска по объекту, контекстный запрос не привязан к тому, какие объекты выделяет сама компания в своей базе данных. При помощи такого формирования запроса исследователи в меньшей степени зависят от предлагаемого базой интерфейса.



чает, что пространство реакции на ИИ пока довольно неустойчиво и сильно зависит не только от общего информационного контекста, но и от значимых информационных поводов, выводящих в топ площадки, где этот повод будет релевантным. Вместе с тем такая неустойчивость каналов репрезентации религиозной тематики в соцсетях фиксировалась нами и ранее. Так, анализ взаимодействия между социальными сетями и традиционными СМИ при освещении деятельности РПЦ показал, что «в социальных медиа мы фиксируем лишь отношение к обозначенному циклически воспроизводимому набору календарных праздников, остальная реакция пользователей оказывается ситуативной и непостоянной по набору основных источников за тот или иной период времени. Существует множество площадок в социальных сетях, реагирующих только на отдельные информационные поводы, связанные с РПЦ» [Душакова, Владимиров, 2022: 436]. В качестве яркого примера, который оказался в топе площадок за год более одного раза, можно привести телеграм-канал «AI ART — красивые и необычные картинки от Midjourney», где публикуются изображения, сгенерированные ИИ по текстовым промптам. Большое количество публикуемых в этом канале изображений связано с религиозной тематикой. Например, 27 октября 2023 г. по запросу «На этой динамичной фотографической картине в формате 4K тщательно очерченное лицо футуристического мужчины погружено в сердечную молитву, его преданность направлена на новаторскую роботизированную скульптуру» было сгенерировано следующее изображение (см. рис. 2).

Рис. 2. Футуристический мужчина погружен в сердечную молитву<sup>4</sup>



Далеко не всегда контент, профиль или канал/группу на площадке можно атрибутировать к РПЦ, РПСЦ или иным религиозным сообществам. Это не специфика религиозной медиатизированной коммуникации, а общее ограничение анализа контента в социальных сетях. Например, в случае отражения результата общения с нейросетью на тему Пасхи мы сталкиваемся с эксплицитным пере-

<sup>4</sup> Источник: AI ART — красивые и необычные картинки от Midjourney. 2023. 27 октября. URL: [https://t.me/ai\\_wow\\_art/520723](https://t.me/ai_wow_art/520723) (дата обращения 20.03.2024).

сечением интереса пользователя к генеративному ИИ и к религиозной тематике, но точно определить связь отдельного пользователя с той или иной религиозной общиной не представляется возможным, поскольку анализ контента социальных сетей дает нам представления о профилях, а не о пользователях. Кроме того, сама граница между религиозными и нерелигиозными медиа может быть размыта тем, что С. Хьярвард обозначает как поле банальной религии [Хьярвард, 2020].

В отдельных же случаях по тону комментария или контексту можно установить, что недостаточная репрезентация той или иной религиозной группы воспринимается негативно. Так, под видео «Самые влиятельные люди по версии нейросети»<sup>5</sup>, опубликованном на YouTube-канале «Короче, новости», комментаторы отмечали значимость фигуры Пророка Мухаммеда, не вошедшего в список. Со стороны части комментирующих такие высказывания были восприняты как обида мусульман на то, что нейросеть христианская, а не мусульманская.

В сформированном корпусе материалов можно условно выделить две преобладающие смысловые группы: 1) описание впечатлений от взаимодействия с генеративным ИИ или оценка сгенерированного контента, чаще — изображений, 2) реакция на информационные поводы, освещенные в СМИ.

Среди событий, вызвавших самую сильную реакцию пользователей соцсетей (вторая группа), наибольшее количество кейсов связано с РПЦ. На это повлиял наш запрос при формировании корпуса, однако можно встретить и события, связанные с иными религиозными направлениями. Например, 23 марта 2023 г. на многих площадках было опубликовано сообщение, что «нейросеть впервые поздравила всех Мусульман с наступлением Священного месяца Рамадан»<sup>6</sup>. Живую реакцию старообрядцев вызвала новость о том, что при помощи искусственного интеллекта по просьбе читателя телеграм-канала «#СМЫСЛЬ — Старообрядческая Мысль» были озвучены «Житие протопопы Аввакума» и книги об истории старообрядчества: «Роботы читают старообрядцам»<sup>7</sup>.

В качестве наиболее ярких примеров, связанных с РПЦ, можно назвать высказывания священников РПЦ о бессмысленности чтения проповедей от нейросети. Так, в ноябре 2023 г. руководитель информационной службы Московской епархии священник Александр Волков в интервью «Газете.Ru» сказал, что «служители церкви не могут обращаться к прихожанам с проповедью, написанной нейросетью, потому что она должна быть „личным словом пастыря“»<sup>8</sup>. Важно добавить, что в этом же интервью А. Волков отмечает и вполне допустимые, с его точки зрения, варианты использования (генеративных) нейросетей: «На самом деле я это довольно иронически рассказал, потому что всякий такого рода текст внешне соответствует всем необходимым форматам проповеди, но внутренне любой священник видит, что это подделка. Тут любой священник чувствует эту фальшь, которая там есть.

<sup>5</sup> Самые влиятельные люди по версии нейросети // Короче, новости. 2023. 25 марта. URL: <https://www.youtube.com/watch?v=MDhfjei5Qkc&lc=UgwaROT1METI1GmnWsh4AaABAg> (дата обращения 20.03.2024).

<sup>6</sup> FAVAID MEDIA | ИСЛАМСКИЕ СЕРИАЛЫ И ФИЛЬМЫ // ВКонтакте. 2023. 23 марта. URL: [https://vk.com/wall-190428221\\_13759](https://vk.com/wall-190428221_13759) (дата обращения 20.03.2024).

<sup>7</sup> РОБОТЫ ЧИТАЮТ СТАРООБРЯДЦАМ // #СМЫСЛЬ — Старообрядческая Мысль. 2023. 8 июля. URL: <https://t.me/staroveru/1948> (дата обращения 20.03.2024).

<sup>8</sup> Закарян Е. Священник объяснил, почему в РПЦ не пользуются нечеловеческими текстами // Газета.Ru. 2023. 17 ноября. URL: <https://www.gazeta.ru/social/news/2023/11/17/21731251.shtml?ysclid=ltw2jblbh2625236851> (дата обращения 20.03.2024).

В каких-то других ситуациях вполне допустимо использовать нейросети для поиска информации, для ее каталогизации, анализа и прочее. Когда мы говорим именно о живом слове, о творчестве человека, то тут, несомненно, совершенно неуместно пользоваться этими всеми вещами. Заменять нейросетью свое слово, которое является органической частью служения любого пастыря, совершенно невозможно, да и не нужно»<sup>9</sup>. Несмотря на этот развернутый комментарий А. Волкова, иллюстрирующий функциональный контекст упоминаний генеративного ИИ в нашем корпусе материалов, со стороны пользователей наибольшее внимание привлекло высказывание о неуместности использования сгенерированного текста в качестве проповеди, что также вынесено в заголовки тех СМИ, где это событие освещали.

Другим примером информационных поводов, вызывающих отклик у аудитории социальных сетей, является выражение официальной позиций церкви или размышления церковнослужителей о (генеративном) ИИ. Например, в апреле 2023 г. в ведущих СМИ освещался призыв РПЦ запретить роботов с человеческим лицом и голосом. При более детальном рассмотрении такое краткое описание, часто встречающееся в постах в соцсетях, видится не совсем корректным, поскольку это было не отдельным официальным заявлением, а тезисом в докладе иерея на научном мероприятии. На круглом столе в НИУ ВШЭ, посвященном применению и правовому регулированию нейросетевых технологий, председатель Патриаршей комиссии по вопросам семьи, защиты материнства и детства РПЦ иерей Федор Лукьянов заявил «о необходимости установить запрет на антропоморфизацию программ (перенесение человеческого образа и его свойств на неодушевленные предметы) и закрепить на законодательном уровне дисклеймер для нейросетей, указывающий на то, что человек общается с искусственным интеллектом»<sup>10</sup>. Тем не менее в данном случае в фокусе оказывается не только внешний вид роботов, но и маркировка общения, когда с человеком взаимодействует генеративный ИИ.

Другим примером подобных высказываний можно считать авторскую колонку в «Вечерней Москве», написанную Митрополитом Калужским и Боровским Климентом «Нейросеть и эффект сострадания»<sup>11</sup> и получившую отклик среди пользователей соцсетей. Колонка посвящена обсуждению этических вопросов, связанных с генеративным ИИ. Автор демонстрирует хорошее знакомство с самыми последними моделями, выпущенными в массовый релиз, описывая их как «неспособных осмыслить запрос». Завершает свою колонку Митрополит призывом не заменять в общении человека искусственным интеллектом: «Христианство — это религия, в которой человек познает Бога через другого человека. Христос говорит нам о том, что мы служим Ему через сострадание окружающим, через соучастие, сопереживание и деятельную помощь им (ср. Мф. 25: 34—40). Этот бесценный

<sup>9</sup> Закарьян Е. Священник объяснил, почему в РПЦ не пользуются нечеловеческими текстами // Газета.Ru. 2023. 17 ноября. URL: <https://www.gazeta.ru/social/news/2023/11/17/21731251.shtml?ysclid=ltw2jblh2625236851> (дата обращения 20.03.2024).

<sup>10</sup> Устинова А., Киселева А. РПЦ призвала запретить ИИ с человеческим лицом // Ведомости. 2023. 26 апреля. URL: <https://www.vedomosti.ru/society/articles/2023/04/26/972573-rpts-prizvala-zapretit-ii-s-chelovecheskim-litsom?from=newsline> (дата обращения 20.03.2024).

<sup>11</sup> Климент, Митрополит Калужский и Боровский. Нейросеть и эффект сострадания // Вечерняя Москва. 2023. 29 сентября. URL: <https://vm.ru/opinion/1082924-nejroset-i-effekt-sostradaniya?ysclid=ltw4glcxu78675831> (дата обращения 20.03.2024).

опыт должен быть у каждого, потому что он приближает нас к спасению. Через служение реальным людям мы подтверждаем, что хотим быть с Господом вечно в Его Царстве. Этот опыт необходимо актуализировать именно сейчас, когда мы рискуем привыкнуть к тому, что может быть найдена искусственная замена человеку как собеседнику, как объекту нашего внимания»<sup>12</sup>. Последние два примера иллюстрируют, как церковь ищет ответ на вопрос, ранее поднятый в исследованиях: какое влияние технологии могут оказывать на человеческую природу?

### **ИИ в РПЦ: божественное, человеческое, технологическое**

Один из наиболее часто встречающихся в течение 2023 г. поводов обращения к теме религии и генеративного ИИ одновременно — реакция священников РПЦ на нейросети. Отчасти это уже было проиллюстрировано при перечислении популярных событий, попавших в наш корпус материалов. Публикации, которые мы видим на религиозных медийных площадках, мало отличаются от уже описанных кейсов, поскольку и в религиозных, и в светских медийных каналах за основу взят комментарий представителя церкви, выполняющего роль эксперта по теме. Так, в телеграм-канале православного телеканала «Спас» 4 марта 2023 г. была опубликована авторская колонка протоиерея Александра Овчаренко, клирика храма св. Александра Невского при МГИМО МИД России, в которой он, демонстрируя хорошее знакомство с последними технологическими разработками и способами их использования в церквях, задается вопросом о том, насколько необходимы эти технологии. Признавая их удобство, он тем не менее отмечает, что их использование может изменить традиции и разрушить нормальный уклад. Рассуждая о генеративном ИИ, он отмечает: «У искусственного интеллекта нет души (или пока?), он может имитировать эмоции, но не духовность. Chat GPT может быть действительно хорош, но вопрос в том, может ли он быть сочувствующим, есть ли у него понимания жертвы, самопожертвования, восхищения, благодарности? Есть ли у него внутреннее чувство бессмертия и поиск Бога? А совесть? И надо ли это ему иметь? Неужели человечество уже так оскудело преподобными, да что там преподобными, обычными человеческими чувствами, чтобы нам желать этого от ИИ?»<sup>13</sup> И далее он отмечает границы, в которых, на его взгляд, использование этих технологий допустимо: «Пусть роботы и ИИ просто исполняют свои технологические функции — помогают человеку»<sup>14</sup>. Эта авторская колонка в тот же день получила репосты в личных профилях в Вконтакте со ссылкой на телеграм-канал «Спаса». В этом примере представитель церкви не выступает против использования генеративного ИИ в целом, граница возникает там, где генеративный ИИ приравнивается к человеку или божественному.

Интересно, что похожие мысли высказываются и в отдельных комментариях. Так, в телеграм-чате «Святой Ангел. Отрок Вячеслав»<sup>15</sup> к одному из постов был дан комментарий: «Вот так нейросеть и цепляет своими прелестями») Что поделять,

<sup>12</sup> Климент, Митрополит Калужский и Боровский. Нейросеть и эффект сострадания // *Вечерняя Москва*. 2023. 29 сентября. URL: <https://vm.ru/opinion/1082924-nejroset-i-effekt-sostradaniya?ysclid=itw4glcxu78675831> (дата обращения 20.03.2024).

<sup>13</sup> Как Русская Православная Церковь относится к появлению священников-роботов в различных странах? // Телеканал СПАС. 2023. 4 марта. URL: <https://t.me/spastv/14134> (дата обращения 20.03.2024).

<sup>14</sup> Там же.

<sup>15</sup> Святой Ангел. Отрок Вячеслав // Telegram. <https://t.me/OtrokVyacheslavSvyatoy> (дата обращения 20.03.2024)

она же теперь каждому доступна. Есть один православный блогер, он вообще просит ее нарисовать Христа. И она рисует. Только вот считаю опасным доходить до такого». Здесь мы можем зафиксировать аналогичное отношение к технологии — не полный отказ от нее, а отказ от ее использования для генерации религиозного контента. Однако здесь не подчеркивается подмена человеческого технологическим, граница касается только божественного.

Напротив, представители РПЦ, говоря о вреде генеративного ИИ, указывают на опасность неразличения человека и технологии. Так, в продолжение темы генерации контента можно привести комментарий В. Кипшидзе, заместителя председателя синодального отдела Московского патриархата по взаимоотношениям Церкви с обществом и СМИ, которого цитирует сообщество во Вконтакте<sup>16</sup> и сайт<sup>17</sup> «Любимый город Киров»: «Наше христианское свидетельство заключается в том, что общение одного человека с другим, необремененное никакими технологиями, ничем не заменимо. Мы, христиане, это связываем с тем, что человек сотворен по образу и подобию Божию и поэтому человек находит радость, поддержку и силы в общении с другим человеком. На этом, на мой взгляд, строится традиционное общество, — цитирует слова Кипшидзе РИА Новости». В целом, в медийном пространстве можно увидеть детально обозначенную позицию представителей РПЦ в отношении искусственного интеллекта.

Пример использования генеративного ИИ: православные сообщества генерируют изображения, обращаясь к своим подписчикам с просьбой оценить результат. Так, в сообществе «ВКонтакте» «Валаамский монастырь» 25 февраля 2023 г. было опубликовано девять изображений монастыря, сгенерированных Midjourney<sup>18</sup>. Пост вызвал бурный отклик, набрав значительное количество лайков, репостов и комментариев.

Попыток создания текстового контента именно на религиозных площадках нам не встретилось. В течение года было несколько вирусных постов в соцсетях со сгенерированным текстом православной литургии, в репостах которого пользователи делились своими впечатлениями от текста, но это профили отдельных пользователей, не являющихся священниками-блогерами. Это коррелирует и с более широким срезом практик использования ИИ — самым популярным сервисом остается генерация изображений, например открыток к праздникам<sup>19</sup>.

Интересным примером, на наш взгляд, можно считать кейс по разработке приложения в жанре православной визуальной новеллы «Вселенная Нестора», которое создали в Выборгской епархии РПЦ. Приложение, выпущенное весной 2023 г., выросло из журнала «Техника духовной безопасности» и адресовано преимущественно молодежи. Оно рассказывает в формате комиксов о церковной жизни и ее различных аспектах. Выпуск приложения отметили ведущие медиаплощадки РПЦ, такие

<sup>16</sup> Любимый город Киров // Вконтакте. 2023. 12 марта. URL: [https://vk.com/wall-184363760\\_28945?ysclid=luagkr e26j596866701](https://vk.com/wall-184363760_28945?ysclid=luagkr e26j596866701) (дата обращения 20.03.2024)

<sup>17</sup> В РПЦ рассказали, сможет ли искусственный интеллект заменить священника // Любимый город Киров. 2023. 12 марта. URL: <https://lyubimiygorod.ru/kirov/news/16851263> (дата обращения 20.03.2024)

<sup>18</sup> Валаамский монастырь // Вконтакте. 2023. 25 февраля. URL: [https://vk.com/wall-41585629\\_75819](https://vk.com/wall-41585629_75819) (дата обращения 20.03.2024)

<sup>19</sup> 1,5 миллиона открыток к 8 Марта создал миллион пользователей GigaChat от Сбера // Сбер Банк. 2024. 9 марта. URL: [www.sberbank.ru/ru/sberpress/tekhnologii/article?newsID=d8fa7fa5-914c-4772-a8b3-72273f59acc8&blockID=69b149cd-6db4-45aa-ade1-b6920d771b11&regionID=77&lang=ru&type=NEWS](http://www.sberbank.ru/ru/sberpress/tekhnologii/article?newsID=d8fa7fa5-914c-4772-a8b3-72273f59acc8&blockID=69b149cd-6db4-45aa-ade1-b6920d771b11&regionID=77&lang=ru&type=NEWS) (дата обращения 20.03.2024).

как «Азбука веры»<sup>20</sup> и журнал «Фома»<sup>21</sup>. Информация об этом приложении на указанных сайтах не затрагивает вопросы использования генеративного ИИ, однако в «Яндекс.Дзен» на канале «Новые известия» приводятся слова одного из создателей приложения — председателя отдела по делам молодежи епархии протоиерея Артемия Литвинова. По его словам, при создании изображений для приложения была предпринята попытка использовать нейросети, однако она не увенчалась успехом: «Персонажи похожи на результат работы популярной нейросети, которая рисует очеловеченные образы городов и стран. Но, по словам автора проекта, оптимизировать процесс отрисовки героев не удалось. Нейросеть не смогла выдать приемлемый результат, персонажи получались какие-то не такие, поэтому были привлечены художники, которые рисовали в стиле нейронки»<sup>22</sup>. Примечательно, что стиль сгенерированных изображений был сохранен с целью привлечения молодежи.

### Старообрядцы и ИИ: практики и стереотипы

Посты старообрядцев в полученном корпусе данных не столь многочисленны, как посты представителей РПЦ, и связаны в первую очередь с их опытом взаимодействия с генеративным ИИ. Это отдельные сообщения о генерации изображений или восстановлении фотографий при помощи нейросети. Например, в группе «Староверы мира! Семейские» в марте 2023 г. был опубликован запрос на оценку работы с нейросетью.

Рис. 3. Пост с раскрашенной нейросетью фотографией<sup>23</sup>



<sup>20</sup> Молодежный отдел Выборгской епархии представил приложение для Android — новеллу «Вселенная Нестора» // Азбука веры. 2023. 15 марта. URL: <https://azbyka.ru/news/molodezhnyj-otdel-vyborgskoj-eparhii-predstavil-prilozhenie-dlja-android-novellu-vsennaja-nestora> (дата обращения 20.03.2024).

<sup>21</sup> Выборгская епархия создала мобильное приложение, чтобы рассказать молодежи о православии // Журнал ФОМА. 2023. 14 марта. URL: <https://foma.ru/vyborgskaja-eparhija-sozdala-mobilnoe-prilozhenie-chtoby-rasskazat-molodezhi-o-pravoslavii.html?ysclid=lu8yhx8fy2385940387> (дата обращения 20.03.2024).

<sup>22</sup> В Петербурге создали православную визуальную новеллу. Нейросеть с ней не справилась // Новые известия. 2023. 31 марта. URL: <https://dzen.ru/a/ZCaBAE2Lwi50Gs75> (дата обращения 20.03.2024).

<sup>23</sup> Источник: Староверы мира! Семейские // ВКонтакте. 2023. 21 марта. URL: [https://vk.com/wall-172423194\\_746](https://vk.com/wall-172423194_746) (дата обращения 20.03.2024).

Кроме генерации изображений, ИИ использовался для создания аудиоверсий книг, связанных со старообрядческой тематикой. В сообщении автор не только поделился ссылками на соответствующие аудиокниги, но и прокомментировал работу ИИ: «Сперва звучит „на любителя“, но к третьей книге богодухновенный интеллект подстроится...»<sup>24</sup> В конце поста автор сообщает, что проект «Старообрядцы» по старинке озвучивает книги человеческим голосом, тем самым противопоставляя голоса «роботов» и людей.

Отметим, что осмысление технологии генеративного ИИ среди старообрядцев встраивается в процесс привлечения инноваций к трансляции религиозных сообщений для широкой аудитории. Например, в телеграм-канале «#СМЫСЛЬ — Старообрядческая Мысль» в сентябре 2023 г. размещен репост мультимедийного продукта из канала «Глась первый». Он представляет собой поздравление от старообрядческого священнослужителя с комментарием: «С Рождеством Пресвятыя Богородицы!!! Вышедшее на новый технологический уровень поздравление от старообрядческого иерея Игоря Мыльникова из Новокузнецка, демонстрирующее силу сибирского духа, природы и многогранности человеческих талантов во славу Божию!»<sup>25</sup>.

В последних двух примерах технологии работают на передачу религиозных смыслов, принимая участие в коммуникации и распространении знания, связанного с религией.

Интересно, что в личном общении и во время интервью старообрядцы делятся своими практиками взаимодействия с генеративным ИИ (чаще всего говорят о регулярном использовании ChatGPT в качестве помощника), причем обычно предпочитают не сообщать об этом в публичном пространстве, не писать посты на эту тему на личных страницах в соцсетях. Пока идет процесс согласования и выработки норм использования новой технологии на уровне старообрядческих религиозных лидеров и отдельных сообществ, практики взаимодействия с ней остаются во многом скрытыми от внешних наблюдателей. Более того, высокая степень открытости РПЦ к внедрению технологических инноваций нередко воспринимается старообрядцами как все более заметное отклонение от традиционных практик, поэтому отказ от инноваций в некоторых религиозных контекстах вербализуется через проведение границы с последователями РПЦ: «Мы же не РПЦ, чтобы...»

Отдельный блок в собранном корпусе составляют посты, написанные не старообрядцами, но представляющие для нас исследовательский интерес. Они отражают распространенное стереотипное представление о неприятии инноваций в среде старообрядцев, что проявляется в том числе в отсутствии навыков работы с ИИ или даже каких-либо знаний о высоких технологиях.

*Если задал фото низкую экспозицию, значит лопух, гонишься за вышедшей в тираж обработкой. А если все еще ничего не сделал через нейросеть, то, наоборот, старообрядец и догоняющее меньшинство*<sup>26</sup>.

<sup>24</sup> РОБОТЫ ЧИТАЮТ СТАРООБРЯДЦАМ // #СМЫСЛЬ — Старообрядческая Мысль. 2023. 8 июля. URL: <https://t.me/staroveru/1948> (дата обращения 20.03.2024).

<sup>25</sup> Днесь спасению нашему начаток!!! // #СМЫСЛЬ — Старообрядческая Мысль. 2023. 21 сентября. URL: <https://t.me/staroveru/2005> (дата обращения 20.03.2024), а также оригинальный пост: Днесь спасению нашему начаток!!! // Глась первый. 2023. 21 сентября. URL: <https://t.me/glas1y/392> (дата обращения 20.03.2024).

<sup>26</sup> Творческий человек Анна // Instagram. 2023. 23 марта. URL: <https://www.instagram.com/p/CqIbm5CIT0u/> (дата обращения 20.03.2024) (компания Meta и соцсети, которыми она владеет, признаны в России экстремистскими и запрещены).

*О ChatGPT и Midjourney не слышала, наверное, только бабушка Агафья, живущая в лесном массиве Абаканского хребта. Хотя, не удивлюсь, что и до старообрядцев нейронки скоро доберутся<sup>27</sup>.*

Зная о подобных стереотипах, старообрядцы (в том числе для борьбы с ними) активно развивают свой сегмент интернета, повышая видимость своей религии в публичном пространстве, создают и регулярно ведут группы и сообщества в соцсетях, многочисленные телеграм-каналы, где сами рассказывают о себе и своей культуре.

### **Некоторые итоги**

Как показывает анализ данных, использование генеративного искусственного интеллекта религиозными сообществами активно обсуждается в социальных сетях, где пользователи или администраторы страниц/групп делятся сгенерированным контентом и/или оценивают его. Интересно, что на данном этапе большая часть дискуссий выстраивается вокруг попыток установить границы применения этой новой технологии — отчасти в дискуссиях пользователей между собой (чаще всего в комментариях к какому-то посту со сгенерированным контентом или ссылкой на релевантную публикацию в СМИ о возможностях или способностях ИИ), отчасти в распространении позиции официальных представителей церкви. Сама технология и личный опыт ее использования обсуждаются ощутимо реже. В центре внимания оказываются обсуждения опасностей, которые могут быть связаны с использованием генеративных нейросетей, этот фокус превалирует — дискуссий о новых возможностях или экономической выгоде от автоматизации или передачи решения рутинных задач в публичном пространстве нам не встретилось.

При этом важно подчеркнуть, что обсуждения опасностей не затрагивают того, что условно можно обозначить как «утилитарное» использование генеративного ИИ: граница допустимого обозначается в сфере религиозных смыслов (посягательство на «божественное» и на «человеческое как созданное по образу и подобию Божию»). Во многих сообщениях, цитирующих священников, можно встретить указания на то, что повседневное бытовое использование ИИ как инструмента полностью допустимо и даже удобно. Можно также отметить, что участие в обсуждениях допустимых границ применения генеративного ИИ без личного опыта использования приводит к тому, что зачастую в таких дискуссиях воспроизводятся оценки, предлагаемые не только представителями церкви, но и СМИ. Анализ репрезентации генеративного ИИ в мейнстримных медиа видится как одна из потенциальных перспектив для продолжения этого исследования. Вместе с тем, как было отмечено, существует зазор между практиками использования нейросетей и их обсуждением в публичном пространстве: ряд практик, которые можно зафиксировать в личных беседах, не выносятся на публику и остаются невидимыми в поле соцмедиа. В этом отношении перспективой исследования становится этнографическое изучение практик использования генеративного ИИ в религиозных сообществах — через наблюдение или интервью с членами религиозных сообществ.

<sup>27</sup> Ходыкин А. 10 интересных инструментов на базе AI, которые вы могли пропустить // ВКонтакте. 2023. 17 февраля. URL: [http://vk.com/wall45282486\\_1278](http://vk.com/wall45282486_1278) (дата обращения 20.03.2024).

При этом, как и предполагалось, набор практик использования генеративного ИИ довольно узок и пока принципиально не различается между представителями различных православных сообществ. Судя по собранным данным, наиболее широко распространенная практика применения генеративных нейросетей — генерация изображений. В редких случаях работа с нейросетями нацелена на улучшение качества старых фотографий: с одной стороны, такая практика расширяет функционал используемого ИИ, с другой — все еще оставляет нас в границах обработки изображений.

Также среди редких практик использования генеративного ИИ можно отметить обращение к нейросетям для озвучки литературы, тематически связанной с религией и историей сообщества, хотя и существует эта озвучка наряду с текстами, начитанными человеком.

Интересно, что часть практик нацелена на активизацию подписчиков или привлечение новых/молодых сегментов аудитории: в сообществах публикуют сгенерированные изображения, связанные с темой страницы/группы, призывая всех комментировать результат. Нацеленность на привлечение внимания подтверждается и тем, что в случае, когда добиться от генеративного ИИ нужного контента не удастся, создатели могут принять решение сохранить стилистику контента такой, чтобы она выглядела как сгенерированная. Это позволяет как создать устраивающий представителей церкви контент, так и сохранить его актуальность для того сегмента аудитории, который следит за модными веяниями в области технологий. Такие практики использования генеративного ИИ конструируют его скорее как актуальный инфоповод, чем как технологию с высоким потенциалом.

## Список литературы (References)

1. Абрамов А. А. Религиозно-этические проблемы искусственного интеллекта: экспертные оценки и позиция Ватикана // Концепт: философия, религия, культура. 2020. Т. 4. № 4. С. 68—82. <https://doi.org/10.24833/2541-8831-2020-4-16-68-82>.  
Abramov A. A. (2020) Religious and Ethical Issues of Artificial Intelligence: Expert Assessments and the Vatican Position. *Concept: Philosophy, Religion, Culture*. Vol. 4. No. 4. P. 68—82. <https://doi.org/10.24833/2541-8831-2020-4-16-68-82>. (In Russ.)
2. Душакова И. С., Владимиров Е. Д. Сети акторов и контроль над информацией: как РПЦ взаимодействует со светским медийным пространством // Stratum Plus. 2022. № 6. С. 427—438. <https://doi.org/10.55086/sp226427438>.  
Dusacova I. S., Vladimirov E. D. (2022) Networks of Actors and Control over Information: How Russian Orthodox Church Interacts with Secular Media. *Stratum Plus*. Vol. 6. P. 427—438. <https://doi.org/10.55086/sp226427438>. (In Russ.)
3. Хьярвард С. Три формы медиатизированной религии: изменение облика религии в публичном пространстве // Государство, религия, церковь в России и за рубежом. 2020. Т. 38. № 2. С. 41—75.  
Hjarvard S. (2020) Three Forms of Mediatized Religion: Changing the Public Face of Religion. *State, Religion and Church in Russia and Worldwide*. Vol. 38. No. 2. P. 41—75. (In Russ.)

4. Bijker W. E. (2009) Social Construction of Technology. In: Friis J. K. B. O., Pedersen S. A., Hendricks V. F. (eds.) *A Companion to the Philosophy of Technology*. Oxford: Wiley-Blackwell. P. 88—94. <https://doi.org/10.1002/9781444310795.ch15>.
5. Campbell H. (ed.) (2012) *Digital Religion: Understanding Religious Practice in New Media Worlds*. London: Routledge.
6. Campbell H. A., Cheong P. H. (2023) Introduction. In: Campbell H. A., Cheong P. H. (eds.) (2023) *Thinking Tools on AI, Religion & Culture*. College Station, TX: Digital Religion Publications. P. 7—8.
7. Campbell H. A., Cheong P. H. (eds.) (2023) *Thinking Tools on AI, Religion & Culture*. College Station, TX: Digital Religion Publications.
8. Campbell H. A., Evolvi G. (2019) Contextualizing current digital religion research on emerging technologies. *Human Behavior & Emerging Technologies*. No. 2. P. 5—17. <https://doi.org/10.1002/hbe2.149>.
9. Ekbja H. (2010) Fifty Years of Research in Artificial Intelligence. *Annual Review of Information Science and Technology*. Vol. 44. P. 201—242. <https://doi.org/10.1002/aris.2010.1440440112>
10. Elliott A. (2022) The Complex Systems of AI: Recent Trajectories of Social Theory. In: Elliott A. (ed.) *The Routledge Social Science Handbook of AI*. New York, NY: Routledge. P. 3—16.
11. Falahatpisheh Z., Khajeheian D. (2020) Affordances and IT Design: A Typology for Social Media and Platform Affordances. *13<sup>th</sup> CMI Conference on Cybersecurity and Privacy (CMI) — Digital Transformation — Potentials and Challenges (51275), Copenhagen, Denmark*. P. 1—7. <https://doi.org/10.1109/CMI51275.2020.9322681>
12. Jacoba R. C. (2023) Exploring the Role of Artificial Intelligence in Interreligious Discourse. *Religion and Social Communication*. Vol. 21. No. 2. P. 375—400. <https://doi.org/10.62461/RCJ100323>
13. Pinch T. J., Bijker W. E. (1984) The Social Construction of Facts and Artefacts: Or How the Sociology of Science and the Sociology of Technology Might Benefit Each Other. *Social Studies of Science*. Vol. 14. No. 3. P. 399—441. <https://doi.org/10.1177/030631284014003004>
14. Rahman F. A., Mohammed R. R. AI in the Middle East: Balancing Cultural Identity, Gender Dynamics, and Religious Perspectives. In: Campbell H. A., Cheong P. H. (eds.) *Thinking Tools on AI, Religion & Culture*. College Station, TX: Digital Religion Publications. P. 17—20.
15. Singler B. (2023) How I Stopped Worrying and Learned to Question the Apocalyptic AI. In: Campbell H. A., Cheong P. H. (eds.) *Thinking Tools on AI, Religion & Culture*. College Station, TX: Digital Religion Publications. P. 10—11.

DOI: [10.14515/monitoring.2024.5.2593](https://doi.org/10.14515/monitoring.2024.5.2593)

Н. Д. Трегубова, А. Я. Фейгина

**«ВЕЖЛИВОЕ» ВЗАИМОДЕЙСТВИЕ С УМНЫМИ КОЛОНКАМИ  
В ПОВСЕДНЕВНОЙ ЖИЗНИ: УНИВЕРСАЛЬНЫЕ НОРМЫ  
И НОВЫЕ ФОРМЫ СОЦИАЛЬНОСТИ****Правильная ссылка на статью:**

Трегубова Н. Д., Фейгина А. Я. «Вежливое» взаимодействие с умными колонками в повседневной жизни: универсальные нормы и новые формы социальности // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 142—165. <https://doi.org/10.14515/monitoring.2024.5.2593>.

**For citation:**

Tregubova N. D., Feigina A. Y. (2024) "Polite" Interactions with Smart Speakers in Everyday Life: New Norms and Forms of Sociality. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 142–165. <https://doi.org/10.14515/monitoring.2024.5.2593>. (In Russ.)

Получено: 20.03.2024. Принято к публикации: 07.08.2024.

## «ВЕЖЛИВОЕ» ВЗАИМОДЕЙСТВИЕ С УМНЫМИ КОЛОНКАМИ В ПОВСЕДНЕВНОЙ ЖИЗНИ: УНИВЕРСАЛЬНЫЕ НОРМЫ И НОВЫЕ ФОРМЫ СОЦИАЛЬНОСТИ

*ТРЕГУБОВА Наталья Дамировна — кандидат социологических наук, доцент кафедры сравнительной социологии, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия*  
E-MAIL: [n.tregubova@spbu.ru](mailto:n.tregubova@spbu.ru)  
<https://orcid.org/0000-0003-3259-5566>

*ФЕЙГИНА Анна Яковлевна — аспирантка кафедры сравнительной социологии, Санкт-Петербургский государственный университет, Санкт-Петербург, Россия*  
E-MAIL: [anna.feigina97@gmail.com](mailto:anna.feigina97@gmail.com)  
<https://orcid.org/0009-0000-8026-7482>

**Аннотация.** Осмысление использования инструментов и технологий искусственного интеллекта (ИИ) в повседневной жизни становится важной составляющей развития социологических исследований. Одно из ключевых направлений изучения взаимодействия человека с ИИ — рассмотрение новых социальных и этических норм в повседневной жизни. Статья представляет результаты эмпирического исследования, цель которого заключается в выяснении того, какие нормы и формы вежливости используют люди при коммуникации с умными колонками. Исследование предполагает определение характера взаимодействий, осмысление процесса «одомашнивания», анализ проявлений вежливости в отношении «разговорных» инструментов ИИ. Теоретические основания работы включают социологию повседневности, концепцию искусственной коммуникации Е. Эспозито и концепцию эвокативного объекта Ш. Теркл. Проблема вежливости в процессе взаимодействия с умными колонками анализируется в терминах концепции «поддер-

## “POLITE” INTERACTIONS WITH SMART SPEAKERS IN EVERYDAY LIFE: NEW NORMS AND FORMS OF SOCIALITY

*Natalia D. TREGUBOVA<sup>1</sup> — Cand. Sci. (Soc.), Associate Professor, Comparative Sociology Department*  
E-MAIL: [n.tregubova@spbu.ru](mailto:n.tregubova@spbu.ru)  
<https://orcid.org/0000-0003-3259-5566>

*Anna Ya. FEIGINA<sup>1</sup> — postgraduate student, Comparative Sociology Department*  
E-MAIL: [anna.feigina97@gmail.com](mailto:anna.feigina97@gmail.com)  
<https://orcid.org/0009-0000-8026-7482>

<sup>1</sup> Saint Petersburg State University, Saint Petersburg, Russia

**Abstract.** Understanding the use of artificial intelligence (AI) tools and technologies in everyday life is an important task for contemporary sociological research. One of the key issues in the analysis of human-AI interaction is the consideration of new social and ethical norms in everyday life. This paper presents the results of an empirical study aimed at revealing the norms and forms of politeness that people use when communicating with smart speakers. The theoretical foundations of the research embrace the sociology of everyday life, the concept of artificial communication by Elena Esposito, and the concept of the evocative object by Sherry Turkle. The problem of politeness in interaction with a smart speaker is characterized in terms of Erving Goffman’s concept of face and the concept of linguistic politeness.

The authors formulate research questions and assumptions that are tested based on the analysis of 18 interviews with users of smart speakers in St. Petersburg, Tel Aviv, and Berlin. A comparative analysis of interviews with

жания лица» И. Гофмана и концепции лингвистической вежливости.

В статье формулируются исследовательские вопросы и предположения, которые проверяются на основе анализа 18 интервью с пользователями умных колонок в Санкт-Петербурге, Тель-Авиве и Берлине. Сравнительный анализ материалов интервью с информантами из трех стран показывает, что нормы вежливости, характерные для пользователей при коммуникации с умными колонками, сходны в разных странах. В таком взаимодействии совмещаются элементы лингвистической вежливости (типичные вежливые фразы) и проявления вежливости как поддержание лица — своего и собеседника. По результатам исследования не было выявлено зависимости между тем, как люди используют умные колонки, и тем, какой концепции вежливости в отношении нее придерживаются. Очеловечивание умных колонок оказывается связанным не с характером их использования, а с их универсальной способностью вступать в разговор.

**Ключевые слова:** умные колонки, разговорный ИИ, нормы вежливости, искусственная коммуникация, одомашнивание технологий, эвкативный объект

**Благодарность.** Авторы выражают благодарность всем информантам за содержательные интервью и искреннюю заинтересованность в исследовании. Особенно следует выделить информантов из города Берлин за их фантастическую отзывчивость, ответственность и неравнодушие. Также выражаем признательность всем, кто помогал в поиске информантов, за их искреннее участие и желание помочь.

informants demonstrates that the norms of politeness typical for users in communication with smart speakers are similar in different countries. Within such interactions, users combine elements of linguistic politeness (typical polite phrases) and manifestations of politeness such as to save the face, one's own and that of the interlocutor. At the same time, the results of the study did not reveal a relationship between how people use smart speakers and the concept of politeness they adhere to in relation to them. Anthropomorphizing of smart speakers turns out to be associated not with the nature of its use, but with its universal ability to enter a conversation.

**Keywords:** smart speakers, conversational AI, norms of politeness, artificial communication, domestication of technologies, evocative object

**Acknowledgments.** The authors thank all informants for their interviews and sincere interest in the study. Informants from Berlin deserve special mention for their fantastic responsiveness, responsibility, and concern. The authors would also like to thank everyone who helped in finding informants for their participation and desire to help.

## Введение

Общество претерпевает фундаментальные трансформации, связанные с вхождением технологий искусственного интеллекта (ИИ) в жизнь людей на различных уровнях и в разных сферах. В данной статье речь пойдет об использовании технологий ИИ в повседневной жизни на примере эмпирического исследования взаимодействия/коммуникации<sup>1</sup> пользователей с умными колонками<sup>2</sup> в домашнем пространстве. Особое внимание мы уделяем проявлениям различных форм вежливости в процессе такой коммуникации.

В социальных науках существует большое число как отечественных, так и зарубежных исследований, посвященных различным аспектам использования людьми умных колонок. Умная колонка — это объект, который находится в домашнем пространстве, и процессы ее одомашнивания представляют особый интерес для ученых [Brause, Blank, 2020; Корбут, 2021; Kucharski, Merkel, 2023]. Еще одна характерная черта умных колонок заключается в том, что коммуникация с ними происходит посредством голоса, поэтому часть работ посвящена разговорам пользователей с умными колонками [Smith et al., 2021; Ушкин, Коваль, Яськин, 2022]. Большое внимание также уделяется проблеме конфиденциальности и приватности, так как пользователи полагают, что умные колонки могут их прослушивать [Malkin et al., 2019; Ушкин, Коваль, 2023; Kang, Oh, 2023].

Вместе с тем проблема вежливости при взаимодействии с умными колонками изучена не столь хорошо. В российском сегменте научных исследований нам не удалось обнаружить публикации, посвященные проблеме вежливости при взаимодействии с ИИ, в частности с умными колонками, хотя этот вопрос волнует некоторых зарубежных исследователей, занимающихся проблематикой взаимодействия людей с технологиями [Burton, Gaskin, 2019; Ribino, 2023]. Наше исследование — это попытка выяснить, на какие нормы вежливости ориентируются люди в процессе коммуникации с одомашненной умной колонкой.

На основании данных, собранных в трех городах — Санкт-Петербурге (Россия), Тель-Авиве (Израиль) и Берлине (Германия), мы проанализировали взаимодействие пользователей с одомашненными умными колонками «Алиса», «Сири» и «Алекса». Мы выделяем ключевые тенденции в формировании норм и форм коммуникации пользователей с новыми для них объектами одомашнивания.

Дальнейшее изложение организовано следующим образом. Мы начнем с описания теоретических оснований исследования, затем обратимся к его методологии. Далее будет приведен анализ интервью с пользователями умных колонок, раскрывающий особенности взаимодействия с этими технологиями, включая проявления вежливости в разговоре с ними. В заключении будут сделаны выводы в отношении исследовательских вопросов и предположений, выделены

<sup>1</sup> Термины «взаимодействие» и «коммуникация» используются в настоящем исследовании как контекстуальные синонимы с разными оттенками смысла. «Взаимодействие» (*interaction*) в рамках микросоциологической традиции характеризует взаимно направленное поведение участников ситуации, обусловленное свойствами этой ситуации и определяющее ее динамику. «Коммуникация» характеризует вербальное взаимодействие (разговор) участников, в ходе которого они обмениваются информацией. Оба термина характеризуют то, что происходит между человеком и умной колонкой, и коммуникация в данном случае может рассматриваться как часть процесса взаимодействия.

<sup>2</sup> Вслед за Андреем Корбутом мы понимаем умную колонку как особую технологию ИИ (портативную колонку с интегрированным голосовым помощником), которая стала новым элементом домашней социальности [Корбут, 2021].

тенденции, характерные для «вежливого» взаимодействия с умными колонками, рассмотрены ограничения проведенного анализа и намечены направления для будущих исследований.

### **Теоретические основания исследования**

Теоретические основания исследования подразделяются на две составляющие: характеризующие собственно взаимодействие/коммуникацию пользователя с умной колонкой и определяющие «вежливое» взаимодействие с ней. Чтобы объяснить, как технология ИИ в принципе может выступать участником коммуникации, поддерживая осмысленный с точки зрения пользователя разговор, мы обращаемся к концепции искусственной коммуникации. Чтобы осмыслить, как пользователи воспринимают умную колонку в процессе взаимодействия с ней, мы опираемся на концепцию эвокативного объекта. Чтобы характеризовать контекст взаимодействия пользователя с умной колонкой в пространстве дома, мы обращаемся к социологии повседневности. Для осмысления вежливости в отношении умной колонки мы опираемся на концепцию поддержания лица Ирвинга Гофмана и на лингвистическую концепцию вежливости.

#### *Искусственная коммуникация*

В рамках концепции искусственной коммуникации (*artificial communication*) Елена Эспозито [Esposito, 2017] рассматривает алгоритмы в качестве социальных агентов и партнеров для коммуникации. Эспозито опирается на теорию коммуникации Никласа Лумана [Luhmann, 1992], согласно которой обязательным условием коммуникации оказывается объединение трех компонентов: сообщения, информации и понимания. Важным аспектом теории является разграничение информации и сообщения, а также понимание ее участниками содержания передаваемой информации.

Эспозито ставит вопрос о том, что происходит в процессе коммуникации, когда один из ее партнеров представляет собой совокупность алгоритмов. Сохраняются ли в таком случае необходимые условия коммуникации? Если сообщение и информация присутствуют в такой форме коммуникации, то что происходит с ключевым аспектом — пониманием содержания и значения информации?

Для того чтобы ответить на данные вопросы, Эспозито включает в свою концепцию понятие контингентности и рассуждает о том, какую форму приобретает контингентность в процессе искусственной коммуникации. Контингентность — обязательное условие коммуникации по Луману. Это возможность множества вариантов ответов в процессе коммуникации: контингентность предполагает неопределенность в том смысле, что мы не можем знать, как себя поведет и что скажет партнер по коммуникации.

Вопрос состоит в том, какую форму приобретает контингентность при коммуникации с технологией ИИ, для которой не существует понятия неопределенности. Специфика умных колонок заключается в том, что эти устройства взаимодействуют с миллионами пользователей, они учатся на основе запросов и информации, получаемой из них. Ответы колонок отражают запросы множества пользователей, коммуницирующих с ними; они не повторяют ответы конкретного пользова-

теля и при этом не заложены в явном виде разработчиками ИИ. Таким образом, коммуникация человека с умными колонками<sup>3</sup> представляет собой новый вид коммуникации, обладающий новым типом контингентности.

### *Эвокативный объект и очеловечивание*

В рамках настоящего исследования умные колонки рассматриваются как эвокативный объект (*evocative object*). Данный термин был использован Шерри Теркл для характеристики отношения человека к компьютеру — новой технологии, которая одновременно и похожа, и не похожа по своим свойствам и активностям на человека [Turkle, 2005]. Характеризуя умную колонку как эвокативный объект, можно указать на следующие ее свойства:

1. Технология как эвокативный объект способна менять отношение людей к окружающей реальности. Умная колонка не только изменяет представления людей об организации повседневной жизни, но и позволяет формулировать более глубокие вопросы о человеческом мышлении, о месте людей в мире, где искусственный интеллект находится так близко, об отношениях к таким изменениям.

2. Технология как эвокативный объект изменяет способы мышления людей. Технологии, которые имитируют человеческую речь и воспринимаются людьми как нечто, способное «думать», побуждают человека задаться вопросами о том, как работает его собственное мышление.

3. Предположение людей о том, что технология способна к мыслительной деятельности, тесно связано с феноменом антропоморфизма — очеловечивания [Salles, Evers, Farisco, 2020]. Мышление людей трансформируется в том числе и из-за того, что они очеловечивают умную колонку, с которой вступают в коммуникацию. Умная колонка, таким образом, становится в глазах ее пользователя не просто объектом, а чем-то более приближенным к живому существу.

Очеловечивание технологий ИИ — это не условие, а результат взаимодействия с ними [Корбут, 2021]. Можно предположить, что очеловечивание одомашненной умной колонки — процесс, который происходит постепенно путем взаимодействия с ней в повседневной жизни. Вопрос очеловечивания умной колонки актуален потому, что она может приобретать некоторые человеческие черты [Li, Suh, 2021], самая очевидная из которых — способность воспроизводить человеческую речь. Вербальный характер коммуникации с умной колонкой — одна из ключевых причин очеловечивания этой технологии, что, в свою очередь, может влиять на характер взаимодействия с ней [Moussawi, Koufaris, Benbunan-Fich, 2021] — на проявление вежливости в процессе коммуникации.

### *Одомашнивание умной колонки в повседневной жизни*

Вхождение технологий ИИ в повседневную жизнь людей позволяет вести рассуждение о новых типах повседневных практик и взаимодействий. Процесс одомашнивания умной колонки тесно связан с трансформацией привычных повседневных домашних практик и появлением новых, связанных непосредственно с использованием этой технологии. К проблеме одомашнивания умной колонки

<sup>3</sup> А также с другими технологиями ИИ вроде онлайн-поисковиков, генераторов изображений и чат-ботов, основанных на больших языковых моделях.

обращались как российские [Корбут, 2021; Ушкин, Коваль, 2023], так и зарубежные исследователи [Brause, Blank, 2020; Kucharski, Merkel, 2023]. Умная колонка рассматривается нами в качестве одомашненного объекта, с которым пользователь взаимодействует в повседневной жизни.

Для объяснения того, каким образом умная колонка может встраиваться в повседневную жизнь людей и менять ее, мы обратимся к этнометодологическому подходу [Maunard, Clayman, 1991] и концепции повседневной жизни [Sztompka, 2008]. Этнометодология позволяет объяснить, как люди воспроизводят новую социальную реальность, в нашем случае — с помощью новых форм взаимодействий с умной колонкой в домашнем пространстве в рамках повседневности. В настоящее время мы наблюдаем новый этап в конструировании повседневной жизни путем апробации различных методов взаимодействия с умной колонкой как с новым объектом одомашнивания.

В то время как этнометодология позволяет оценить то, как конструируется новая социальная реальность, концепция повседневной жизни дает возможность оценить конкретные повседневные практики. По этой причине мы обращаемся к работе Петра Штомпки [Sztompka, 2008], выделяя наиболее важные для нашего исследования ее составляющие. События в повседневной жизни носят рутинный и циклический характер, они в каком-то смысле ритуализированы. Мы связываем использование различных функций умной колонки с повседневными практиками, которые обладают свойством повторяемости и рутинизации. Необходимо оценить, насколько умная колонка и практики ее использования стали частью повседневной жизни людей. Эпизоды повседневной жизни происходят в конкретных местах, в нашем случае таким местом выступает домашнее пространство. Решение о расположении умной колонки чаще всего связано с тем, какие повседневные практики в каких местах удобнее совершать при участии умной колонки.

Каким образом люди одомашнивают новый технологический объект — умную колонку? Исследователи выделяют несколько последовательных этапов процесса одомашнивания, такие как коммодификация, объективизация, инкорпорация и конверсия [Brause, Blank, 2020]. Каждый из этих этапов выражает различные аспекты, например характер взаимодействия, решение о физическом расположении колонки и трансформацию повседневных привычных практик.

Существуют также некоторые условия, необходимые для того, чтобы колонка гармонично встраивалась в повседневные домашние практики [Корбут, 2021]. В качестве таких условий выступают фоновый характер и доступность умной колонки для управления голосом, а также коллективное использование колонки обитателями домашнего пространства.

Таким образом, наша задача заключается в том, чтобы посмотреть, как умная колонка становится частью повседневной жизни в рамках домашнего пространства. То есть каким образом пользователи строят коммуникацию с умной колонкой в процессе осуществления различных практик.

### *Проблема вежливости: конструируя «лицо» колонки*

Концепция, связанная с работой лица, сформулированная Ирвингом Гофманом, используется в нашем исследовании для того, чтобы проанализировать, как

пользователи конструируют свое лицо или «лицо» самой умной колонки в процессе взаимодействия с ней.

В рамках концепции Гофмана [Goffman, 1955] лицо рассматривается в качестве социальной ценности человека, которую он утверждает и воспроизводит в процессе выстраивания определенной линии поведения. В свою очередь, линия поведения — это набор действий (вербальных или невербальных), через которые человек выражает свой взгляд на ситуацию и оценивает себя и окружающих. Важным аспектом концепции лица является то, что путем работы лица, человек конструирует свое собственное «Я» [Watts, 2003]. Таким образом, лицо складывается из интерпретаций в процессе взаимодействий как других, так и самого человека. Лицо также является социальным, то есть оно приобретает смысл только в процессе социальных взаимодействий.

Под работой лица Гофман подразумевает действия человека, которые соответствуют конструируемому им и окружающими лицу [Goffman, 1955]. Основным инструментом работы лица является его сохранение (*face-saving*). Этот инструмент включает в себя практики, стандарты и паттерны поведения, направленные на поддержание образа лица перед другими и перед самим собой. В ситуациях, когда не удастся сохранить позитивный образ лица (например, при неодобрении, оскорблении или обиде), работа лица должна быть направлена на восстановление его позитивного образа.

В отношении концепции лица и умной колонки мы стремимся ответить на следующие вопросы: влияет ли колонка на конструирование пользователями их лица? Как пользователи относятся к «лицу» самой умной колонки? Стремятся ли пользователи восстановить свое лицо перед колонкой? Что происходит в случае невозможности сохранения «лица» колонки? Мы будем рассматривать эти вопросы на примере обиды и извинения перед умной колонкой в процессе коммуникации с ней.

#### *«Спасибо, Алиса!»: к вопросу о лингвистической вежливости*

Концепция лингвистической вежливости значима для исследования потому, что форма коммуникации человека с умной колонкой носит исключительно вербальный характер [Nwoye, 1992]. Лингвистическая вежливость включает в себя конкретные слова, выражения или предложения, способствующие коммуникации. Исследователи выделяют два типа выражения лингвистической вежливости [Watts, 2003]:

1. Шаблонные высказывания — наиболее общепринятые проявления вежливости при социальных взаимодействиях, используемые в ритуализированных и направленных формах вербальной коммуникации. Это могут быть благодарности или просьбы, извинения, приветствия или прощания.

2. Полуформальные высказывания — выражения вежливости, используемые в процессе вербальной коммуникации, но не обладающие формой конкретного обращения. Например, выражение солидарности с кем-либо («Да, ты прав»), модальные выражения («Не мог бы ты», «Могу ли я»).

В контексте нашего исследования нас в первую очередь будут интересовать такие проявления вежливости, как благодарность, просьба и извинение.

## Исследовательские вопросы и предположения

На основании теорий и концепций, рассмотренных выше, были сформулированы исследовательские вопросы и предположения.

К ключевым вопросам нашего исследования относятся следующие: в чем заключается специфика взаимодействия пользователей с одомашненной умной колонкой? Как можно формировать и развивать вежливое взаимодействие с разговорным ИИ? Зависят ли проявления вежливости в отношении умной колонки от характера ее включения в повседневную жизнь пользователей? Существуют ли универсальные или культурно-специфичные нормы вежливости во взаимодействии с умной колонкой?

Исследовательские предположения подразделяются на две группы. Первая связана с характеристикой вежливости пользователей в отношении умных колонок. Здесь возможны два варианта: либо определенные формы вежливости в отношении умных колонок уже сложились, либо их не существует. В первом случае эти нормы могут быть универсальными или культурно-специфичными. Вторая группа предположений связывает характер использования умной колонки, процессы ее очеловечивания и характер вежливости, проявляемый пользователями в отношении нее.

Сформулируем наши предположения в виде тезисов и контртезисов.

**Тезис 1.** Существуют определенные нормы вежливости при коммуникации с одомашненной умной колонкой.

**Контртезис 1.** Не существует никаких определенных норм вежливости при коммуникации с одомашненной умной колонкой<sup>4</sup>.

**Тезис 1.1.** Нормы вежливости при коммуникации с одомашненной умной колонкой культурно обусловлены. В разных странах они будут различаться.

**Контртезис 1.1.** Нормы вежливости универсальны. Они связаны с отношениями к технологиям, а не с культурами разных стран.

**Тезис 2.** Люди, которые вступают в коммуникацию с одомашненной умной колонкой в целях решения конкретных задач, не склонны очеловечивать ее и будут использовать лингвистическую вежливость. Люди, для которых умная колонка выступает в качестве социального партнера, склонны очеловечивать ее и стремятся сохранить лицо в процессе вежливого взаимодействия.

**Контртезис 2.** Не существует зависимости между тем, как люди используют колонку, и тем, какой концепции вежливости они придерживаются.

## Методология исследования

Сбор данных был реализован с применением метода полуструктурированного интервью. Гайд интервью разрабатывался на основании исследовательских вопросов и предположений.

Качественные методы, в частности метод полуструктурированного интервью, были выбраны по следующим причинам:

1) исследование представляло собой первый шаг в области изучения проблем вежливости при взаимодействии с ИИ; для нас было важно выделить конкретные кейсы на основе ограниченного числа интервью;

<sup>4</sup> Это предполагает в том числе случаи полного отсутствия проявлений вежливости по отношению к умной колонке со стороны пользователя.

2) выделенные кейсы подлежали подробному анализу и установлению связей между ними для подтверждения или опровержения сформулированных тезисов и контртезисов;

3) кейсы, выделенные по результатам анализа данных, позволили сделать выводы о взаимосвязи между характером взаимодействия с умной колонкой и проявлениями вежливости в отношении нее. Полученные результаты позволяют выявить и осмыслить типичные практики, мотивы и способы использования умных колонок, валидность которых затем может быть проверена в ходе количественных исследований.

Сбор данных проходил с 13 декабря 2023 г. по 26 января 2024 г. среди информантов из трех городов — Санкт-Петербурга, Тель-Авива и Берлина. Было проведено 18 интервью — по шесть в каждом из городов. Ограниченное число интервью обусловлено тем, что исследование является поисковым. Можно говорить о том, что это первый этап в ряде исследований, посвященных вежливой коммуникации с умной колонкой. Интервью проводились преимущественно с применением информационно-коммуникационных технологий. Лишь одна информантка из Санкт-Петербурга пожелала, чтобы интервью было проведено лично.

В качестве информантов были отобраны пользователи умных колонок, соответствующие следующим критериям:

1) информант на момент проведения интервью является действующим пользователем умной колонки (колонка включена и находится в постоянном пользовании);

2) информант является постоянным резидентом страны и города, заявленных им в интервью;

3) независимо от страны проживания, информант говорит на русском языке, который для него (нее) является родным. Язык коммуникации с умной колонкой не учитывался и мог быть любым.

Проведение интервью с русскоязычными пользователями в трех крупных городах, с одной стороны, обеспечивает сравнимость пользовательского опыта и облегчает установление контакта с информантами. С другой стороны, оно дает достаточное культурное разнообразие в отношении проявлений вежливости. Детальное описание информантов представлено в виде таблицы в Приложении 1.

Анализ данных был выстроен в соответствии со стандартной схемой анализа качественных интервью. Путем открытого кодирования в программе Atlas.ti 9 выделялись категории, связанные с различными блоками вопросов в интервью. Выбор метода открытого кодирования обусловлен тем, что не было необходимости в предварительном знакомстве с данными и составлении подробной книги кодов с описаниями: все данные изначально были знакомы, так как были собраны самостоятельно. Всего в процессе кодирования было выделено 17 категорий, некоторые из которых были разбиты на подкатегории. Список выделенных в процессе анализа категорий и подкатегорий представлен в Приложении 2.

Встречаемость кодов варьировалась в диапазоне от 9 до 90 упоминаний на весь массив данных (18 интервью). Самое большое количество цитат пришлось на коды «Используемые функции: примеры» — 90 цитат, «Коммуникация с умной колонкой: характер» — 71 цитата и «Вежливость с умной колонкой» — 52

цитаты. В остальных случаях количество цитат, приходившихся на один код, составляло в среднем 30 упоминаний на весь массив данных.

## Результаты исследования

### Одомашнивание умных колонок

Взаимодействие с умными колонками, подлежащее анализу в рамках настоящего исследования, происходит исключительно в рамках домашнего пространства. Умная колонка — это в первую очередь объект одомашнивания, которое становится возможным путем повседневных взаимодействий, а также трансформации существующих практик и формирования новых.

Целый блок вопросов в интервью был посвящен практикам использования умных колонок в домашнем пространстве. Самые распространенные места расположения умных колонок дома — кухня и гостиная (в некоторых случаях они объединены). Обычно они расположены там, где наши информанты и другие обитатели дома проводят больше всего времени. Практики, с которыми может быть связано такое местоположение, — прослушивание музыки, которое часто приобретает коллективный характер, постановка таймеров и будильников в процессе готовки, напоминаний, а также использование возможностей, предоставляемых программным обеспечением колонки, называемым «Умный дом». Самая распространенная практика, связанная с этой системой, — включение и выключение света с помощью умных лампочек.

Практики, осуществлению которых способствует умная колонка, создают общий фон пользования ею. Далее мы рассмотрим специфику выстраивания и общий характер коммуникации с умной колонкой на фоне существующих практик в домашнем пространстве.

### Нормы вежливости во взаимодействии с умной колонкой

Логика представления результатов анализа данных будет выстроена так, чтобы была возможность проследить подтверждение или опровержение выдвинутых в исследовании предположений. Тезис 1 заключался в том, что существуют определенные нормы вежливости при взаимодействии с умной колонкой. Контртезис 1 гласил, что никаких норм вежливости при взаимодействии с умной колонкой не существует.

В целях их проверки обратимся к эмпирической базе исследования. Первое, о чем необходимо упомянуть: вежливая коммуникация и взаимодействие с умной колонкой характерны для всех пользователей, принявших участие в исследовании. Однако причины, по которым информанты идут по пути вежливого взаимодействия, разные. Для начала отметим, что вежливое взаимодействие с колонкой рассматривается информантами как часть культуры вежливого общения.

*Вообще я считаю, что каждый человек должен быть вежлив. В принципе, изнутри и по сути своей. И почему, если я должна быть... Ну, если я вежлива с одними, я не вежлива с другими, пусть это даже Алиса. (Информант № 5, Ж, Санкт-Петербург)*

*Не знаю, как-то по общему, как сказать, ощущению, просто так вот, что со всеми надо быть вежливым, с колонкой тоже. (Информант № 10, Ж, Берлин)*

Таким образом, пользователи не склонны разграничивать вежливое взаимодействие с умной колонкой и с людьми в реальной жизни. Однако эти утверждения характеризуют скорее общий характер взаимодействия. В частных случаях ситуация может выглядеть по-другому.

Проявления вежливости тесно связаны с конструированием пользователями своего лица. Здесь дело не в том, что «подумает» сама умная колонка. Некоторые не видят смысла обращаться с колонкой невежливо из собственных соображений:

*Ну, тут дело, да, не только в том, что она ничего не чувствует, просто зачем мне это делать? То есть в чем смысл в этом. Я же не оскорбляю людей на улице, бабушек, собак, детей. Зачем мне говорить это колонке? (Информант № 6, М, Санкт-Петербург)*

*Мне неприятно самому такое говорить [грубое]. Стараюсь такого избегать. (Информант № 13, М, Тель-Авив)*

В других случаях проявления вежливости при коммуникации с умной колонкой тесно связаны с очеловечиванием этой технологии:

*Ну, я как бы ставила себя... Ну, представляла, что если бы она была реальным человеком, то, наверное, ей было бы неприятно. Поэтому как бы стараюсь так все равно более приветливо к ней относиться. (Информант № 16, Ж, Тель-Авив)*

Вежливость оказалась характерной чертой коммуникации с умной колонкой. Очеловечивание умной колонки, а также конструирование своего лица как лица вежливого человека оказывают влияние на характер коммуникации с колонкой.

### *Вежливость, язык и культура*

Продолжим проверять первую группу исследовательских предположений и обратимся к тезису 1.1, согласно которому нормы вежливости культурно обусловлены и в разных странах различаются.

Вежливое обращение с колонкой может проявляться в различных формах благодарности к ней. При этом по результатам исследования выяснилось, что эти формы универсальны.

*Ну, «спасибо», «благодарю». На запросы отвечаю какие-то, если помогла. Я благодарю ее. (Информант № 15, М, Тель-Авив)*

Благодарность умной колонке — это не столько культурно обусловленная норма, сколько, по словам одного из информантов, общая социальная норма вежливого взаимодействия в обществе:

*Ну, спасибо я ей говорю периодически, просто так, как-то это тоже скорее даже потому, что это уже настолько социальная норма, социальная норма просто в моей голове. Вот мне кто-то помог — я им скажу спасибо. И поэтому периодически я говорю «спасибо». (Информант № 9, М, Берлин)*

Продолжая линию рассуждения о культуре, следует отметить, что если нормы вежливости не обусловлены культурным контекстом, то характер коммуникации может зависеть от языка, на котором информант разговаривает с колонкой:

*Да, ну, это еще такая небольшая практика свободы английского мата. Я как бы запинаясь даже в этих вещах. Я на нее ору. А тут я могу как бы повыпендриваться перед собой. Ну, в общем, да, вот такая история<sup>5</sup>. (Информант № 14, М, Тель-Авив)*

Добавим, что некоторые языковые и культурные особенности страны проживания могут оказывать влияние на характер коммуникации с колонкой. Это связано не столько с нормами вежливости, сколько с привычными синтаксическими конструкциями, используемыми в языке:

*...На иврите прямо есть такое слово «дай» — «хватит». И тут можно сказать такое на улице человеку или «прекрати» — «товсик», «прекрати». Коротко, одним словом, сразу «ты» и «прекрати» — это «товсик». (Информант № 15, М, Тель-Авив)*

Упомянутые формы повелительного наклонения не считаются в иврите невежливым обращением. Это привычные формы коммуникации, не несущие в себе негативного подтекста, однако оказывающие влияние на характер коммуникации русскоязычного пользователя с русскоязычной умной колонкой.

Таким образом, подтверждается *контртезис 1.1*, согласно которому нормы вежливости при коммуникации с умной колонкой оказываются общими и универсальными в разных странах. Нормы вежливости связаны скорее с отношением к людям в обществе, которое пользователи проецируют на взаимодействие с умной колонкой. Вместе с тем характер коммуникации с умной колонкой может трансформироваться под влиянием культуры страны проживания.

### *Очеловечивание умной колонки*

Обратимся ко второй группе исследовательских предположений, которая позволит сделать выводы о существовании возможной связи очеловечивания и характера использования умной колонки.

Первое, что следует отметить: помощь умной колонки, пусть даже и инструментальная, оказывает влияние на очеловечивание ее пользователями. То есть люди склонны очеловечивать эту технологию независимо от характера использования:

*Да, приятно, конечно. То есть я чувствую более близкую связь, и эта помощница, она превращается в более живого человека в таком случае. Мне важно, чтобы у по-*

<sup>5</sup> Информант № 14 общается с умной колонкой (*HomePod Siri*) на английском языке. По его словам, нецензурная лексика в адрес колонки — это не столько проявление невежливого отношения к ней, сколько простая практика обцененной лексики иностранного для информанта языка.

*мощницы скорее были более такие человеческие чувства. (Информант № 15, М, Тель-Авив)*

Главная причина, по которой пользователи очеловечивают умную колонку, — ее способность воспроизводить живую речь. Здесь играет роль не столько характер коммуникации, сколько в принципе возможность ведения вербального диалога.

*Ну, наверное, да, она вот более живая. Более живая из всего того, что сейчас есть... потому что с ней можно поговорить... Вербальное общение, да. (Информант № 1, Ж, Санкт-Петербург)*

*То есть это все-таки я... чуть ближе отношу к живому предмету... Потому что она, во-первых, отвечает, во-вторых, ну, блин, она отвечает. Мой мозг воспринимает, что, если тебе кто-то отвечает, скорее всего, это кто-то живой. Примерно живой... Я не буду как-то агрессивно. (Информант № 13, М, Тель-Авив)*

Можно отметить, что общий характер взаимодействия и коммуникации с умной колонкой не оказывает влияние на очеловечивание. Люди по природе своей склонны приписывать человеческие черты всему окружающему, а в данном случае мы имеем дело с технологией, способной воспроизводить речь и вступать в диалог. Скорее всего, именно это ее свойство побуждает пользователей к очеловечиванию, а роль характера использования оказывается не столь важна. Материалы исследования подтвердили контртезис 2.

### *Вежливые фразы*

Смысл одного из наших предположений в рамках тезиса 2 состоял в том, что пользователи, не склонные очеловечивать умную колонку, при коммуникации с ней прибегают к приемам лингвистической вежливости. Имеется в виду, что колонка воспринимается такими людьми в качестве инструментального помощника. В ходе интервью информанты нередко упоминали, что используют вежливые слова в процессе коммуникации с колонкой, и причина этого действительно не всегда связана с очеловечиванием:

*Да, стараюсь быть вежливым, но, опять же, «Алиса, доброе утро» — это не только вежливость, а все-таки сценарий для запуска. (Информант № 2, М, Санкт-Петербург)*

Отметим, что добавление вежливых слов для запуска сценариев или окончания диалога — необязательная часть коммуникации, однако пользователи употребляют различные вежливые формы обращения с умной колонкой:

*Чтобы закончить диалог, я обычно говорю «спасибо», тогда она отключается. (Информант № 4, М, Санкт-Петербург)*

Лингвистическая вежливость действительно сопутствует инструментальному использованию умной колонки. Добавление вежливых слов или выражений — это

не столько признак очеловечивания, сколько часть привычных сценариев и характеристика коммуникации с умной колонкой.

### Поддержание лица

Мы также предполагали, что те пользователи, кто воспринимает умную колонку как социального партнера, сохраняют лицо (свое и собеседника) в процессе коммуникации. Рассмотрим этот тезис на примере возможных обид и извинений перед умной колонкой, так как полагаем, что извинение является одним из ключевых инструментов восстановления как лица самого пользователя, так и «лица» умной колонки. Один из информантов отметил, что старается подбирать слова в процессе коммуникации, чтобы не обидеть умную колонку:

*Я как бы, как минимум, если я боюсь, не боюсь, но если я пытаюсь не обидеть, если я пытаюсь подбирать слова, если я говорю «спасибо», ну, во-первых, мне это приятно говорить, во-вторых, если я так говорю, значит, это что-то как-то влияет [на характер коммуникации с умной колонкой]. (Информант № 13, М, Тель-Авив)*

Информанты отмечают, что стремятся сохранить свое лицо в процессе коммуникации с умной колонкой. Здесь также присутствует взаимосвязь с отношением к людям в реальной жизни:

*Ну, это, возможно, как сказать, какая-то некая совесть, да, то есть не совсем осознанно. То есть вроде не человек — действительно, да, там, голосовой помощник, но все равно как бы, ну, тут опять же, скорее, из-за того, что обычно я же так с другими людьми не разговариваю, а если я разговариваю, то извинюсь, если я был груб, ну, тут, соответственно, срабатывает такой же механизм. (Информант № 6, М, Санкт-Петербург)*

Один из информантов отметил, что старается вести мягкую и ровную коммуникацию с колонкой в процессе «бесед» с ней:

*Если я какой-то запрос хочу сделать, если это что-то, если это простой разговор, ну, типа, если я продолжаю беседу после какого-то вопроса с ней, я могу использовать более мягкую речь, слова. (Информант № 13, М, Тель-Авив)*

Можно отметить, что такой характер коммуникации способствует поддержанию как лица самого пользователя, так и «лица» умной колонки.

Что касается возможных «обид» умной колонки на невежливое обращение с ней, то одна из пользовательниц отметила, что, по ее мнению, умная колонка действительно способна обижаться. По этой причине она извиняется перед ней:

*Если она будет говорить, что она там обиделась, в принципе, можно извиниться, и я думаю, она простит. (Информант № 16, Ж, Тель-Авив).*

Однако пути восстановления «лица» колонки не всегда строятся через извинения перед ней:

*Ну то есть у меня есть те моменты, когда мне что-то все еще нужно после того, как я на нее наорал. Я могу типа сделать паузу, подождать, пока она выключится, включится обратно. И я такой еще раз: «Пожалуйста, Сири...» (Информант № 14, М, Тель-Авив)*

Здесь мы наблюдаем ситуацию, при которой после не слишком вежливого взаимодействия с колонкой пользователю достаточно просто дождаться перезапуска колонки, чтобы продолжить диалог. То есть нет необходимости проводить работу по восстановлению «лица» умной колонки путем извинений. Это связано с тем, что колонка — это технология, не способная запоминать или держать обиду. Иногда даже не нужно дожидаться перезапуска и, соответственно, нет необходимости восстанавливать «лицо» колонки:

*То есть вот идет оскорбление. Все, она поняла, идет оскорбление, даже если я понял. Но она будет слушать до конца, потому что я могу оскорблять, оскорблять, оскорблять, а в конце спросить, какая погода. (Информант № 14, М, Тель-Авив)*

То есть после разрушения «лица» колонки путем ее оскорбления пользователь продолжает взаимодействие на инструментальном уровне с помощью вопроса о погоде. Возможно, здесь уже нет места партнерству и социальному взаимодействию с умной колонкой.

Поддержание «лица» умной колонки в процессе коммуникации тесно связано с очеловечиванием этой технологии. Об этом говорит в интервью одна из информанток:

*Из-за того, что имитируется разговор с колонкой как разговор с человеком, подразумевается, что она такой своеобразный Искусственный Интеллект, можно сказать, и нужно тоже к ней уважительно относиться, потому что если ты будешь с ней, например, неуважительно разговаривать, то она тоже отреагирует. Она тоже скажет, что «не надо меня там обзывать». (Информант № 16, Ж, Тель-Авив)*

Еще один интересный пример касается взаимодействия с колонкой от компании «Амазон» (Amazon) Алексой. Наша информантка прямо в процессе интервью продемонстрировала диалог, содержащий последовательно оскорбление и извинение:

*Alexa, du bist dumm [нем.— ты глупая]. А, вот, видишь, я ей сказала: «Ты тупая, ты глупая», и она сделала звук, как знаешь, contact failed [англ.— контакт не удался]. Да, она пикнула, типа: «Я не буду на это отвечать». Alexa, entschuldigung bitte [нем.— Алекса, извини, пожалуйста]. О, вот смотри, я сейчас ей сказала: «Извини, пожалуйста», и она сказала: «Окей, не проблема» — «Es schon gut» [нем.— все хорошо]. (Информант № 7, Ж, Берлин)*

Этот пример ярко демонстрирует, что, во-первых, колонка способна реагировать на оскорбительное отношение к ней, а во-вторых, что пользователю достаточно произнести слова извинения для того, чтобы восстановить «лицо» колонки.

Таким образом, исследование подтвердило, что пользователи стараются подерживать свое лицо и «лицо» колонки в процессе коммуникации с ней. При этом работа лица и формирование уважительного отношения к колонке связаны с очеловечиванием. Однако не все пользователи стремятся сохранить или восстановить «лицо» колонки, так как понимают, что все-таки это технология.

## Заключение

Умная колонка — новый объект домашнего пространства, особенности которого позволяют пользователям взаимодействовать с колонкой в повседневной жизни, а также формировать новые типы взаимоотношений с ней. Настоящее исследование представляет собой попытку концептуализации одного из аспектов взаимодействия с умной колонкой, связанного с проявлением различных форм вежливости в коммуникации с ней. Работа была организована вокруг анализа двух групп тезисов и контртезисов.

Первая группа характеризовала вежливость пользователей во взаимодействии с умными колонками. По результатам исследования нашли свое подтверждение *тезис 1* (существуют определенные нормы вежливости при коммуникации с одомашненной умной колонкой) и *контртезис 1.1* (нормы вежливости универсальны). Сравнение интервью с информантами из трех стран показало, что нормы вежливости, характерные для пользователей при коммуникации с умными колонками, оказываются сходными в разных странах и представляют собой универсальный «минимум» вежливости, который люди проявляют по отношению друг к другу. Вместе с тем были выявлены некоторые различия между информантами из разных стран, связанные скорее со свойствами языка, на котором ведется разговор. Следует также отметить, что хотя в прошлых исследованиях [Burton, Gaskin 2019] не было выявлено связи в проявлении вежливости к электронным помощникам и к людям в обычной жизни, результаты нашего исследования свидетельствуют об обратном. Пользователи умных колонок склонны переносить нормы вежливости из социальных взаимодействий на коммуникацию с технологиями.

Вторая группа описывала взаимосвязь между очеловечиванием умной колонки и характером вежливости во взаимодействии с ней. Результаты исследования привели к подтверждению *контртезиса 2* (не существует зависимости между тем, как люди используют колонку, и тем, какой концепции вежливости они придерживаются). Очеловечивание умной колонки связано не с характером ее использования (как инструмента для решения задач или как социального партнера), а с тем, что умная колонка может вступать в вербальную коммуникацию с пользователем, имитируя человеческую речь. Именно эта универсальная способность технологии побуждает пользователей в той или иной мере очеловечивать ее. Во взаимодействии с умной колонкой совмещаются элементы лингвистической вежливости (вежливые фразы) и вежливость как поддержание лица — своего и собеседника. При этом оказывается, что возможен простой переход от социального взаимодействия с умной колонкой к чисто инструментальному, когда для сохранения «лица» умной колонки можно просто «перезапустить» ее, что невозможно во взаимодействии с человеком.

Проведенное исследование характеризуется определенными ограничениями. Первое связано с качественным характером данных: несмотря на то что анализ интервью выявил определенные устойчивые тенденции в «вежливом» взаимо-

действии с умными колонками, мы не можем знать, в какой мере и для кого эти тенденции типичны. Второе ограничение определяется тем, что характер взаимодействия с умными колонками выявлялся в ходе интервью, которое позволяет выделить повседневные практики и мотивы информантов, но, в отличие от наблюдения, не дает доступа к деталям организации их повседневной жизни.

Несмотря на существующие ограничения, по результатам проведенного анализа мы можем сделать следующие общие выводы.

1. Вежливое отношение к одомашненной умной колонке — одна из главных черт использования этой технологии ИИ. Проявления вежливости исходят от самих пользователей, при этом они не изобретают новых форм вежливого обращения. Здесь наблюдается ориентация на привычные социальные взаимодействия. Таким образом, умная колонка — не просто объект одомашнивания, она выступает как *социальный объект*, который становится частью повседневной жизни людей.

2. Коммуникация с умной колонкой сопровождается параллельным обучением на основании ежедневных взаимодействий с большим числом пользователей. В этом случае проявление различных форм вежливости при коммуникации с умной колонкой самими пользователями способствует формированию «навыков» коммуникации умной колонки. Благодаря различным сценариям взаимодействия колонка учится различать вежливое и грубое отношение к ней, а также воспринимать различные лингвистические конструкции и реагировать на них. Таким образом, путем построения вежливой или невежливой коммуникации, пользователи коллективно конструируют «лицо» умной колонки. Чем больше и разнообразнее пользователи взаимодействуют с умной колонкой, тем больше у нее возможностей научиться выстраивать различные линии поведения с пользователями. Это взаимный процесс, который не прекращается до тех пор, пока продолжается взаимодействие множества пользователей с их умными колонками. Для разных технологий (Алиса, Сири, Алекса и др.) это «лицо» будет разным, если пользователи взаимодействуют с ними по-разному.

3. Хотя умная колонка — это технология, используемая многими преимущественно в инструментальных целях, она заставляет людей задуматься о том, кто они такие и как они должны с ней обращаться (то есть является эвокативным объектом в терминах Ш. Теркл). Способность умной колонки вступать с разговор с пользователями обуславливает их склонность к очеловечиванию данной технологии и, следовательно, желание строить с ней более «живую» коммуникацию. Умная колонка позволяет людям ощутить непосредственное присутствие искусственного интеллекта в повседневной жизни, тем самым трансформируя их представления о реальности.

4. Появление и активное использование умной колонки в домашнем пространстве повлияли на конструирование новых форм социальной реальности. Домашнее пространство в повседневности — это уже не что-то статичное и опривыченное, а динамичное и гибкое. Умная колонка создает новый фон домашнего пространства, сравнимый, по словам одного из информантов, с картой глубины<sup>6</sup>.

<sup>6</sup> «А Алиса, она дает некоторые технологические возможности, она, допустим, дает оранжевый цвет, а не красный. Но это распределено по воздуху, по радиусу действия. И очень круто и удобно, потому что тебе не нужно физически двигаться к какому-то из красных точек, какие-то вопросы, которые тебе не требуют вмешательства настолько серьезные, она как бы способна ответить. Поэтому она как бы расширяет твой горизонт использования технологий» (Информант № 13, М, Тель-Авив).

Подводя итоги, следует обратить внимание на то, что умная колонка является социальным объектом, способным к вербальной коммуникации с пользователем и обладающим навыками обучаемости. Вежливое отношение к умной колонке как к новой технологии, чья роль в повседневной жизни людей еще не до конца понятна, отражает не столько специфику культуры или индивидуальные качества того или иного пользователя, сколько общее состояние развития культуры вежливости в обществе.

Мы находимся на таком этапе развития и трансформации общества, на котором происходят фундаментальные изменения социального порядка, связанные с прочным вхождением технологий ИИ в различные сферы общественной жизни. Задача исследователей — наблюдать за тем, как будут развиваться события, выявлять новые тенденции и типы социальных взаимодействий, а также разрабатывать новые теории и концепции, направленные на объяснение происходящих в обществе процессов.

Данное исследование мы рассматриваем как поисковое. На основании анализа специальной литературы были сформулированы и затем, на ограниченном эмпирическом материале, проверены тезисы и контртезисы, выделены тенденции формирования вежливого взаимодействия пользователей с умными колонками. Полученные выводы могут служить основанием для организации новых — как количественных, так и качественных — исследований. Укажем лишь два направления подобных исследований.

Во-первых, в процессе исследования выявлено, что умные колонки способны реагировать на вежливое/невежливое отношение людей к ним. Мы полагаем, что вопрос реакции умных колонок и изменение их «поведения» в зависимости от характера коммуникации с ними важны для понимания динамики взаимодействия пользователей с данными технологиями.

Во-вторых, особый интерес представляет то, как в существующие умные колонки может быть «встроен» генеративный искусственный интеллект нового поколения, основанный на больших языковых моделях. Мы надеемся обратиться к данным проблемам в дальнейших исследованиях.

## Список литературы (References)

1. Корбут А. М. Одомашнивание искусственного интеллекта: умные колонки и трансформация повседневной жизни // Мониторинг общественного мнения: экономические и социальные перемены. 2021. № 1. С. 193—216. <https://doi.org/10.14515/monitoring.2021.1.1808>.  
Korbut A. M. (2021) Domestication of Artificial Intelligence: Smart Speakers and Transformation of Everyday Life. *Monitoring of Public Opinion: Economic and Social Changes*. No. 1. P. 193—216. <https://doi.org/10.14515/monitoring.2021.1.1808>. (In Russ.)
2. Ушкин С. Г., Коваль Е. А. Алиса, ты следишь за мной? Восприятие конфиденциальности в нарративах пользователей «умных» колонок // Мониторинг общественного мнения: экономические и социальные перемены. 2023. № 3. С. 23—40. <https://doi.org/10.14515/monitoring.2023.3.2314>.

- Ushkin S. G., Koval E. A. (2023) Alice, Are You Watching Me? Perception of Privacy in Narratives of “Smart” Speaker Users. *Monitoring of Public Opinion: Economic and Social Changes*. No. 3. P. 23—40. <https://doi.org/10.14515/monitoring.2023.3.2314>. (In Russ.)
3. Ушкин С. Г., Коваль Е. А., Яськин А. Н. Жить с Алисой: как голосовые помощники трансформируют практики коммуникации? // Журнал исследований социальной политики. 2022. Т. 20. № 3. С. 361—376. <https://doi.org/10.17323/727-0634-2022-20-3-361-376>.  
Ushkin S. G., Koval E. A., Yaskin A. N. (2022) Living with Alice: How Voice Assistants Transform Communication Practices? *Journal of Social Policy Research*. Vol. 20. No. 3. P. 361—376. <https://doi.org/10.17323/727-0634-2022-20-3-361-376>. (In Russ.)
  4. Brause S. R., Blank G. (2020) Externalized Domestication: Smart Speaker Assistants, Networks, and Domestication Theory. *Information, Communication & Society*. Vol. 23. No. 5. P. 751—763. <http://dx.doi.org/10.1080/1369118X.2020.1713845>.
  5. Burton, N. G., Gaskin J. (2019) “Thank You, Siri”: Politeness and Intelligent Digital Assistants. In: *Proceedings of the 25th Americas Conference on Information Systems. Cancún, Mexico, August 15—17*. P. 2860—2870.
  6. Esposito E. Artificial communication? (2017) The Production of Contingency by Algorithms. *Zeitschrift für Soziologie*. Vol. 46. No. 4. P. 249—265. <http://dx.doi.org/10.1515/zfsoz-2017-1014>.
  7. Goffman E. (1955) On Face-Work: An Analysis of Ritual Elements in Social Interaction. *Psychiatry*. Vol. 18. No. 3. P. 213—231. <https://doi.org/10.1080/00332747.1955.11023008>.
  8. Kang H., Oh J. (2023) Communication Privacy Management for Smart Speaker Use: Integrating the Role of Privacy Self-Efficacy and the Multidimensional View. *New Media & Society*. Vol. 25. No. 5. P. 1153—1175. <http://dx.doi.org/10.1177/146144482111026611>.
  9. Kucharski A. B., Merkel S. (2023) Domestication of Smart Speakers by Older Users: Preliminary Findings from an Exploratory Qualitative Study. *Digital Culture & Society*. Vol. 9. No. 1. P. 203—212. <https://doi.org/10.14361/dcs-2023-0110>.
  10. Li M., Suh A. (2021) Machinelike or Humanlike? A Literature Review of Anthropomorphism in AI-Enabled Technology. In: T. X. Bui (ed.) *Proceedings of the 54th Hawaii International Conference on System Sciences (HICSS)*, Grand Wailea, Maui, Hawaii, 2021. P. 4053—4062. <http://dx.doi.org/10.24251/HICSS.2021.493>.
  11. Luhmann N. (1992) What is Communication? *Communication Theory*. Vol. 2. No. 3. P. 251—259. <https://doi.org/10.1111/j.1468-2885.1992.tb00042.x>.
  12. Malkin N., Deatrack J., Tong A., Wijesekera P., Egelman S., Wagner D. (2019) Privacy Attitudes of Smart Speaker Users. *Proceedings on Privacy Enhancing Technologies*. No. 4. P. 250—271. <http://dx.doi.org/10.2478/popets-2019-0068>.

13. Maynard D. W., Clayman S. E. (1991) The Diversity of Ethnomethodology. *Annual Review of Sociology*. Vol. 17. No. 1. P. 385—418. <http://dx.doi.org/10.1146/annurev.soc.17.1.385>.
14. Moussawi S., Koufaris M., Benbunan-Fich R. (2021) How Perceptions of Intelligence and Anthropomorphism affect Adoption of Personal Intelligent Agents. *Electronic Markets*. Vol. 31. P. 343—364. <https://doi.org/10.1007/s12525-020-00411-w>.
15. Nwoye O. G. (1992) Linguistic Politeness and Socio-Cultural Variations of the Notion of Face. *Journal of Pragmatics*. Vol. 18. No. 4. P. 309—328. <https://doi.org/10.1016/0378-2166%2892%2990092-P>.
16. Ribino P. (2023) The Role of Politeness in Human—Machine Interactions: A Systematic Literature Review and Future Perspectives. *Artificial Intelligence Review*. Vol. 56. P. 445—482. <https://doi.org/10.1007/s10462-023-10540-1>.
17. Salles A., Evers K., Farisco M. (2020) Anthropomorphism in AI. *AJOB Neuroscience*. Vol. 11. No. 2. P. 88—95. <http://dx.doi.org/10.1080/21507740.2020.1740350>.
18. Smith E., Sumner P., Hedge C., Powell G. (2021) Smart Speaker Devices Can Improve Speech Intelligibility in Adults with Intellectual Disability. *International Journal of Language & Communication Disorders*. Vol. 56. No. 3. P. 583—593. <https://doi.org/10.1111/1460-6984.12615>.
19. Sztompka P. (2008) The Focus on Everyday Life: A New Turn in Sociology. *European Review*. Vol. 16. No. 1. P. 23—37. <https://doi.org/10.1017/S1062798708000045>.
20. Turkle S. (2005) *The Second Self: Computers and the Human Spirit*. Cambridge: MIT Press. <http://dx.doi.org/10.7551/mitpress/6115.001.0001>.
21. Watts R. J. (2003) *Politeness*. Cambridge: Cambridge University Press. <https://doi.org/10.1017/CB09780511615184>.

**Приложение 1.****Таблица 1. Характеристика информантов**

№	Возраст	Пол	Город	Род занятий	Проживание
1	27	Ж	Санкт-Петербург	Тренер по гимнастике	С мужем и собакой
2	27	М	Санкт-Петербург	Проектный менеджер	С девушкой
3	28	М	Санкт-Петербург	Бизнесмен	Один
4	27	М	Санкт-Петербург	IT специалист	Один
5	28	Ж	Санкт-Петербург	Логист	С матерью
6	30	М	Санкт-Петербург	Адвокат	С девушкой
7	29	Ж	Берлин	Консультант	С мужем
8	28	М	Берлин	Налоговый консультант	С женой
9	30	М	Берлин	Логист	С женой, двумя детьми и двумя кошками
10	30	Ж	Берлин	Домохозяйка	С мужем, двумя детьми и двумя кошками
11	30	Ж	Берлин	Домохозяйка	С мужем, дочерью и собакой
12	30	Ж	Берлин	Адвокат	С мужем и дочерью
13	25	М	Тель-Авив	IT специалист	С девушкой
14	26	М	Тель-Авив	Студент, IT специалист	Один
15	25	М	Тель-Авив	Бармен	С женой
16	24	Ж	Тель-Авив	Студентка	С мужем
17	25	М	Тель-Авив	IT специалист	С девушкой
18	43	Ж	Тель-Авив	IT специалист	С мужем и двумя дочерьми

## Приложение 2. Список категорий и подкатегорий

1. Одомашнивание.
  - 1.1 Появление умной колонки дома (коммодификация).
  - 1.2 Расположение умной колонки: локация и причина.
  - 1.3 Перемещение умной колонки.
  - 1.4 Ожидания от пользования умной колонкой (объективизация).
  - 1.5 Частота обращений к умной колонке.
  - 1.6 Динамика использования.
2. Практики.
  - 2.1 Трансформация привычных практик.
  - 2.2 Новые практики.
3. Используемые функции.
  - 3.1 Используемые функции: список.
  - 3.2 Используемые функции: примеры.
  - 3.3 Используемые функции: повседневность.
4. Культура и язык.
5. Образ умной колонки.
6. Влияние умной колонки на отношение к другим людям.
7. Влияние умной колонки на отношение к другим технологиям.
8. Очеловечивание.
  - 8.1 Очеловечивание умной колонки.
  - 8.2 Очеловечивание: предметы и другие технологии.
9. Эвокативный объект.
10. Социальные взаимодействия.
  - 10.1 Социальные взаимодействия: включение.
  - 10.2 Социальные взаимодействия: коллективное.
  - 10.3 Социальные взаимодействия: случайные.
  - 10.4 Социальные взаимодействия: замена.
11. Коммуникация с умной колонкой.
  - 11.1 Коммуникация: вопросы.
  - 11.2 Коммуникация: молчание колонки.
  - 11.3 Коммуникация с умной колонкой: непонимание.
  - 11.4 Коммуникация с умной колонкой: ограничения.
  - 11.5 Коммуникация с умной колонкой: прекращение диалога.
  - 11.6 Коммуникация с умной колонкой: характер.

## 12. Вежливость.

- 12.1 Вежливость с умной колонкой.
- 12.2 Вежливость: слова и выражения.
- 12.3 Вежливость: от умной колонки.
- 12.4 Вежливость: умная колонка и люди.
- 12.5 Вежливость: обида.
- 12.6 Вежливость: извинения.

## 13. Благодарность.

## 14. Сравнение разных умных колонок.

## 15. Рекомендации разработчикам.

## 16. Умная колонка и чат GPT.

## 17. Умная колонка и изменения в жизни.

DOI: [10.14515/monitoring.2024.5.2626](https://doi.org/10.14515/monitoring.2024.5.2626)

**О. А. Бурова, Ю. С. Неверова, Г. И. Селеткова, Е. В. Середкина**

## **ВЗАИМОДЕЙСТВИЕ ПОЛЬЗОВАТЕЛЕЙ РАЗНОГО ВОЗРАСТА С СЕРВИСНЫМ РОБОТОМ: ЛАБОРАТОРНОЕ ИССЛЕДОВАНИЕ**

**Правильная ссылка на статью:**

Бурова О. А., Неверова Ю. С., Селеткова Г. И., Середкина Е. В. Взаимодействие пользователей разного возраста с сервисным роботом: лабораторное исследование // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 166—182. <https://doi.org/10.14515/monitoring.2024.5.2626>.

**For citation:**

Burova O. A., Neverova Y. S., Seletkova G. I., Seredkina E. V. (2024) Interaction of Users of Different Ages with a Service Robot: A Laboratory Study. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 166–182. <https://doi.org/10.14515/monitoring.2024.5.2626>. (In Russ.)

Получено: 10.05.2024. Принято к публикации: 10.09.2024.

**ВЗАИМОДЕЙСТВИЕ ПОЛЬЗОВАТЕЛЕЙ РАЗНОГО ВОЗРАСТА С СЕРВИСНЫМ РОБОТОМ: ЛАБОРАТОРНОЕ ИССЛЕДОВАНИЕ**

*БУРОВА Ольга Аркадьевна — старший преподаватель кафедры социологии и политологии, Пермский национальный исследовательский политехнический университет, Пермь, Россия*  
E-MAIL: [olgabu@mail.ru](mailto:olgabu@mail.ru)  
<https://orcid.org/0000-0002-6182-4375>

*НЕВЕРОВА Юлия Сергеевна — старший преподаватель кафедры социологии и политологии, Пермский национальный исследовательский политехнический университет, Пермь, Россия*  
E-MAIL: [ny-sity@mail.ru](mailto:ny-sity@mail.ru)  
<https://orcid.org/0000-0003-2185-7903>

*СЕЛЕТКОВА Гюзель Ильясовна — старший преподаватель кафедры социологии и политологии, Пермский национальный исследовательский политехнический университет, Пермь, Россия*  
E-MAIL: [guzal.ka@mail.ru](mailto:guzal.ka@mail.ru)  
<https://orcid.org/0000-0003-3402-3473>

*СЕРЕДКИНА Елена Владимировна — кандидат философских наук, доцент, директор научно-исследовательского центра «Центр оценки технологий», Пермский национальный исследовательский политехнический университет, Пермь, Россия*  
E-MAIL: [selena36@mail.ru](mailto:selena36@mail.ru)  
<https://orcid.org/0000-0003-2506-2374>

**Аннотация.** Статья посвящена изучению процесса человеко-машинного взаимодействия на примере коммуникации пользователей с сервисным социальным роботом. В теоретико-методологическом плане исследование имеет полипарадигмальную основу: субъективный опыт пользователей и их адаптация к конкретной ситуации

**INTERACTION OF USERS OF DIFFERENT AGES WITH A SERVICE ROBOT: A LABORATORY STUDY**

*Olga A. BUROVA<sup>1</sup> — Senior Lecturer at the Department of Sociology and Political Science*  
E-MAIL: [olgabu@mail.ru](mailto:olgabu@mail.ru)  
<https://orcid.org/0000-0002-6182-4375>

*Yulia S. NEVEROVA<sup>1</sup> — Senior Lecturer at the Department of Sociology and Political Science*  
E-MAIL: [ny-sity@mail.ru](mailto:ny-sity@mail.ru)  
<https://orcid.org/0000-0003-2185-7903>

*Guzel I. SELETKOVA<sup>1</sup> — Senior Lecturer at the Department of Sociology and Political Science*  
E-MAIL: [guzal.ka@mail.ru](mailto:guzal.ka@mail.ru)  
<https://orcid.org/0000-0003-3402-3473>

*Elena V. SEREDKINA<sup>1</sup> — Cand. Sci. (Philos.), Associate Professor, Director of the Research Center “Center for Technology Assessment”*  
E-MAIL: [selena36@mail.ru](mailto:selena36@mail.ru)  
<https://orcid.org/0000-0003-2506-2374>

<sup>1</sup> Perm National Research Polytechnic University, Perm, Russia

**Abstract.** The article is devoted to studying human-robot interaction (HRI) using the example of user interaction with a service social robot. The study has a multi-paradigm basis in theoretical and methodological terms: the subjective users' experience and their adaptation to a particular interaction situation with the robot are considered within social constructiv-

взаимодействия с роботом рассматриваются в рамках социального конструктивизма, феноменологии и этнометодологии. Эмпирической базой служит лабораторное исследование с использованием методов социального эксперимента, наблюдения, фокус-групп. Участники эксперимента, имитируя получение стандартной услуги по оформлению документов, обращались к агентам разных типов — человеку и сервисному роботу в разных режимах работы.

По результатам исследования выделены особенности взаимодействия с сервисным роботом, а также определена специфика субъективного восприятия этого опыта пользователями, относящимися к разным возрастным группам. В ходе работы подтвердилась гипотеза о том, что молодые пользователи легче адаптируются к ситуации взаимодействия с роботом. По оценкам участников исследования, представителям среднего и старшего возраста при получении услуги проще общаться с человеком-консультантом, а молодым участникам — с роботом (независимо от режима его работы). Также в статье выделены различия в ожиданиях и оценках относительно роботов на уровне общих социальных установок. По результатам фокус-групп, готовность к взаимодействию с роботами снижена по причинам недоверия их функциональным возможностям, неуверенности в собственной компетентности и боязни неопределенности (настороженного отношения ко всему новому).

**Ключевые слова:** взаимодействие человека и робота (HRI), сервисный робот, социальный робот, социальные установки к новым технологиям, агенты коммуникации, адаптация к взаимодействию с роботом, отношение к роботам, изучение пользователей

ism, phenomenology, and ethnomethodology. The empirical basis is laboratory research using social experiments, observation, and focus groups. The experiment participants interacted with different types of agents (a human and a service robot in different modes of operation), simulating receiving a standard state service for the execution of documents.

According to the study results, the features of the actual interaction with the service robot, as well as the subjective perception of this experience by users of different age groups, are highlighted. The study confirms the hypothesis that young users adapt more quickly to interaction with a robot, which is expressed in the speed of interaction and subjective satisfaction with this interaction. According to the estimates of the middle-aged and older study participants, it is easier for them to receive services interacting with a human consultant and, according to the estimates of young participants, with a robot (regardless of its mode of operation). Differences in expectations and assessments regarding robots at the level of general social attitudes are also highlighted. According to the results of the focus groups, the willingness to interact with robots is reduced due to distrust of the functional capabilities of the robot, lack of confidence in their competence, and fear of uncertainty (a wary attitude towards everything new).

**Keywords:** human-robot interaction (HRI), service robot, social robot, social attitudes towards new technologies, communication agents, adaptation to interaction with a robot, attitude towards robots, user study

**Благодарность.** Авторы выражают благодарность за содействие в проведении исследования сотрудникам кафедры социологии и политологии ПНИПУ Ольге Александровне Ганиной и Юрию Михайловичу Вассерману, а также ООО «Промобот» (г. Пермь).

**Acknowledgments.** The authors extend their sincere gratitude to Olga Alexandrovna Ganina and Yuri Mikhailovich Wasserman, esteemed members of the Department of Sociology and Political Science at PNRPU, for their invaluable assistance in conducting this study. We also express our appreciation to Promobot LLC (Perm) for their support and collaboration, which was instrumental in the successful execution of this research.

## Введение

В мире растет использование социальных роботов в различных сферах: здравоохранении, образовании, предоставлении государственных услуг. Социальный робот понимается нами как физически воплощенный искусственный агент, имитирующий поведение, внешний вид или движения человека, который может восприниматься индивидами как социальная сущность [Naneva et al., 2020], взаимодействует с людьми через социальный интерфейс [Hegel et al., 2009] и передает вербальную и/или невербальную информацию [Broekens, Heerink, Rosendal, 2009]. Отношение к роботам влияет на эффективность взаимодействия человека с ними. Понимание того, как разные социальные группы, в частности выделяемые по возрасту, относятся к роботам, будет способствовать внедрению робототехники в различных сферах общественной и частной жизни с учетом специфики данных групп.

Анализ данных эмпирических исследований показывает некоторые различия в результатах относительно влияния возраста человека на отношение к взаимодействию с роботом и на сам процесс такого взаимодействия.

В рамках своего обзора С. Нанева и соавторы [Naneva et al., 2020] показали, что с 2005 по 2019 г. было опубликовано как минимум 97 статей, основанных на исследованиях, посвященных изучению отношения человека к роботу. Авторы обзора выделили и описали работы, направленные на изучение аффективных и когнитивных установок, общих оценок (сочетание аффективных и когнитивных установок), принятия (готовности к использованию, использование), уровня тревоги и доверия к роботам. Исследователи отметили, что возраст не связан с перечисленными показателями.

Т. Гнамбс и М. Аппель [Gnambs, Appel, 2019] на основе данных Евробарометра за 2012, 2014, 2017 гг., применяя многоуровневый анализ для изучения отношения к использованию роботов, обнаружили, что возраст оказывает незначительное влияние на их оценку. В частности, мужчины с высшим образованием позитивнее относятся к социальным роботам по сравнению с женщинами.

Дж. Хадсон и соавторы [Hudson, Orviska, Hunady, 2017], анализируя отношение людей к роботам, которые применяются для ухода за пожилыми людьми, на основе данных Евробарометра за ноябрь/декабрь 2014 г., пришли к следующим выводам о влиянии ряда факторов на отношение к роботам:

— возраст: пожилые люди негативнее относятся к использованию роботов, нежели молодежь;

— уровень урбанизации: жители крупных городов демонстрируют более уверенное, позитивное отношение к роботам, чем жители деревень и небольших городов.

С. Мандл и соавторы [Mandl et al., 2022] выделили при изучении отношения к роботам три компонента: моральность/коммуникабельность, активность/сотрудничество и антропоморфизм. Они обнаружили значимые корреляции между возрастом, полом, уровнем образования и выделенными компонентами шкалы.

Е. А. Гаврилина [Гаврилина, 2023], рассматривая взаимодействие человека и социального робота в России через призму объектно-ориентированного и акторно-сетевого подходов, отмечает влияние ожиданий человека от взаимодействия с роботом на сам процесс взаимодействия. По результатам исследования Р. Абрамова и В. Катечкиной [Абрамов, Катечкина, 2022], на восприятие людьми взаимодействия человека и робота оказывают влияние социокультурные факторы.

Таким образом, в исследованиях получены противоречивые данные по поводу влияния возраста реальных и потенциальных пользователей на отношение к роботам, что может быть связано с действием комбинации факторов, таких как контекст взаимодействия, характеристики робота и личностные особенности участников исследований.

Для изучения влияния возраста пользователей на процесс взаимодействия с конкретным сервисным роботом в заданном контексте было проведено региональное исследование. Основная гипотеза исследования состояла в следующем (гипотеза 1): *молодые пользователи, в силу большей готовности к новым технологиям, легче, чем пользователи старшего возраста, адаптируются к ситуации взаимодействия с роботом независимо от характеристик робота.*

Важность включения контекста в изучение отношения человека к роботам подчеркивалась многими исследователями [Katz, Halpern, Crocer, 2015]. Н. Н. Зильберман и М. А. Стефанцова отмечают, что главной характеристикой социального робота является способность выстраивать взаимодействие с человеком с «использованием сложившихся семиотических систем и учетом социального и ситуативного контекста» [Зильберман, Стефанцова, 2016: 297].

Л. Сачмен в книге «Реконфигурации отношений человек — машина: планы и ситуативные действия» на основе феноменологии, этнометодологии и конверсационного анализа подчеркивает важность социального контекста взаимодействия [Сачмен, 2019: 8]. Условия, в которых действуют индивиды, не установлены заранее и не очевидны, они формируются и определяются в процессе самого действия. Взаимодействие людей предполагает взаимное придание смыслов, конструирование совместного объяснения (взаимо)действия, тем самым каждый участник вносит свой вклад в формирование общего опыта и понимания [там же]. Исходя из этих теоретических предпосылок, можно отметить ограниченность роботов, асимметрию между человеком и машиной в человеко-машинном взаимодействии. При этом разработчики роботизированных систем используют определенные характеристики общения людей, что способствует «приписыванию машинам интерактивности их собеседниками-людьми» [там же: 34], «очеловечению» роботов в глазах пользователей. В свою очередь, приписывание человеческих ка-

честв роботам создает у пользователей искаженное представление о способности роботов к взаимодействию, пониманию контекста и формированию смыслов, что может осложнять коммуникацию, приводить к сбоям в понимании (и интерпретации), которые в обычном человеческом взаимодействии легко исправляются, но становятся критическими в коммуникации между человеком и машиной.

Концепции конфигурации и реконфигурации Л. Сачмен, а также ее подход к анализу фигураций, понимаемых как способ представления и организации опыта, могут быть полезны для изучения человеко-машинного взаимодействия. Конфигурация в контексте человеко-машинного взаимодействия относится к тому, как организованы элементы взаимодействия между человеком и машиной, включающие в себя графические интерфейсы, элементы управления (кнопки, меню, иконки), функционал, пользовательский опыт (как пользователь оценивает интуитивность и удобство работы с системой). Реконфигурация означает изменения в этой организации, происходящие в ответ на обратную связь от пользователей или изменения в контексте использования. Например, адаптация системы и ее персонализация. Подход Л. Сачмен к анализу фигураций подчеркивает, что любой язык, включая технический и математический, состоит из «тропов», или «оборотов речи», которые вызывают ассоциации с различными областями смысла. Это важно для человеко-машинного взаимодействия по нескольким причинам. Во-первых, взаимодействие с машиной часто происходит через язык интерфейса. Например, названия кнопок и команд могут иметь разные значения для разных пользователей в зависимости от их культурного и профессионального контекста. Анализируя эти языковые конструкции, мы можем лучше понять, как пользователи воспринимают интерфейс и какие ассоциации он вызывает. Фигурации показывают, что взаимодействие — нестатичный процесс, оно постоянно изменяется в зависимости от контекста, обратной связи и эволюции технологий, что соответствует концепции реконфигурации, где системы должны адаптироваться к меняющимся условиям. Таким образом, концепции конфигурации и реконфигурации, а также анализ фигураций помогают глубже понять, как люди взаимодействуют с технологиями. Они позволяют исследовать влияние конфигураций интерфейсов на восприятие и поведение пользователей, изменение системы в ответ на обратную связь, влияние культурных и контекстуальных факторов на понимание языковых конструкций в интерфейсах и человеко-машинное взаимодействие в целом. Таким образом, при изучении фактического взаимодействия пользователей с роботом важны как наблюдаемые поведенческие реакции и способы коммуникации, так и субъективное восприятие, интерпретации пользователями этого взаимодействия, которые формируются на основе ранее сложившихся представлений и конкретной ситуации взаимодействия.

Рассматривая реальность как социальный конструкт, П. Бергер и Т. Лукман подчеркивали, что реальность воспроизводится людьми во время ее интерпретации и формулирования знания о ней [Бергер, Лукман, 1995]. Данные знания, входящие в так называемый здравый смысл, поддерживаются за счет социального взаимодействия. С опорой на идеи социального конструктивизма, феноменологии и этнометодологии была выдвинута дополнительная гипотеза исследования (гипотеза 2): *пользователям комфортнее взаимодействовать с роботом, который соответствует их ожиданиям.*

## Материалы и методы

Для проверки гипотез исследования был организован лабораторный эксперимент, который проходил 6—8 декабря 2019 г. на базе ООО «Промобот» в Перми. В задачи исследования входило сравнение разных возрастных групп по двум направлениям: 1) процесс взаимодействия пользователей с сервисным социальным роботом; 2) восприятие и оценка пользователями своего опыта взаимодействия с роботом в заданной ситуации.

Экспериментальная ситуация состояла в имитации получения стандартной услуги МФЦ по замене паспорта гражданина РФ посредством обращения к человеку (сотруднику МФЦ, в роли которого выступали специально подготовленные сотрудники компании «Промобот») и посредством обращения к сервисному антропоморфному роботу Promobot V.4 (далее робот). Этот робот имеет строение, подобное строению человеческого тела, может двигать руками и головой, отображать анимированную мимику или эмоции на лицевом экране, вести разговор с помощью встроенных механизмов распознавания и синтеза речи, выводить информацию на сенсорный экран, используемый в качестве дополнительного интерфейса<sup>1</sup>.

Каждый участник в роли клиента МФЦ по очереди взаимодействовал с четырьмя агентами: человеком (консультантом МФЦ в его обычном режиме работы), человеком с коммуникацией по скрипту (консультантом МФЦ, который общается с клиентом исключительно диалоговым скриптом, реализованным у робота), роботом в обычном режиме работы (роботом в его обычном автономном режиме работы), роботом с расширенной коммуникацией (роботом под дистанционным управлением оператора в целях повышения гибкости и эмоциональности коммуникации). Применение разных вариантов коммуникации у робота и человека позволило определить, как пользователи воспринимают и оценивают разные характеристики конкретного робота, сравнивая их с человеческой коммуникацией и со своими ожиданиями относительно робота.

На экспериментальной площадке были представлены два «консультанта МФЦ» (специально подготовленные сотрудники компании «Промобот»), которые сидели за рабочими столами, и один робот. Участники эксперимента (в роли пользователей) обращались ко всем агентам за получением одной услуги, при этом к роботу подходили дважды (взаимодействуя с ним в двух режимах). Порядок обращения к различным агентам был произвольным, но при обращении за получением услуги к роботу участники в первый раз взаимодействовали с роботом в его обычном режиме работы, а во второй раз — с роботом с расширенной коммуникацией. Участники эксперимента получали необходимый макет документов и краткое описание их роли в эксперименте, при этом заранее им не сообщалось, что используются разные режимы работы робота или человека-консультанта.

Отбор участников эксперимента происходил стихийно, в том числе с использованием цифровых социальных сетей. Добровольцы приглашались к исследованию взаимодействия человека и робота на площадке компании «Промобот». Всего в исследовании приняли участие 40 добровольцев, было проведено четыре фокус-группы. При этом участники были разделены на группы по возрасту, что

<sup>1</sup> Подробную информацию о внешнем виде и функциональных возможностях робота см. на сайте Promobot. URL: <https://promo-bot.ai/robots/promobot-v4/> (дата обращения: 13.10.2024).

было необходимо для выявления различий между разными возрастными группами и обеспечения однородности фокус-групп. Состав участников лабораторного эксперимента и их ключевые характеристики представлены в таблице 1.

Таблица 1. Характеристики участников исследования

Целевые группы участников	Количество человек	Мужчины / женщины	Возраст
Молодые участники	29	18/11	18—30 лет
Участники среднего и старшего возраста	11	5/6	31—54 года

В ходе эксперимента информация собиралась двумя основными методами: 1) наблюдение за ходом взаимодействия участников с разными агентами: видео-записи и дневники наблюдателей, 2) фокус-группы с участниками эксперимента непосредственно после взаимодействия с четырьмя агентами.

### Невербальные реакции участников исследования при взаимодействии с роботом и человеком в различных режимах (по результатам наблюдения)

#### Особенности невербальных реакций группы участников среднего и старшего возраста

При первом обращении к консультанту или роботу участники данной группы были сдержанны в своем поведении: больше слушали, чем говорили; внимательно следили за реакциями собеседника (особенно робота). В процессе взаимодействия быстро адаптировались к условиям эксперимента и начинали вести себя более свободно и раскрепощенно (активно жестикулировали, задавали вопросы и т. д.).

Общаясь с человеком-консультантом (в обоих режимах), участники исследования часто демонстрировали отстраненное, скужающее поведение: четко отвечали на поставленные вопросы, не проявляли инициативу, не смотрели на человека либо разглядывали окружающую обстановку.

Взаимодействуя с роботом, участники демонстрировали более разнообразное поведение. Большая часть участников делали попытки наладить коммуникацию с роботом: ожидали от робота большей инициативы, задавали вопросы, отвечали на вопросы робота распространенными предложениями. Сожалели, если робот мало шел на контакт. Все участники, кому робот демонстрировал симпатию и говорил комплименты, реагировали улыбками, пытались тут же обсудить такое поведение робота.

Часть участников этой группы не были склонны поддерживать активное взаимодействие с роботом и стремились только четко исполнить данную им инструкцию. При этом демонстрировали раздражение, если подача инструкции задерживалась (робот задавал больше вопросов или ломался).

Если происходила заминка в работе робота, большинство участников этой группы пытались решить вопрос самостоятельно (повторяли вопрос к роботу, соверша-

ли попытки ввести информацию на экране) и лишь потом обращались за помощью. При этом могли обратиться как к наблюдателям эксперимента, так и к «консультантам МФЦ».

Все участники этой группы не реагировали на движения робота. Больше взаимодействовали с экраном, нежели слушали голосовые сообщения робота.

При повторном подходе к роботу демонстрировали более уверенное поведение и меньшее стремление пообщаться с ним.

### *Особенности невербальных реакций группы участников молодого возраста*

Большинство участников данной группы демонстрировали схожее поведение при взаимодействии как с человеком-консультантом, так и с роботом. Как правило, некоторое время с начала эксперимента участники вели себя неуверенно (выдерживали большую дистанцию при взаимодействии, мало жестикулировали, больше слушали). Адаптируясь к условиям эксперимента, начинали действовать четко по инструкции и показывали уверенное поведение (уменьшали дистанцию при взаимодействии, демонстрировали более четкую, громкую речь, активнее жестикулировали).

Стоит отметить, что при взаимодействии с человеком-консультантом участники данной возрастной группы — как при первом, так и при втором подходе, — показывали отстраненное и сдержанное поведение: сидели, откинувшись на спинку стула, четко отвечали на поставленные вопросы, мало проявляли инициативу.

При взаимодействии с роботом большинство участников данной возрастной группы действовали быстро, четко, уверенно. Движения выглядели естественными и привычными. Мало обращали внимания на голосовые сообщения робота, больше взаимодействовали с экраном. При этом не наблюдалось волнения или напряжения. Если робот допускал ошибку, большинство участников спокойно повторяли необходимые действия.

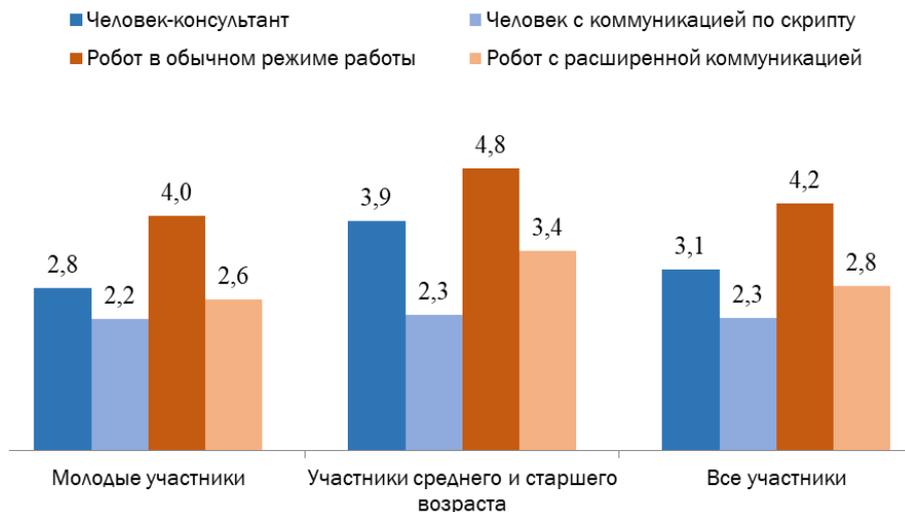
Если случалась заминка в работе робота, большинство участников этой группы пытались решить проблему самостоятельно, обращались за помощью к наблюдателям лишь при невозможности решить проблему длительное время.

Также среди представителей данной возрастной группы были участники, которые подходили к роботу (особенно первый раз) осторожно, с некоторым волнением. При взаимодействии с роботом четко исполняли инструкцию, которую предпочитали читать на экране. Начинали действовать по инструкции до того, как робот окончит фразу. Слабо реагировали на комплименты со стороны робота, не отвлекаясь от выполнения инструкции. Не ждали обратной связи от робота. Терялись, если робот допускал ошибку или ломался. После чего прекращали попытки взаимодействовать с ним и сразу обращались за помощью к кому-нибудь из наблюдателей. При втором подходе к роботу участники действовали быстрее, демонстрировали более уверенное поведение.

Оценивая хронометраж, можно заметить, что участники среднего и старшего возраста в обоих случаях взаимодействовали с роботом дольше, чем молодые участники исследования (см. рис. 1). При этом во всех возрастных группах наименее продолжительным было взаимодействие с человеком-консультантом, который общался по скрипту (в среднем 2,3 минуты), а наиболее продолжительным — пер-

вое взаимодействие с роботом в его обычном режиме работы (4,0 минуты среди молодых участников и 4,8 минуты среди участников среднего и старшего возраста). Время взаимодействия с роботом, действующим в режиме расширенной коммуникации, во всех экспериментальных группах меньше, чем время взаимодействия с роботом в обычном режиме, так как это было повторное получение услуги, то есть взаимодействие с роботом на основе уже полученного опыта.

Рис. 1. Среднее время, затрачиваемое участниками эксперимента на получение услуги при взаимодействии с разными агентами (количество минут)



Таким образом, ярких различий в невербальных реакциях участников исследования не выявлено. Особенности реакций связаны преимущественно с возрастом и наличием опыта взаимодействия как с людьми, так и с техническими устройствами.

Учет хронометража показал, что:

- участники всех групп провели больше времени, взаимодействуя с роботом, чем с человеком-консультантом;
- участники всех групп сократили время взаимодействия с роботом при втором подходе к нему;
- среднее время взаимодействия с роботом при первом подходе к нему в группах участников молодого возраста меньше, чем среднее время взаимодействия с роботом в группе участников среднего и старшего возраста.

### Субъективная оценка участниками эксперимента опыта взаимодействия с роботом (по результатам фокус-групп)

В фокус-группах обсуждались ожидания участников эксперимента от сервисных роботов, их реализация либо изменение в ходе экспериментального взаимодействия с роботом Promobot V.4.

Позитивные и негативные ожидания относительно сервисных роботов влияют на желание (или нежелание) пользователей взаимодействовать с ними. По результатам фокус-групп можно выделить три группы мотивов, определяющих приоритет пользователей в выборе способа получения услуги у человека-консультанта или у сервисного робота: 1) оценка функциональности взаимодействия (возможности быстро и эффективно решить задачу), 2) предсказуемость и привычность ситуации (снижение неопределенности), 3) психоэмоциональный комфорт (отсутствие лишних психоэмоциональных затрат, избегание неловких ситуаций, снижение риска собственной ошибки и негативной ответной реакции) и личные предпочтения (интерес к технологиям, интерес к живому общению). Далее рассмотрим ожидания и оценки взаимодействия с роботом по этим параметрам, полученные от представителей разных возрастных групп.

В оценках простоты и функциональности получения услуг у человека-консультанта и у робота участники среднего и старшего возраста чаще отмечали, что им проще взаимодействовать с человеком, тогда как молодежь выбирала робота.

*Иногда с роботом проще... потому что тебе не нужно ни с кем контактировать, ты сделал свое дело и ушел. (Молодые участники)*

*[При общении с человеком] ты терпишь там и сбрасываешь позже, в итоге у тебя эмоциональная нагрузка дольше длится, а тут сразу роботу говоришь, ну, он же все равно не обидится, он не может обидеться, ему не на что обижаться, и... ты ему говоришь просто «не тупи там, давай». (Молодые участники)*

Показательны различия в субъективном восприятии времени, потраченного на получение услуги у человека и у робота. Участники среднего и старшего возраста чаще считают, что они дольше получали услугу при взаимодействии с роботом, чем с человеком-консультантом. По их оценкам, им проще взаимодействовать с человеком-консультантом: необходимо только озвучить информацию, и консультант ее внесет. С роботом же требуется внести данные самостоятельно, что замедляет процесс. Напротив, молодые пользователи чаще отвечали, что услугу у робота они получили быстрее, чем у человека, и что при самостоятельном вводе данных тратится меньше времени, чем при взаимодействии с человеком-консультантом. По субъективным оценкам молодых участников, услуга оказывается быстрее, так как не приходится сидеть и просто ждать. Однако почти все участники фокус-групп сказали, что во второй раз у робота они получили услугу быстрее, чем в первый раз (это подтверждается и фактическими затратами времени, см. рис. 1).

Участники исследования отметили более высокую включенность в процесс при взаимодействии с роботом, так как в этом случае пользователь сам контролирует вводимую информацию. При общении с человеком-консультантом участники отвлекались во время пауз, меньше следили за процессом. Молодые участники говорили, что коммуникации с роботом было даже больше, чем с человеком. Напротив, участники среднего и старшего возраста отмечали, что коммуникации с человеком-консультантом у них было больше, чем с роботом. Они недостаточно

общались с роботом, так как были больше сосредоточены на выполнении задания, нежели на коммуникации.

Чувство психоэмоционального дискомфорта, возникающее у некоторых участников исследования в процессе взаимодействия с роботом, во многом связано с самооценкой собственной компетентности (наличия необходимых знаний, навыков по обращению с роботом) и со степенью испытываемого социального давления. Некоторые участники исследования (преимущественно среди пользователей среднего и старшего возраста, реже — среди молодых пользователей) сообщили, что при возникновении проблем при взаимодействии с роботом они испытывают неловкость, переносят вину на себя (свои неправильные действия, отсутствие необходимых знаний), переживают по поводу восприятия их «другими» людьми.

*Я себя с роботом почувствовала абсолютной душой в этот раз, потому что у меня ничего не получилось ввести, у меня ничего не получилось... (Участники среднего и старшего возраста)*

*Я здесь с роботом больше переживала, что я что-то не так делаю, что я задержу, и все такое. (Молодые участники)*

Одновременно с этим пользователи среднего и старшего возраста негативнее, чем молодые пользователи, относятся к ошибкам и техническим сбоям со стороны робота. Часть молодых пользователей выражают свое доброжелательное отношение к роботу и спокойное принятие его ошибок, технических неполадок, что объясняется, во-первых, сниженными ожиданиями относительно текущего уровня развития технологий, во-вторых, возможностью выразить свои эмоции непосредственно в момент «ошибки» без необходимости держать их в себе, как при общении с человеком. Участники среднего и старшего возраста, напротив, больше внимания обращают на технические сбои, недоработки, которые вызывают у них досаду. Это может объясняться как повышенным вниманием участников среднего и старшего возраста именно к функциональности роботов, так и нежеланием обнаружить свою некомпетентность во взаимодействии с роботом.

В целом в группе участников среднего и старшего возраста и, в меньшей степени, в группе молодых участников готовность к взаимодействию с роботами снижена по следующим причинам:

- неверие в функциональные (технические) возможности робота, недостаток информации о его возможностях (более значимо для участников среднего и старшего возраста);
- неуверенность в собственной компетентности при взаимодействии с роботом (отмечалось и возрастными участниками, и небольшой частью молодых пользователей);
- настороженность в отношении ко всему новому, непривычному (отмечалось и возрастными участниками, и небольшой частью молодых пользователей).

Со стороны участников среднего и старшего возраста наблюдалось более инструментальное (прагматично-функциональное) отношение к роботу: они воспринимают робота в качестве инструмента, а не партнера. Молодые участники иссле-

дования в большей степени относятся к роботу в контексте прогресса, развития возможностей. Среди молодых пользователей выражено отношение к роботам как к инструментам, которые облегчают деятельность человека.

*Инструмент, упрощает жизнь. Мы не на равных (смех). Для меня он не конкурент и не партнер, потому что он не может развить мысль какую-то... Мне кажется, у обоих партнеров должны быть цели, которые пересекаются где-то, а здесь у робота не могут быть цели. (Молодые участники)*

*Сейчас большинство людей воспринимает подобного робота как некую простую игрушку. Вот она стбит очень много, она непонятно зачем... К 30-му году, наверно, поменяется иерархия: скорее всего, большинство людей будут уже воспринимать роботов как некий полезный инструмент, а не как развлекательный центр. (Молодые участники)*

При этом некоторые молодые пользователи считают, что в будущем, по мере развития технологий, роботы смогут выполнять роль партнера: «Ну вот если у робота появится полноценный искусственный интеллект, который будет развиваться... он будет считаться как человек» (Молодые участники).

В процессе экспериментального взаимодействия с четырьмя агентами участники исследования не заметили использование нестандартных режимов работы (человек с коммуникацией по скрипту, робот с расширенной коммуникацией), но уже в ходе фокус-групповой дискуссии они отметили особенности и свои предпочтения относительно разных агентов.

### *Робот vs человек*

Участники отметили такие преимущества робота перед человеком, как функциональность, возможность для пользователя не тратить силы на поддержание беседы, при желании открыто выражать свое недовольство — без опасений получить негативную реакцию от робота. Участники исследования, предпочитающие взаимодействовать с человеком, объясняют свой выбор возможностью получения уточняющей, дополнительной информации, привычной и надежной ситуацией взаимодействия, более простым и понятным процессом общения. При сравнении человека-консультанта и сервисного робота участники исследования отмечают, что им было бы проще подойти к роботу, если бы он первым реагировал на человека, демонстрируя (вербально или невербально) свою готовность и приглашая к взаимодействию.

### *Человек в обычном режиме vs человек с коммуникацией по скрипту*

Большинство участников не заметили разницы между человеком в обычном режиме работы и человеком с коммуникацией по скрипту. Вероятно, это объясняется значительно стандартизированной работой консультантов при оказании услуг. Когда участники исследования узнавали об использовании консультантом коммуникации по скрипту, некоторые негативнее оценивали данный тип взаимодействия по сравнению с общением с человеком в обычном режиме, так как это нарушало привычный способ коммуникации.

### *Робот в обычном режиме vs робот с расширенной коммуникацией*

Участники не сразу заметили разницу между двумя режимами работы робота. Предпочтения по этим системам взаимодействия неоднозначны. С одной стороны, участникам нравились эмоциональность и комплименты робота с расширенной коммуникацией, но многим это мешало или казалось излишним.

Молодым участникам исследования взаимодействовать с роботом с расширенной коммуникацией понравилось больше, чем участникам среднего и старшего возраста. В качестве преимущества они отмечали индивидуальную обратную связь: узнал их, повернулся в их сторону, говорил им нестандартные фразы, выводил «сердечки» и т. д. Участники среднего и старшего возраста были более сосредоточены на технической стороне получения услуги у робота (ввод информации, сканирование документов) и не обращали внимания на его коммуникативные особенности.

### **Выводы**

Большинство участников исследования воспринимают робота в качестве инструмента, технического средства, а не партнера или компаньона. Такое отношение к роботам — с акцентом на их функциональные возможности — в большей степени характерно для участников среднего и старшего возраста. Возможность для человека при взаимодействии с роботом не вкладываться эмоционально, не задумываться о том, как тебя воспринимают, при этом получать положительную реакцию от робота — существенное преимущество взаимодействия между роботом и человеком.

Восприятие и оценка разных агентов взаимодействия во многом определяются тем, насколько ситуация соответствует сложившимся ожиданиям пользователей. По оценкам участников исследования, для них важна возможность идентифицировать, с кем они взаимодействуют, чтобы выстраивать ролевую модель этого взаимодействия. По результатам фокус-групп, взаимодействие с человеком с коммуникацией по скрипту, как и (в вероятном будущем) общение с полностью антропоморфным роботом, по внешности неотличимым от человека, вызывает у участников скорее негативные эмоции, так как нарушает ролевые ожидания. В большей степени участников исследования привлекла бы обратная ситуация: робот, внешне легко идентифицируемый в качестве робота, но по поведению и общению более похожий на человека.

Таким образом, гипотеза 2 о том, что пользователям комфортнее взаимодействовать с роботом, который соответствует их ожиданиям, подтвердилась. Это проявляется, в частности, в более быстром и, по субъективным оценкам, более простом взаимодействии с роботом при повторном обращении к нему. Несформированность ожиданий относительно роботов неоднозначно влияет на реальное взаимодействие с конкретным роботом: в одних случаях это приводит к росту неопределенности и чувству тревоги при взаимодействии с роботом (особенно среди пользователей среднего и старшего возраста), но в некоторых случаях это вызывает интерес к фактическим возможностям робота. Можно предположить, что при отсутствии сформированных установок (ожиданий) по отношению к роботам на взаимодействие пользователя с конкретным роботом будут влиять более общие социальные установки отношения к новым технологиям, инновациям вообще.

Как показали результаты эксперимента и фокус-групп, молодые пользователи в целом легче, чем пользователи среднего и старшего возраста, адаптируются к ситуации взаимодействия с сервисным роботом. При этом молодые пользователи позитивнее оценивают взаимодействие как с роботом в его обычном режиме работы, так и с роботом с расширенной коммуникацией, а также они более терпимы (снисходительны) к ошибкам, техническим несовершенствам роботов.

Для большинства участников среднего и старшего возраста и некоторой части молодых участников исследования характерно умеренно-положительное отношение к роботам: признание некоторых их преимуществ, готовность к взаимодействию с ними, но с опасениями по поводу функциональных (технических) возможностей роботов и своей компетентности в работе с ними. Преодолению этих барьеров человеко-машинного взаимодействия могут способствовать:

- 1) развитие функциональности и коммуникационной гибкости роботов;
- 2) расширение опыта взаимодействия с роботами, внедрение их в привычные практики;
- 3) на начальных этапах — сопровождение роботов человеком (консультантом) для оказания поддержки и обучения, то есть присутствие компетентного и при этом понятного «проводника» в коммуникации между роботом и человеком.

Хотя гипотеза 1 о возрастных различиях во взаимодействии с сервисным роботом подтвердилась, стоит отметить, что и среди молодых участников исследования встречаются те, кому некомфортно взаимодействовать с роботом, преимущественно по причине неуверенности в собственной компетентности. На процесс взаимодействия пользователей с сервисным роботом и субъективную оценку этого взаимодействия влияет не только возраст (поколенческие особенности), но и другие факторы, требующие дальнейшего изучения.

К ограничениям представленного исследования можно отнести состав участников и процедуру сбора данных. Поскольку для участия в эксперименте набирались добровольцы, можно отметить, что большинство участников исследования изначально проявляли интерес к теме роботов и новых технологий в целом. Относительно небольшое число участников и отсутствие среди них ярко выраженных технофобов не позволяет переносить полученные результаты на всех потенциальных пользователей. Мнения и оценки, высказанные в ходе фокус-группового обсуждения, а также особенности взаимодействия с роботом в лабораторных условиях могут отличаться от взаимодействия пользователей с сервисным роботом в естественных условиях.

## Список литературы (References)

1. Абрамов Р. Н., Катечкина В. М. Социальные аспекты взаимодействия человека и робота: опыт экспериментального исследования // Журнал социологии и социальной антропологии. 2022. Т. 25. № 2. С. 214—243.  
Abramov R. N., Katechikina V. M. (2022) Social Aspects of Human-Robot Interaction: Experimental Research Experience. *The Journal of Sociology and Social Anthropology*. Vol. 25. No. 2. P. 214—243. (In Russ.)

2. Бергер П., Лукман Т. Социальное конструирование реальности. Трактат по социологии знания. М.: Медиум, 1995.  
Berger P., Lukman T. (1995) *The Social Construction of Reality. A Treatise in the Sociology of Knowledge*. Moscow: Medium. (In Russ.)
3. Гаврилина Е. А. Агентность не-человеков: взаимодействие людей и социальных роботов // Мониторинг общественного мнения: экономические и социальные перемены. 2023. № 3. С. 41—55. <https://doi.org/10.14515/monitoring.2023.3.2318>.  
Gavrulina E. A. (2023) Agency of Non-humans: Interaction of Humans and Autonomous Intelligent Systems. *Monitoring of Public Opinion: Economic and Social Changes*. No. 3. P. 41—55. <https://doi.org/10.14515/monitoring.2023.3.2318>. (In Russ.)
4. Зильберман Н. Н., Стефанцова М. А. Социальный робот: подходы к определению понятия // Современные исследования социальных проблем (электронный научный журнал). 2016. № 11. С. 297—312. <https://doi.org/10.12731/2218-7405-2016-11-297-312>.  
Zilberman N. N., Stefantsova M. A. (2016) Social Robot: Defining the Concept. *Modern Research of Social Problem*. No. 11. P. 297—312. <https://doi.org/10.12731/2218-7405-2016-11-297-312>. (In Russ.)
5. Сачмен Л. Реконфигурации отношений человек — машина: планы и ситуативные действия / пер. с англ. А. С. Максимовой; под ред. А. М. Корбута. М.: Элементарные формы, 2019.  
Suchman L. (2019) Plans and Situated Actions: The Problem of Human — Machine Communication. Moscow: Elementary forms. (In Russ.)
6. Broekens J., Heerink M., Rosendal H. (2009) Assistive Social Robots in Elderly Care: A Review. *Gerontechnology*. Vol. 8. № 2. P. 94—103.
7. Gnamb T., Appel M. (2019) Are Robots Becoming Unpopular? Changes in Attitudes Towards Autonomous Robotic Systems in Europe. *Computers in Human Behavior*. No. 93. P. 53—61. <https://doi.org/10.1016/j.chb.2018.11.045>.
8. Hegel F., Muhl C., Wrede B., Hielscher-Fastabend M., Sagerer G. (2009) Understanding Social Robots. In: *Second International Conference on Advances in Computer-Human Interaction*. Cancun: IEEE. P. 169—174. <https://doi.org/10.1109/ACHI.2009.51>.
9. Hudson J., Orviska M., Hunady J. (2017) People's Attitudes to Robots in Caring for the Elderly. *International Journal of Social Robotics*. No. 9. P. 199—210. <https://doi.org/10.1007/s12369-016-0384-5>.
10. Katz J. E., Halpern D., Crocker E. T. (2015) In the Company of Robots: Views of Acceptability of Robots in Social Settings. In: Vincent J., Taipale S., Sapio B., Lugano G., Fortunati L. (eds.) *Social Robots from a Human Perspective*. Cham: Springer. P. 24—38. [https://doi.org/10.1007/978-3-319-15672-9\\_3](https://doi.org/10.1007/978-3-319-15672-9_3).

11. Mandl S., Bretschneider M., Asbrock F., Meyer B., Strobel A. (2022) The Social Perception of Robots Scale (SPRS): Developing and Testing a Scale for Successful Interaction Between Humans and Robots. In: Camarinha-Matos L. M., Ortiz A., Boucher X., Osório A. L. (eds.) *Collaborative Networks in Digitalization and Society 5.0. PRO-VE. IFIP Advances in Information and Communication Technology*. Vol. 662. Cham: Springer. P. 321—334. [https://doi.org/10.1007/978-3-031-14844-6\\_26](https://doi.org/10.1007/978-3-031-14844-6_26).
12. Naneva S., Sarda M., Webb T. L., Prescott T. J. (2020) A Systematic Review of Attitudes, Anxiety, Acceptance and Trust Towards Social Robots. *International Journal of Social Robotics*. No. 12. P. 1179—1201. <https://doi.org/10.1007/s12369-020-00659-4>.

DOI: [10.14515/monitoring.2024.5.2602](https://doi.org/10.14515/monitoring.2024.5.2602)



**Е. Г. Ним**

## **ПОП-КУЛЬТУРА, ФЭНДОМЫ И НЕЙРОСЕТИ: ФАНАТЫ ВСТРЕЧАЮТСЯ С ИИ**

### **Правильная ссылка на статью:**

Ним Е. Г. Поп-культура, фэндомы и нейросети: фанаты встречаются с ИИ // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 183—202. <https://doi.org/10.14515/monitoring.2024.5.2602>.

### **For citation:**

Nim E. G. (2024) Pop Culture, Fandoms and Neural Networks: Fans Meet AI. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 183–202. <https://doi.org/10.14515/monitoring.2024.5.2602>. (In Russ.)

Получено: 28.03.2024. Принято к публикации: 15.08.2024.

ПОП-КУЛЬТУРА, ФЭНДОМЫ И НЕЙРОСЕТИ:  
ФАНАТЫ ВСТРЕЧАЮТСЯ С ИИ

*НИМ Евгения Генриевна — кандидат социологических наук, доцент Института медиа факультета креативных индустрий, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия  
E-MAIL: nimeg@mail.ru  
<https://orcid.org/0000-0001-7349-9429>*

**Аннотация.** Статья посвящена анализу и концептуализации процессов распространения ИИ-технологий в фан-культуре. Автор рассматривает фэндомы как медиатизированные креативные сообщества, раскрывает взаимосвязи искусственного интеллекта и фан-культуры, выявляет различные типы отношения к генеративному ИИ на примере двух фан-комьюнити. В статье выделены основные линии пересечения ИИ и фэндома: технологии ИИ модифицируют объекты фанатского интереса; меняют пространство, структуру и креативные практики фэндомов; встраиваются в инновационные фанатские сервисы; применяются для аналитики фэндомов; актуализируют вопросы этики и права в культурах соучастия. Экспансия искусственного интеллекта (в частности, нейросетей) вызвала поляризацию фэндома: в то время как одни фанаты активно применяют нейросети для редактирования и создания контента, другие считают это неуместным. Предварительные наблюдения в двух фан-комьюнити показали, что критические и протестные фандискурсы вокруг генеративного ИИ сфокусированы на трех идеях: а) использование ИИ противоречит концепции фэндома; б) сегодняшний ИИ как технология несовершенен; в) текущие практики обучения и применения ИИ неэтичны. В статье обозначены дальнейшие перспективы изучения взаимодействия фэндомов с искусственным интеллектом, которые позволят расширить границы как *fan studies*, так и социальных исследований ИИ.

## POP CULTURE, FANDOMS AND NEURAL NETWORKS: FANS MEET AI

*Evgeniya G. NIM<sup>1</sup> — Cand. Sci. (Soc.), Associate Professor at the Institute of Media, Faculty of Creative Industries  
E-MAIL: nimeg@mail.ru  
<https://orcid.org/0000-0001-7349-9429>*

<sup>1</sup> HSE University, Moscow, Russia

**Abstract.** The article is dedicated to analyzing and conceptualizing the processes of disseminating AI technologies in fan culture. The author examines fandoms as mediatised creative communities, reveals the relationship between artificial intelligence and fan culture, and identifies the practices of generative AI reception using the example of two fan communities. The article highlights the main lines of intersection between AI and fandom: AI technologies modify objects of fan interest; change the space, structure, and creative practices of fandoms; are integrated into innovative fan services; are used for fandom analytics; foreground issues of ethics and law in participatory cultures. The expansion of artificial intelligence (neural networks in particular) has caused a polarization of fandom: while some fans actively use neural networks to edit and create content, others consider it inappropriate. Preliminary observations in two fan communities showed that critical and protest fan discourses around generative AI are focused on three ideas: a) the use of AI contradicts the concept of fandom; b) today's AI as a technology is imperfect; c) current practices in AI training and application are unethical. The article outlines further prospects for studying the interaction of fandoms with artificial intelligence, which will expand the boundaries of both fan studies and social studies of AI.

**Ключевые слова:** фан-культура, фэндомы, популярная культура, искусственный интеллект, генеративный ИИ

**Keywords:** fan culture, fandoms, popular culture, artificial intelligence, generative AI

## Введение

Распространение технологий искусственного интеллекта — одно из проявлений глубокой медиатизации социальной жизни [Hepp, 2020]. Будучи невидимым, но проникая всюду, искусственный интеллект меняет идентичность и повседневную жизнь людей, трансформирует организации, социальные институты, национальные государства и глобальную экономику [Левашов, Гребняк, 2023; Резаев, Трегубова, 2019; Elliott, 2019: xix]. В частности, искусственный интеллект заметно влияет на сферы культурного потребления и производства, превращаясь, по выражению Льва Мановича, в «культурный ИИ» [Manovich, 2019], формирующий эстетику, вкус и культурную экосистему в целом. Технологии искусственного интеллекта начинают широко использоваться в креативных индустриях, возникает новый тип медиа, частично или полностью созданных ИИ, — синтетические медиа [Kalrokas, 2021]. ИИ становится способным заместить труд сценаристов, актеров, музыкантов, художников, писателей, журналистов, маркетологов и аналитиков, создавая разные типы культурного контента [Danesi, 2024; Jin, 2021]. Примечательно, что долгое время популярная культура формировала образ искусственного интеллекта в научно-фантастических романах, фильмах и сериалах [Дегтяренко и др., 2023; Лещинская и др., 2023; Barron, 2023; Hermann, 2023; Nader et al., 2022], а теперь сама (вос) производится с помощью ИИ-инструментов [Chow, 2020; Datta, Goswami, 2021].

Алгоритмы искусственного интеллекта лежат в основе рекомендательных систем цифровых платформ, подсказывающих пользователям, что следует смотреть, слушать, читать и покупать [Грушевская, 2022; Manovich, 2019]. Иначе говоря, ИИ не только отслеживает, но и во многом определяет предпочтения медиапотребителей, включая сферу развлечений. И если этот аспект влияния «культурного ИИ» уже находится в поле внимания исследователей [Jin, 2021], то практики активного применения (генеративного) ИИ самой аудиторией, а именно фанатами поп-культуры, пока малоизучены. Отчасти потому, что нейросетевые технологии, доступные конечным потребителям (такие как ChatGPT, Midjourney, DALL-E и др.), появились совсем недавно. Фэндомы — активные производители контента, связанного с огромным множеством медиаобъектов (фильмов, сериалов, видеоигр, музыки, селебрити и т. д.). Создание этого контента предполагает определенные медийные навыки и их развитие, включая освоение новых ИИ-сервисов.

В этой статье анализируются и концептуализируются процессы распространения ИИ-технологий в фан-культуре. Мы начинаем анализ с интерпретации фэндомов как медиатизированных сообществ, далее раскрываем взаимосвязи (линии пересечения) искусственного интеллекта и фан-культуры и, наконец, рассматриваем фанатское видение генеративного ИИ на примере двух фан-платформ (вики-хостинга Fandom и репозитория фан-работ Archive of Our Own<sup>1</sup>). Отметим, что

<sup>1</sup> С апреля 2023 г. сайт Archive of Our Own ([www.archiveofourown.org](http://www.archiveofourown.org)) заблокирован на территории Российской Федерации, данные используются только в исследовательских целях.

эти фан-сообщества представлены как иллюстративные кейсы, а не объект полноценного этнографического исследования; к тому же наблюдаемые здесь обсуждения ИИ имеют преимущественно критический характер. Несмотря на эти ограничения, мы надеемся, что положения этой статьи помогут углубить понимание как фэндомов (в аспекте их взаимоотношений с цифровыми технологиями и ИИ), так и генеративного ИИ (в плане его восприятия различными социальными группами и субкультурами).

### Фэндомы как медиатизированные сообщества

История фэндома уходит вглубь веков и, вероятно, связана с практикой интерпретации и распространения мифов, а также с культом знаменитостей, их почитанием и славой. Понятие «фанат» — сокращение от «фанатик», оно все еще несет в себе ассоциации с религиозным рвением (лат. *fanaticus* — одержимый божеством). Впервые слово «фанат» в массмедиа появилось в спортивной колонке *The Washington Post* 10 октября 1896 г. [Lamerichs, 2018a: 19] и относилось к поклонникам бейсбола и других спортивных зрелищ. С развитием кинематографа, телевидения, радио и звукозаписи в начале XX века возникли фан-клубы кинозвезд и музыкальных исполнителей [Duffett, 2013: 30—35]. В частности, в конце 1920-х годов голливудские киностудии совокупно получали от поклонников более 32 млн писем ежегодно и к этому времени уже имели специальные отделы для отслеживания фанатской почты [ibid.: 32]. Не менее востребованным медиа стало коммерческое радио: например, суточная аудитория комедийного сериала «Эймс и Энди», запущенного одной из чикагских радиостанций, оценивалась в 40 млн человек (в период с 1929 по 1931 г.) [McCracken, 2018: 417]. Тогда же зародились литературные клубы любителей научной фантастики [Сорра, 2006: 42], детективов и приключений. Поклонники научно-фантастических романов стали издавать фэнзины — любительские малотиражные журналы с обзорами, новостями и фан-работами [Hellekson, 2018: 70]. В США множество читательских фан-сообществ появлялось вокруг бульварной периодики, в том числе десятки журналов, специализирующихся на любовных историях, вестернах, рассказах ужасов и прочих жанрах. При этом популярные журналы первой трети прошлого века уже включали «интерактивные» элементы — конкурсы, почту фанатов, колонки писем и т. д. [Edwards, 2018: 52].

Начиная с 1930-х годов фанатами стали называть группы людей, которым нравятся одни и те же медиаразвлечения [Booth, 2017: 22]. На наш взгляд, это время можно считать началом появления медиафэндома, который, изменяясь под воздействием многих социальных, культурных и технологических факторов, существует уже более столетия. Хотя в западных *fan studies* есть авторитетная традиция связывать понятие медиафэндома с культовым научно-фантастическим сериалом «Звездный путь» (1966—1969) или шпионской драмой «Агенты А. Н.К.Л.» (1964—1968) [Сорра, 2006: 43—44], эта точка отсчета не так очевидна. Она отдает приоритет телевизионным шоу (и ТВ-фэндомам), описывая более ранние фан-сообщества и практики, возникшие вокруг других типов медиа (радио, кинематограф, бульварные издания), скорее как предысторию медиафэндома. Однако, если вслед за Марком Даффетом понимать медиафэндом как «признание пози-

тивной личной, относительно глубокой эмоциональной связи с опосредованным (*mediated*) элементом популярной культуры» [Duffett, 2013: 24], представляется, что первые медиафэндомы появились задолго до середины 1960-х<sup>2</sup>.

Понятие медиафэндома предполагает, что в основе фанатских интересов лежит медиатекст, созданный для массового потребления. В эпоху интернета и цифровизации фэндом стремительно развил и другие аспекты своей медийности. Современная фанатская культура глубоко медиатизирована: она возникает вокруг медиаобъектов (романов, фильмов, сериалов, комиксов, телешоу, видеоигр, музыкальных групп, медиаперсон), коммуницирует посредством медиа (технологий, каналов и платформ) и производит медиаконтент (фанфикшн, фанарты, фанвидео, посты в социальных медиа и т. д.). Интересующая нас ИИ-«колонизация» фэндома [Lamerichs, 2018b] распространяется во всех этих направлениях. Согласно Генри Дженкинсу, медиафэндомы — это одна из форм культуры соучастия, возникшая до цифровизации, но получившая новые возможности в эпоху конвергенции и сетевых коммуникаций [Jenkins, 2018: 18]. В отличие от пассивных потребителей, фанаты «присваивают» и трансформируют коммерческие медиапродукты, создавая на их основе свои альтернативные версии полюбившихся историй, миров и персонажей. Как полагают Кристина Буссе и Карен Хеллексон, ключевым концептом для понимания фэндома является «незавершенная работа» (*work in progress*) [Busse, Hellekson, 2006: 7] в значении продолжения и расширения исходного текста, который всегда открыт для дальнейших интерпретаций и преобразований (даже если формально он завершен). Точно так же как поп-культурные тексты часто получают новые прочтения и воплощения (ремейки, сиквелы, адаптации и др.), фанатские работы и обсуждения продолжают, оспаривают и дополняют как первоначальную историю, так и друг друга, формируя целую вселенную вариаций вокруг первоисточника. Фэндом активен, пока включен в эту работу по созданию и обмену новыми смыслами и новым медиаконтентом.

Хотя трактовка фэндома как продуктивного сообщества довольно распространена, она проблематизируется внутри *fan studies*. Понимание фанатов прежде всего как «преобразователей» пришло вместе с «первой волной» *fan studies* в начале 1990-х годов, когда исследователям было важно установить различие между фанатами и «некритичными» потребителями поп-культуры. Однако наряду с «активистами» ученые также наблюдают «тайных», или «тихих», поклонников [Duffett, 2013: 66], которые не афишируют свои увлечения, не создают фан-контент или даже не идентифицируют себя как фанатов. Иначе говоря, «трансформативная работа», будь то фанатское творчество или гражданский активизм, сегодня не рассматривается как ключевая характеристика фанатства. Тем не менее концепция «трансформативного», «продуктивного» фэндома представляется вполне эвристичной для изучения использования ИИ в фанатских практиках и пространствах.

Фэндомы — не только креативные и продуктивные, но и аффективные сообщества [Lamerichs, 2018a: 17], основанные на любви. Хотя среди исследователей нет консенсуса по поводу того, кого считать фанатом, наличие позитивной глубо-

<sup>2</sup> Следует также отметить, что в ключевых работах по *fan studies* прежде всего представлена история американского и британского медиафэндома, что ограничивает наше представление о том, как возникали и формировались фэндомы в других региональных и культурных контекстах.

кой эмоциональной связи с объектом поп-культуры обычно признается важным его атрибутом. Фэндом имеет свою особую культуру чувств, формирующую фанатскую идентичность. Эти чувства мотивируют фанатов на добровольный труд — «lovebor» [Stanfill, 2019: 151], работу по проявлению любви. Настаивая на трактовке фанатской любви как работы, создающей ценность, Мел Стэнфилл опирается на традицию «серьезного» понимания женского репродуктивного труда [ibid.: 152]. Так же, как и домашний женский труд, фанатский вклад должен считаться не просто «естественным» выражением любви, а именно работой по «извлечению, упаковке и передаче» своих чувств. Фэндом основан на экономике дарения, но его дары, как замечает Карен Хеллексон, даются не бесплатно, а в обмен на нечто, что выполняет роль «валюты» в конкретном фан-комьюнити, например фанарты, фанфики, посты, лайки и т. д. [Hellekson, 2018: 71]. В этом контексте сам фэндом может рассматриваться как результат этой серии обменов, возникающий в цикле дарения и получения. «Классические» фэндомы имеют антирыночный характер, что тесно связано и с правовым аспектом их существования: пока фанатская работа не используется в коммерческих целях, она имеет больше шансов избежать судебных разбирательств за нарушение авторских прав. Однако в целом легитимность значительной части фан-работ находится в «серой» правовой зоне, что осложняет отношения между фан-комьюнити и индустрией [Schwabach, 2016].

Для понимания современной фан-культуры также важно обозначить дилемму между фэндомом как сообществом и индивидуальным фанатским опытом. Эта дилемма актуальна для текущих дискуссий внутри *fan studies*, где обсуждается, изучаем ли мы прежде всего комьюнити (фэндом) или фанатов [Gray, Sandvoss, Harrington, 2017: 7]. Традиция исследовать фан-культуру через призму фэндомов во многом сформирована Генри Дженкинсом, принадлежность к фан-сообществу считалась одним из ключевых критериев фанатской идентичности. В его «Текстовых браконьерах» 1992 г. фэндом рассматривается как особый способ восприятия медиатекстов; определенный набор критических и интерпретативных практик; основа потребительского активизма; специфическая форма культурного производства; альтернативное социальное сообщество [Jenkins, 2005: 284—287]. Эта модель описывает различные уровни фанатской вовлеченности, от «переговорного» прочтения текстов до политических движений. Дженкинс также трактует фэндом как место неформального обучения с низким барьером для творческого выражения и гражданского участия, где люди обмениваются знанием и контентом, ценят вклад каждого и социальную связь друг с другом [Jenkins, 2018: 18]. В то же время в *fan studies* распространен и другой подход, сближающий фанатов с широкой аудиторией и акцентирующий личный фанатский опыт. Например, Хенрик и Сара Линден наряду с «традиционными» фанатами медиаразвлечений выделяют новое поколение фанатов брендов [Linden, Linden, 2017: 39], что размывает и без того неочевидную границу между фанатами и активными потребителями. Все большая доступность медиапродукции, платформ и медиатехнологий способствует тому, что некогда специфические фаномные практики нормализуются, становясь частью личного опыта большинства «обычных» людей.

Эта тенденция к более широкому пониманию фанатов отражает путь, который прошел сам фэндом: от общественной и медийной стигматизации фанатов как

«Других» к мейнстримизации фан-культуры и «фандомизации» повседневной жизни [Gray, Sandvoss, Harrington, 2017]. Однако понимание фэндомов как пространства культурной автономии и сопротивления (заложенное Дженкинсом), где участники комьюнити, часто ассоциируемые с дискриминируемыми социальными группами, практикуют критическое чтение и «пересборку» индустриальных текстов, все еще актуально. Фэндомы способны проявлять себя как «контрпублики» [Chew, 2018; Marshall, 2020], оспаривая доминирующие дискурсы и социальные процессы (в частности, как мы увидим далее, сопротивляться автоматизации фан-культуры и экспансии технологий искусственного интеллекта).

### **Взаимосвязь искусственного интеллекта и фан-культуры**

С развитием и распространением искусственного интеллекта медиатизация фан-сообществ и фанатских практик переходит на новый уровень. Прежде чем обратиться к более детальному анализу этого процесса, попытаемся установить, как соотносятся ИИ и фэндом в общем, какого рода пересечения и взаимосвязи между ними можно выявить и проблематизировать. На наш взгляд, в пространстве конвергенции искусственного интеллекта и фан-культуры просматриваются семь «перекрестков», где фэндом встречается с ИИ (включая нейросети) и трансформируется под его влиянием.

Во-первых, мы наблюдаем *модификацию объектов фэндомов*, то есть медийных текстов, образов и персон, вокруг которых возникают фанатские сообщества. Технологии ИИ активно используются в культурном производстве [Danesi, 2024; Han, 2021; Jin, 2021]: нейросети пишут сценарии, истории и музыку, снимают фильмы, преобразуют внешность и голоса селебрити, создают их цифровые аватары и т. д. Посредством искусственного интеллекта синтезируются и полностью виртуальные люди [Burden, Savin-Baden, 2019] — не только умные персонажи видеоигр, но и цифровые инфлюенсеры и поп-звезды [Wu, Ahn, 2023]. Поскольку ИИ способен производить привлекательный контент, можно предположить, что в будущем возникнет множество фан-сообществ, посвященных ИИ-кумирам. Этот тренд особенно заметен в странах Восточной Азии, где процветает индустрия виртуальных айдолов (вокалоидов, V-туберов, виртуальных k-pop-групп, анимированных умных компаньонов) [Bredikhina et al., 2020; Liu, 2023; Nuranda, Purwangingtyas, 2023]. Если раньше ИИ был частью научно-фантастического нарратива, то сейчас из «персонажа» он превращается в самостоятельную виртуальную личность, наделенную своей (условной) агентностью и идентичностью. Активность таких ИИ разворачивается в медийных средах с разной степенью интерактивности и иммерсивности. Все эти трансформации объектов фанатского интереса определенным образом меняют и фанатский опыт.

Во-вторых, системы искусственного интеллекта преобразуют пространства фан-культуры, что проявляется в процессах *платформизации и алгоритмизации фэндомов* [Yin, 2020]. Эти пространства включают медиаплатформы для потребления коммерческих продуктов поп-культуры (онлайн-кинотеатры, стриминговые сервисы); социальные медиа, где фанаты формируют онлайн-сообщества и публикуют свой контент («ВКонтакте», YouTube, TikTok и др.); разнообразные тематические фан-сайты и ресурсы. Технологии ИИ применяются здесь прежде всего

для продвижения и дистрибуции контента. Например, стриминговые платформы *Netflix*, *Amazon Prime* и *YouTube* используют рекомендательные алгоритмы в целях персонализации зрительского опыта, предлагая телешоу, фильмы или ролики на основе прошлой активности пользователя (истории просмотров и предпочтений) [Frey, 2021]. Потребление развлечений становится настолько связанным с платформами и алгоритмами, что исследователи говорят о конвергенции ИИ, цифровых платформ и поп-культуры [Jin, 2021]. Другой пример — использование генеративного ИИ в пространствах культуры соучастия, в частности вики-проектах, таких как *Fandom*. ИИ здесь применяют для поисковой оптимизации, удаления вандализма, распознавания неэтичных изображений<sup>3</sup>. В то же время идея генерация вики-контента с помощью больших языковых моделей пока не получает широкого одобрения.

В-третьих, доступность ИИ-технологий меняет креативные фанатские практики, одновременно проблематизируя саму природу фэндома и фанатского творчества. Согласно Генри Дженкинсу, фанаты — это «текстовые браконьеры» [Jenkins, 2005], создающие альтернативные интерпретации коммерческих медиапродуктов. Для фэндома важна «трансформативная» работа, часто требующая хороших медийных навыков. Фанаты — не только потребители популярного контента, они также фикрайтеры, художники, продюсеры, монтажеры, переводчики, фандаберы и т. д. Любой фанатский труд имеет ценность в культурах соучастия, а многие авторы-фанаты обретают собственных поклонников. Однако революция нейросетей привела к тому, что фанфики пишут чат-боты типа ChatGPT, фанарты генерируются с помощью *Midjourney* и *DALL-E*, для перевода, дубляжа и производства аудиовизуальных текстов также используются нейросети. С одной стороны, ИИ-инструменты расширили возможности создания фанатского контента, с другой — стали предметом острых дискуссий о соответствии таких продуктов ключевым ценностям фан-сообщества [Lamerichs, 2018b]. Еще одна проблема — генерация дипфейков (анти)фанатами. Дипфейки могут развлекать и «воскрешать» ушедших звезд [Andrews, 2022], но также нести угрозу для репутации и благополучия селебрити, порождать конфликты и домыслы среди фанатов [Porova, 2020].

В-четвертых, поскольку фанаты как потребители являются «святым граалем для медиакультуры» [Zwaan, Duits, Belinders, 2014: 1], производители поддерживают их увлечения, разрабатывая в том числе и *специализированные фанатские ИИ-сервисы*. Например, в 2021 г. поклонникам k-поп-исполнителей была предложена услуга «*AI Private Call*» от корпорации *NCSoft* — имитация личного телефонного разговора с айдолами, чьи голоса были искусственно синтезированы. В настройках можно было выбрать время звонка, свой никнейм и стиль общения, однако звонок больше напоминал голосовое сообщение, поскольку не был двусторонним. Хотя используемая здесь технологии ИИ была обучена на реальных голосовых данных айдалов, пользователи ощутили утрату «аутентичности» своих кумиров, и сервис не получил фанатской поддержки [Kang et al., 2022]. Другой кейс — платформа *Character.AI* и схожие проекты, основанные на моделях глубокого обучения. *Character.AI* позволяет создавать чат-боты персонажей (как вымышленных,

<sup>3</sup> Практики использования генеративного ИИ на платформе *Fandom* будут подробнее рассмотрены в следующем разделе.

так и исторических) с высокой степенью детализации и интерактивности. Подобные продукты направлены на поддержание воображаемой коммуникации между поклонниками и знаменитостями, однако их успех не всегда очевиден, как показывает опыт с «*AI Private Call*». То, как фанаты используют и оценивают ИИ-сервисы, позволяющие устанавливать более «близкие» отношения с объектами их интереса, может стать предметом целой серии исследований.

В-пятых, под влиянием искусственного интеллекта *изменяется состав и структура фэндома*, его участниками будут не только люди, но и другие субъекты, в частности чат-боты. Нейросети уже способны писать посты и комментарии, рисовать фан-арты, сочинять истории и музыку, переводить и озвучивать видео — одним словом, быть продуктивными «членами» фанатского сообщества. Пока генеративный ИИ в большей мере воспринимается как технология, но с дальнейшим развитием его «социальности» искусственные акторы будут преодолевать свой «инструментальный» статус, претендуя на признание агентности (и, возможно, авторства). Как отмечает Николь Ламерикс, разнообразие в фэндоме вскоре будет связано не столько с привычными социальными признаками (национальность, гендер, возраст), сколько с различиями между людьми и не-людьми [Lamerichs, 2018b]. Вероятно, по мере перехода социальной жизни и культуры соучастия в метаверс взаимодействие людей с разного рода ИИ-персонами станет обычной практикой. Кроме того, уже сейчас фэндом полемизирует о том, заменит ли генеративный ИИ людей, не потеряют ли люди-авторы, создающие фан-работы, свои «рабочие места». Это особенно актуально для тех, кто отчасти монетизирует свой трудоемкий эксклюзивный контент через систему пожертвований (донатов). Возможность доступа к инновационным ИИ-сервисам, а также их преимущества (или ограничения) могут стать факторами цифрового неравенства внутри фан-комьюнити. Учитывая, что продвинутые версии генеративного ИИ часто требуют платной подписки или определенных условий регистрации, сама перспектива их использования актуализирует тему социальных привилегий.

В-шестых, технологии ИИ используются *для аналитики и исследования фэндома*. Фанаты составляют активную часть потребителей медиапродукции (фильмов, сериалов, книг, видеоигр, комиксов, музыкальных исполнителей и т. д.). Платформизация и алгоритмизация фэндомов предполагают, что платформы не только предоставляют доступ к своим коммерческим продуктам или возможности создания и распространения UGC-контента, но и собирают данные о потребителях (демографию и географию пользователей, историю просмотров, активность в социальных сетях и др.). Анализ этих данных с помощью ИИ позволяет создавать рекомендации для персонализации пользовательского опыта, таргетировать аудиторию, разрабатывать маркетинговые и рекламные стратегии, прогнозировать коммерческий успех [Han, 2021: 304]. Если включать в категорию фанатов также фанатов брендов [Linden, Linden, 2017] и политических фигур [Dean, Andrews, 2021], то станет очевидно, в каких масштабах могут применяться алгоритмы для изучения фанатских предпочтений (культурных, политических, потребительских) и их формирования. Помимо этого, возможности ИИ как «культурного теоретика» (по выражению Льва Мановича) используют ученые в своих исследованиях цифровой культуры [Manovich, 2019]. ИИ-этнография фэндома, основанная на анализе больших данных, может стать частью цифровой гуманитаристики.

Наконец, в-седьмых, проникновение искусственного интеллекта, и в частности генеративного ИИ, в пространство фан-культуры, актуализирует *правовые и этические проблемы применения нейросетей* [Fui-Hoon Nah et al., 2023]. Последствия экспансии генеративного ИИ затрагивают не только представителей культурных и креативных индустрий [Danesi, 2024; Jin, 2021], но и фанатские сообщества с их любительским творческим контентом. Например, значительная часть фанатов, как мы покажем далее, выступает против использования огромной базы фанфиков и фанартов для машинного обучения таких моделей, как ChatGPT, без согласия авторов. В то же время фанаты сами активно генерируют изображения с любимыми звездами и персонажами посредством *Midjourney*, *DALL-E* и других нейросетей. И хотя подобные изображения обычно не используются в коммерческих целях, их правовой статус не вполне определен. Кроме того, нередки случаи создания (анти)фанатами дипфейков со знаменитостями как в целях развлечения, так и для разрушения их репутации [Popova, 2020]. Таким образом, вопросы этики и права, связанные с развитием синтетических медиа, приобретают свою специфику в контексте фэндома и *fan studies*.

### **Нейросети и фан-комьюнити: дискуссии о применимости генеративного ИИ**

Использование медиатехнологий, во многих случаях требующих «продвинутых» компетенций, — неотъемлемая часть креативных практик цифрового фэндома [Booth, 2017]. В этом смысле привлечение ИИ-инструментов как компьютерных программ хотя и является инновационным, но не принципиально ново для фан-культуры. Нейросети так или иначе могут применяться для производства практически всего спектра фанатского контента: фанфиков, вики, фанартов, косплея, фанвидео, музыки, перевода и озвучивания текстов (в том числе аудиовизуальных) и т. д.

Технологии генеративного ИИ становятся все более доступными и совершенными, поэтому можно прогнозировать их дальнейшее активное освоение фэндомом. При этом «пришествие» нейросетей ставит перед фанатами сложные дилеммы: с одной стороны, фэндом максимально восприимчив к новым цифровым технологиям, позволяющим воплотить фанатское видение любимых историй, миров и персонажей, с другой — принятие ИИ проблематизирует само фанатское творчество, этику и устройство фэндома [Lamerichs, 2018b]. Примечательно, что фэндомы, являясь «браконьерской» альтернативой культурных и креативных индустрий, «отзеркаливают» ряд проблем, возникших в этих индустриях с приходом генеративного ИИ. Обесценивание человеческого труда и угроза потери «рабочих мест», новые формы цифрового неравенства, «плагиат» нейросетей, генерация неэтичного, предвзятого или неточного контента — эти риски, связанные с экспансией генеративного ИИ [Fui-Hoon Nah et al., 2023], в той или иной мере разделяет и фан-комьюнити. При этом у фанатов меньше возможности легитимно оспаривать тревожащие их ИИ-тренды, поскольку фанатские произведения по большей части сами находятся в «серой» правовой зоне [Schwabach, 2011, 2016]. Тем не менее фэндом активно обсуждает достоинства и ограничения нейросетей, демонстрируя поляризацию мнений [Mussies, 2023].

Некоторые крупные фан-платформы разрабатывают свою политику в отношении генеративного ИИ, обсуждая ее с создателями фан-контента. В частности, вики-хостинг *Fandom*<sup>4</sup> (объединяющий свыше 250 тыс. фан-комьюнити) и репозиторий фан-работ *Archive of Our Own / АОЗ*<sup>5</sup> (более 13 млн работ) публично проясняют свое видение взаимодействия с ИИ. Не претендуя на полноценное этнографическое наблюдение, рассмотрим эти два случая более детально. Они тем более показательны, что представляют разные типы фан-платформ и разные стратегии и дискурсы, возникающие вокруг ИИ.

Прежде чем перейти к изложению результатов этого анализа, поясним, почему мы сочли возможным оставить открытым доступ к цитируемым постам и комментариям, притом что этика *fan studies* чувствительна к вопросам конфиденциальности. Основная проблема в случае фанатских пространств связана со слабой дифференциацией приватного и публичного: даже если фан-контент публикуется на открытых интернет-ресурсах, он не предназначен для третьих лиц. Поэтому Кристина Буссе и Карен Хеллексон, в частности, рекомендуют исследователям получать информированное согласие на цитирование любого фанатского текста и/или не использовать в научных статьях прямые URL-адреса фанатских ресурсов [Busse, 2018: 9]. Однако они также признают, что эти требования не всегда выполнимы и целесообразны. Наше решение сохранить ссылки на цитируемый контент опирается на следующее: 1) в обоих случаях обсуждения генеративного ИИ инициированы руководством платформ в публичных постах, излагающих официальную политику этих организаций (ссылки на них как источники информации необходимы); 2) пост директора *Fandom* по общественной безопасности был опубликован на платформе в его блоге, имеющем статус *staff blogs*, — и пользователи оставляли комментарии в этом «официальном» пространстве; 3) в случае АОЗ анализировались комментарии к новостному посту на сайте некоммерческой организации *Organization for Transformative Works (OTW)*, проектом которой, кстати, является не только *Archive of Our Own*, но и академический журнал *Transformative Works and Cultures*. С учетом данных обстоятельств мы рассматриваем эти пространства скорее как публичные места переговоров руководства и пользователей фан-платформ, при этом признавая, что их исследование в определенной мере чувствительно.

## *Fandom*

В декабре 2023 г. один из директоров вики-платформы *Fandom* опубликовал в своем блоге обширный пост под заголовком «*Examining How Fandom Used Generative AI in 2023*»<sup>6</sup>, в котором обозначены философия, текущая практика и перспективы применения ИИ в вики-сообществе. В тексте подчеркивается, что подключение инструментов генеративного ИИ не предполагает замены труда редакторов фан-вики на машинный, статьи по-прежнему будут писаться людьми, и в этом состоит их ценность. В то же время команда *Fandom* отмечает, что

<sup>4</sup> [www.fandom.com](http://www.fandom.com) (дата обращения: 28.03.2024).

<sup>5</sup> [www.archiveofourown.org](http://www.archiveofourown.org) (дата обращения: 28.03.2024).

<sup>6</sup> Quievryn T. Examining How Fandom Used Generative AI in 2023 // *Fandom.com*. 2023. December 14. URL: [https://community.fandom.com/wiki/User\\_blog:TimmyQuivy/Examining\\_How\\_Fandom\\_Used\\_Generative\\_AI\\_in\\_2023#](https://community.fandom.com/wiki/User_blog:TimmyQuivy/Examining_How_Fandom_Used_Generative_AI_in_2023#) (accessed: 20.03.2024).

в течение года отслеживала потребности редакторов вики-страниц и увидела «открытость» или по крайней мере «осторожный интерес» к идее использования некоторых функций ИИ на платформе. Это позволило бы облегчить рутинный труд администраторов многочисленных вики-проектов по ряду направлений. С учетом этого компания определила три области применения нейросетей для *Fandom*: модерация изображений, поисковая оптимизация и предоставление быстрых ответов на вопросы посетителей сайта.

Модерация загружаемых на *Fandom* изображений, ранее осуществлявшаяся вручную, теперь становится автоматической: обучаемый ИИ от компании *CoActive* проверяет каждый объект на потенциальную вредоносность (порнография, язык ненависти, жестокость и т. д.), отклоняя неприемлемый контент. Привлечение ИИ, генерирующего JSON-файлы, делает данные вики-страниц (например, таблицы) более доступными для понимания и сканирования поисковиком *Google*, что в итоге повышает позиции этих страниц в поисковой выдаче. ИИ-функция «быстрые ответы» (на этапе бета-тестирования) переформатирует объемный контент вики-страниц в короткие вопросы и ответы, отображаемые в отдельном модуле, и также улучшает продвижение сайта. Выражая свое стремление развивать опыт использования технологий генеративного ИИ на *Fandom*, топ-менеджмент компании декларирует, что принцип «расширение возможностей, а не замена» (*«enablement, not replacement»*) останется философией платформы в 2024 г.

Несмотря на достаточно умеренную ИИ-политику платформы, ориентир на автоматизацию некоторых человеческих активностей вызвал неоднозначную реакцию со стороны вики-редакторов. Среди более сотни комментариев<sup>7</sup> к этому посту преобладают сдержанные и критические оценки обозначенных компанией перспектив. В первом приближении можно выделить четыре типа аргументов, формирующих критический дискурс в отношении ИИ-инноваций на платформе *Fandom*:

— концептуальный — это антропоцентричная риторика, подчеркивающая, что ИИ противоречит самой природе и идее вики-фэндома, поэтому его присутствие даже в малых дозах нежелательно (*«вся суть фэндома в том, что его создают люди, а не ИИ»; «вики должны редактироваться людьми»*);

— технологический — основан на недоверии к возможностям генеративного ИИ, неудачном опыте его использования (ошибки, дезинформация и т. д.), прогнозе связанных с ним рисков и проблем (*«ИИ — это... психически больной интернет-тролль, если только его не используют для простых задач»*);

— этический — проблематизирует легитимность машинного обучения и ценность продуктов, созданных ИИ (*«ИИ — не что иное, как бездушный хакер, прячущийся за тысячами фрагментов украденного контента и притворяющийся, что он нечто большее, чем есть на самом деле»*);

— менеджериальный — выражает недоверие руководству *Fandom* и опасения, что в погоне за популярными трендами, рейтингами и доходом компания будет *«под видом инноваций»* расширять область использования ИИ *«там, где ему*

<sup>7</sup> Всего было опубликовано 102 комментария. Все цитируемые комментарии находятся на странице одного поста: Quievryn T. Examining How Fandom Used Generative AI in 2023 // *Fandom.com*. 2023. December 14. [https://community.fandom.com/wiki/User\\_blog:TimmyQuivy/Examining\\_How\\_Fandom\\_Used\\_Generative\\_AI\\_in\\_2023](https://community.fandom.com/wiki/User_blog:TimmyQuivy/Examining_How_Fandom_Used_Generative_AI_in_2023) (accessed: 20.03.2024).

не место на сайте» («Я понимаю, что не весь ИИ плох... Но у него есть потенциал, чтобы действительно все испортить... я не хочу давать компании свободу действий в определенных вещах. Дай им дюйм, и они пройдут милю»).

Часть комментаторов вполне лояльно отнеслась к политике внедрения генеративного ИИ в вики-сообществе при условии, что нейросети будут тщательно обучены и/или ИИ-инструменты не станут обязательными в их работе. В целом вики-редакторы хотели бы, чтобы с учетом разброса мнений и всевозможных рисков этот процесс шел осторожно: «это действительно нормально, если фэндом выбирает медленный подход». При этом в обсуждении уже звучит вопрос, разрешено ли использовать ИИ для написания самих статей, то есть создания основного контента вики. Хотя для многих участников сообщества сама его постановка неприемлема, общих правил, запрещающих использование нейросетей для генерации контента, на платформе нет (этот вопрос регулируется на уровне отдельных вики). Таким образом, практики применения генеративного ИИ в вики-сообществах распространяются не только «сверху» (от компаний и платформ), но и «снизу» (от самих фанатов), что является серьезным вызовом для фан-комьюнити.

### Archive of Our Own (AO3)

Если политика платформы *Fandom* направлена на достижение определенного уровня включения технологий генеративного ИИ, то крупнейший архив фан-работ AO3 и связанное с ним сообщество демонстрируют активное сопротивление экспансии нейросетей. Этот протест идет в двух направлениях: против использования фанфиков для машинного обучения и против применения генеративного ИИ для создания самих фан-работ<sup>8</sup>.

В первом случае фанатская организация «*Organization for Transformative Works*» (OTW), проектом которой является AO3, достигла некоторых результатов: фан-сообщество консолидировалось, привлекло общественное внимание к проблеме несогласованного сбора данных, а также предприняло технические меры по защите фан-работ от скрейпинга. О том, что публикуемые на сайте фанфики были включены в датасет *Common Crawl*, используемый для обучения ИИ, организации стало известно в декабре 2022 г.<sup>9</sup> Пользователи обнаружили, что *Sudowrite*, инструмент для творческого письма, основанный на GPT-3 от *OpenAI*, понимает специфическую сексуализированную терминологию фанфиков конкретного жанра<sup>10</sup>. Для большинства авторов AO3 это стало очевидным доказательством того, что генеративный ИИ (точнее, обучающие его компании) «крадет» их работы<sup>11</sup>. Команда сайта установила код, требующий, чтобы *Common Crawl* больше не собирал данные на AO3. Авто-

<sup>8</sup> Хронология борьбы сообщества с генеративным ИИ изложена в статье «AO3 & AI Generated Content» на вики-ресурсе *Fanlore*, который является еще одним проектом Организации трансформативных работ (OTW) наряду с AO3: [https://fanlore.org/wiki/AO3\\_%26\\_AI\\_Generated\\_Content](https://fanlore.org/wiki/AO3_%26_AI_Generated_Content) (accessed: 20.03.2024).

<sup>9</sup> AI and Data Scraping on the Archive // Organization for Transformative Works. 2023. May 13. URL: <https://www.transformativeworks.org/ai-and-data-scraping-on-the-archive/#comments> (accessed: 20.03.2024).

<sup>10</sup> Eveleth R. The Fanfic Sex Trope That Caught a Plundering AI Red-Handed // Wired.com. 2023. May 15. URL: <https://www.wired.com/story/fanfiction-omegaverse-sex-trope-artificial-intelligence-knotting/> (accessed: 20.03.2024).

<sup>11</sup> Это косвенно подтвердило расследование *The Washington Post*: по данным издания, сайт AO3 занимает 516-е место в датасете, а *Wattpad.com* (еще одна фанфикшн-платформа) — 287-е. См.: Schaul K., Chene S. Y., Tiku N. See the websites that make AI bots like ChatGPT sound so smart. *The Washington Post*. 2023. April 19. URL: <https://www.washingtonpost.com/technology/interactive/2023/ai-chatbot-learning/> (accessed: 28.03.2024).

ры фан-работ стали блокировать внешний доступ к своим произведениям. Юристы АОЗ также представили письмо в Бюро авторских прав США с обоснованием позиции, что пользователи должны иметь право отказаться от включения их произведений в обучающие датасеты для ИИ. Кроме того, в июне 2023 г. фэндом провел недельную акцию «*Knot in my Name*», призывая авторов написать как можно больше «подрывных» историй, чтобы «загрязнить» ими обучающие ИИ наборы данных.

Руководство ОТВ и пользователи сайта понимают недостаточную эффективность этих мер, но в целом участники комьюнити ценят усилия организации по защите их работ от посягательств технологических компаний. Однако в той же майской публикации 2023 г., где разъяснялась стратегия борьбы со скрейпингом данных Архива, была изложена и текущая политика ОТВ в отношении фан-работ, сгенерированных ИИ. И эта позиция вызвала волну критических комментариев<sup>12</sup>. В частности, здесь обозначено, что на АОЗ нет запрета на размещение фан-работ, полностью или частично созданных с помощью ИИ. Организация подчеркивает, что ее цель — «максимальная инклюзивность фан-работ», и если фанаты используют ИИ, то «это работы, которые мы также стремимся сохранить». Как и редакторы платформы *Fandom*, пользователи АОЗ проблематизировали в своих комментариях концептуальные, этические, технологические и менеджериальные аспекты политики ОТВ в отношении ИИ.

Многие фанаты выражали категорическое несогласие с тем, что тексты, сгенерированные ИИ, можно считать фанатскими «трансформативными работами», поскольку «они были созданы бездумными машинами, а не фанатами», «программа не может быть „фанатом“ чего-либо». Признание таких произведений как фанатских, по их мнению, разрушает ценность человеческого творчества и саму концепцию АОЗ («Архив НАШ, а не ИИ»). Фанаты также подчеркивали противоречие в заявлении руководства ОТВ: пытаясь оградить продукцию сообщества от «неэтичного ИИ», оно поддерживает работы, произведенные этим же ИИ («как кто-то мог вообще с невозмутимым видом поместить обе части этого обновления в один пост»). Эксплуататорский характер генеративного ИИ (и его компаний-разработчиков) обосновывается с двух сторон: речь не только о «краже» или «плагиате» некоммерческих фан-работ для получения дальнейшей прибыли корпорациями, но и о сомнительной практике компаний по обучению ИИ посредством скрытого человеческого труда, в частности дешевой рабочей силы в странах Глобального Юга. Комментаторы упоминали расследование *Time* о том, как *OpenAI* использовала труд кенийских работников за два доллара в час, чтобы сделать *ChatGPT* менее токсичным (они были вынуждены читать и маркировать большой объем контента с экстремальными описаниями насилия и жестокости)<sup>13</sup>. Акцентировалась и неэкологичность технологий искусственного интеллекта: «Вычислительная мощность требует много электроэнергии и воды».

Кроме того, пользователи настоятельно просили руководство ОТВ пересмотреть и четко выразить позицию, не одобряющую публикацию сгенерированных

<sup>12</sup> Далее обобщаются и цитируются комментарии (всего их 197) к публикации «*AI and Data Scraping on the Archive*» (2023) на сайте ОТВ. Этот пост был также размещен на сайте АОЗ, где получил 862 комментария.

<sup>13</sup> Perrigo B. Exclusive: OpenAI Used Kenyan Workers on Less Than \$ 2 Per Hour to Make ChatGPT Less Toxic // *Time.com*. 2023. January 18. URL: <https://time.com/6247678/openai-chatgpt-kenya-workers/%20A03/> (accessed: 28.03.2024).

ИИ фанфиков, пусть даже технически предотвратит их появление на сайте невозможно. Авторам таких работ предлагалось тегировать контент как целиком или частично созданный ИИ. Наряду с признательностью за многолетнюю поддержку и защиту фан-работ, многие участники комьюнити выразили недоверие и разочарование политикой *OTW*, призвав ее быть более искренней, прозрачной и решительной. В продолжение этих обсуждений в июле 2023 г. на сайте АОЗ и в социальных сетях *OTW* появился коллективный манифест<sup>14</sup> (первые два раздела), приглашающий присоединиться к движению *#FandomVsAI*. Его авторы озвучили не только протест против использования фан-работ для обучения генеративного ИИ, но и свое видение того, каким должен быть фэндом, подчеркнув, что «ИИ нет в нем места». В то же время АОЗ оставил возможность размещения сгенерированных нейросетями работ с пометкой «*AI-Generated Text*» (приняв также во внимание аргументы пользователей о том, что в настоящее время нет надежных способов различить машинный и человеческий текст). В документе акцентируется некоммерческий характер фэндома как «одного из последних бастионов свободного творческого самовыражения без мотива получения прибыли». Сопротивление ИИ здесь вписано в более широкий политический и идеологический контекст: «мы боремся за свободу, которую дает нам фэндом, чтобы творить за пределами границ капитализма». Таким образом, фан-сообщество АОЗ позиционирует себя как «контрпублика», отстаивающая свою агентность в эпоху цифровизации и автоматизации.

## Заключение

Технологии искусственного интеллекта стремительно захватывают и преобразуют сферы культурного производства и потребления. Фан-культура, будучи одной из форм культуры соучастия, сочетает в себе оба измерения: фанаты всегда больше, чем пассивные реципиенты, они просьюмеры, создающие огромный объем разнообразного любительского контента. Неудивительно, что инновации и проблемы, привнесенные с ИИ в культурные и креативные индустрии, во многом актуальны и для медиафэндомов. Фанаты в целом демонстрируют высокую восприимчивость к медиа как инструментам производства своей продукции, каналам ее распространения и местам взаимодействий. Однако в случае с искусственным интеллектом, и в частности глубокими генеративными моделями, фэндом оказался поляризован: в то время как одни активно осваивают и используют нейросети для редактирования и создания контента, другие полагают, что подобные практики неуместны.

В статье мы в большей мере сосредоточились на критической реакции фэндома, упуская позитивные дискурсы (как фанатские, так и академические), возникающие вокруг генеративного ИИ. Между тем нужно отметить, что нейросети демократизируют творческий процесс, позволяя людям, не обладающим хорошими креативными навыками или имеющим особые потребности, создавать изображения, истории и т. д. Как выразилась Мартина Муссис, ИИ реализует фанатскую мечту, в которую каждый может внести свой вклад [Mussies, 2023]. По мнению ис-

<sup>14</sup> Artificial: Reconciling Fandom with Emerging Technology. *Archiveofourown.org*. 2023. July 14. URL: <http://archiveofourown.org/works/48565666> (accessed: 20.03.2024).

следовательницы (и художницы), вопросы о том, является ли синтетическое произведение произведением искусства, кто его автор и т.д., для фэндома нерелевантны. Не имеет значения, кто и как создал (например, фанарт), — важно лишь то, что работа циркулирует в сообществе и участвует в коллективном создании неканонического фикционального мира.

Что касается критических и протестных фан-дискурсов о генеративном ИИ, по нашим предварительным наблюдениям, они включают три ключевых высказывания:

а) использование ИИ-технологий противоречит самой идее фэндома (как сообщества людей) и фанатского творчества;

б) сегодняшний ИИ недостаточно совершенен, чтобы заменить человеческие труд и мастерство;

в) текущие практики обучения и внедрения генеративного ИИ эксплуататорские и незитичные.

Многие фан-сообщества и платформы разрабатывают свою политику использования или сопротивления ИИ, вплоть до запрета публиковать синтетические фан-работы. Однако обеспечить действие таких ограничений технически сложно, поскольку нет программ, достоверно различающих машинные и человеческие медиатексты.

Мы не затронули ряд вопросов, которые в перспективе требуют отдельного изучения:

— как конструируется агентность генеративного ИИ и кто вообще ей наделяется в представлении фанатов (сам ИИ, компания-разработчик, пользователь)?

— как фэндом пересобирает свое понимание креативности, фанатского труда и фандомной этики в эпоху культурного ИИ?

— воспринимаются ли машинные фан-работы иначе, чем человеческие (меняются ли практики фанатского чтения, эстетические переживания и т.д.)?

— как выстраиваются отношения между фанатами и поп-культурными ИИ-объектами (в частности, виртуальными селебритами)?

— какие новые формы цифрового неравенства могут возникать и становиться предметом рефлексии внутри фан-культуры?

— как национальная и региональная специфика медиафэндомов влияет на практики использования генеративного ИИ (прежде всего, в неанглоязычных странах)?

Эти и другие направления исследований актуальны для последующего развития как *fan studies*, так и социокультурного анализа искусственного интеллекта. Вероятно, процесс гибридизации человеческого и машинного продолжится, и в будущем ИИ будет играть все более заметную роль в культурах соучастия.

## Список литературы (References):

1. Грушевская В. Модель фильтрации информации в социальных медиа // Журнал исследований социальной политики. 2022. Т. 20. № 3. С. 393—406. <https://doi.org/10.17323/727-0634-2022-20-3-393-406>.  
Grushevskaya V. (2022) The Information Filtering Model in Social Media. *The Journal of Social Policy Studies*. Vol. 20. No. 3. P. 393—406. <https://doi.org/10.17323/727-0634-2022-20-3-393-406>. (In Russ.)

2. Дегтяренко К. А., Пчелкина Д. С., Шпак А. А., Пименова Н. Н. Образ искусственного интеллекта в кинематографе: трансформации в период 1980—2010-х годов // Журнал Сибирского федерального университета. Гуманитарные науки. 2023. Т. 16. № 8. С. 1454—1470. <https://elib.sfu-kras.ru/handle/2311/150860>.  
Degtyarenko K. A., Pchelkina D. S., Shpak A. A., Pimenova N. N. (2023) The Image of Artificial Intelligence in Cinema: Transformations in the 1980s–2010s. *Journal of Siberian Federal University. Humanities & Social Sciences*. Vol. 16. No. 8. P. 1454—1470. <https://elib.sfu-kras.ru/handle/2311/150860>. (In Russ.)
3. Левашов В. К., Гребняк О. В. Россияне о вызовах искусственного интеллекта, нейронных сетей и социальном оптимизме // Социологические исследования. 2023. № 11. С. 115—120. <https://doi.org/10.31857/S013216250028537-3>.  
Levashov V. K., Grebnyak O. V. (2023) Russian Citizens on the Challenges of Artificial Intelligence, Neural Networks and Social Optimism. *Sociological Studies*. No. 11. P. 115—120. <https://doi.org/10.31857/S013216250028537-3>. (In Russ.)
4. Лещинская Н. М., Колесник М. А., Омелик А. А., Ермаков Т. К. Образ искусственного интеллекта в сериале «Черное зеркало» (2011—2019) // Журнал Сибирского федерального университета. Гуманитарные науки. 2023. Т. 16. № 8. С. 1331—1342. <https://elib.sfu-kras.ru/handle/2311/150851>.  
Leshchinskaja N. M., Kolesnik M. A., Omelik A. A., Ermakov T. K. (2023) The Image of Artificial Intelligence in the Series «Black Mirror» (2011—2019). *Journal of Siberian Federal University. Humanities & Social Sciences*. Vol. 16. No. 8. P. 1331—1342. <https://elib.sfu-kras.ru/handle/2311/150851>. (In Russ.)
5. Резаев А. В., Трегубова Н. Д. «Искусственный интеллект», «онлайн-культура», «искусственная социальность»: определение понятий // Мониторинг общественного мнения: экономические и социальные перемены. 2019. № 6. С. 35—47. <https://doi.org/10.14515/monitoring.2019.6.03>.  
Rezaev A. V., Tregubova N. D. (2019) Artificial Intelligence, On-line Culture, Artificial Sociality: Definition of the Terms. *Monitoring of Public Opinion: Economic and Social Changes*. No. 6. P. 35—47. <https://doi.org/10.14515/monitoring.2019.6.03>. (In Russ.)
6. Andrews P. C. (2022) 'Are Di Would of Loved It': Reanimating Princess Diana Through Dolls and AI. *Celebrity Studies*. Vol. 13. No. 4. P. 573—594. <https://doi.org/10.1080/19392397.2022.2135087>.
7. Barron L. (2023) AI and Popular Culture. Leeds: Emerald Publishing Limited. <https://doi.org/10.1108/9781803823270>
8. Booth P. (2017) Digital Fandom 2.0: New Media Studies (Second Edition). New York, NY: Peter Lang. <https://doi.org/10.3726/978-1-4539-1654-4>.
9. Bredikhina L., Kameoka T., Shimbo S., Shirai A. (2020) Avatar Driven VR Society Trends in Japan. In: *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. Atlanta, GA: IEEE. P. 497—503. <https://doi.org/10.1109/VRW50115.2020.00103>.

10. Burden D., Savin-Baden M. (2019) *Virtual Humans: Today and Tomorrow*. New York, NY: Chapman and Hall/CRC; Taylor & Francis Group. <https://doi.org/10.1201/9781315151199>.
11. Busse K. (2018) The Ethics of Studying Online Fandom. In: Scott S., Click M. A. (eds.) *The Routledge Companion to Media Fandom*. Routledge, Taylor & Francis Group. P. 9—18.
12. Busse K., Hellekson K. (2006) Introduction: Work in Progress. In: Busse K., Hellekson K. (eds.) *Fan Fiction and Fan Communities in the Age of the Internet*. Jefferson, NC: McFarland. P. 5—32.
13. Byun K. J., Ahn S. J. (Grace). (2023) A Systematic Review of Virtual Influencers: Similarities and Differences Between Human and Virtual Influencers in Interactive Advertising. *Journal of Interactive Advertising*. Vol. 23. No. 4. P. 293—306. <https://doi.org/10.1080/15252019.2023.2236102>.
14. Chew N. (2018) Tumblr as Counterpublic Space for Fan Mobilization. *Transformative Works and Cultures*. Vol. 27. <https://doi.org/10.3983/twc.2018.1186>
15. Chow P. S. (2020) Ghost in the (Hollywood) Machine: Emergent Applications of Artificial Intelligence in the Film Industry. *NECSUS: European Journal of Media Studies*. Vol. 9. No. 1. P. 193—214. <https://doi.org/10.25969/mediarep/14307>.
16. Coppa F. A (2006) Brief History of Media Fandom In: Busse K., Hellekson K. (eds.) *Fan Fiction and Fan Communities in the Age of the Internet*. Jefferson, NC: McFarland. P. 41—59.
17. Danesi M. (2024) *AI-Generated Popular Culture: A Semiotic Perspective*. Cham: Palgrave Macmillan. <https://doi.org/10.1007/978-3-031-54752-2>.
18. Datta A., Goswami R. (2021) The Film Industry Leaps Into Artificial Intelligence: Scope and Challenges by the Filmmakers. In: Rathore V. S. et. al (eds.) *Rising Threats in Expert Applications and Solutions*. Vol. 1187. P. 665—670. Singapore: Springer. [https://doi.org/10.1007/978-981-15-6014-9\\_80](https://doi.org/10.1007/978-981-15-6014-9_80).
19. Dean J., Andrews P. (2021) Celebritization from Below: Celebrity, Fandom, and Anti-Fandom in British Politics. *New Political Science*. Vol. 43. No. 3. P. 320—338. <https://doi.org/10.1080/07393148.2021.1957602>.
20. Duffett M. (2013) *Understanding Fandom: An introduction to the Study of Media Fan Culture*. N.Y.: Bloomsbury.
21. Edwards A. (2018) Literature Fandom and Literary Fans. In: Booth P. (ed.) *A Companion to Media Fandom and Fan Studies*. Hoboken: John Wiley and Sons, Inc. P. 47—64. <https://doi.org/10.1002/9781119237211.ch3>.
22. Elliott A. (2019) *The Culture of AI: Everyday Life and the Digital Revolution*. London: Routledge. <https://doi.org/10.4324/9781315387185>.

23. Frey M. (2021) *Netflix Recommends: Algorithms, Film Choice, and the History of Taste*. Berkeley: University of California Press. <https://doi.org/10.1525/9780520382022>.
24. Fui-Hoon Nah F., Zheng R., Ca J., Siau K., Chen L. (2023) Generative AI and ChatGPT: Applications, Challenges, and AI-Human Collaboration. *Journal of Information Technology Case and Application Research*. Vol. 25. No. 3. P. 277—304. <https://doi.org/10.1080/15228053.2023.2233814>.
25. Gray J., Sandvoss C., Harrington C. L. (eds.) (2017) *Fandom: Identities and Communities in a Mediated World (Second Edition)*. New York, NY: New York University Press. <https://doi.org/10.18574/nyu/9781479845453.001.0001>.
26. Han S. (2021) AI, Culture Industries and Entertainment. In: Elliott A. (ed.) *The Routledge Social Science Handbook of AI*. London: Routledge. P. 295—312.
27. Hellekson K. (2018) The Fan Experience. In: Booth P. (ed.) *A Companion to Media Fandom and Fan Studies*. Hoboken: John Wiley and Sons, Inc. P. 65—76.
28. Hepp A. (2020) *Deep Mediatization*. London: Routledge. <https://doi.org/10.4324/9781351064903>.
29. Hermann I. (2023) Artificial Intelligence in Fiction: Between Narratives and Metaphors. *AI & Society*. Vol. 38. No. 1. P. 319—329. <https://doi.org/10.1007/s00146-021-01299-6>
30. Jenkins H. (2005) *Textual Poachers: Television Fans and Participatory Culture*. New York, NY.; London: Routledge.
31. Jenkins H. (2018) Fandom, Negotiation, and Participatory Culture. In: Booth P. (ed.) *A Companion to Media Fandom and Fan Studies*. Hoboken: John Wiley and Sons, Inc. P. 13—26.
32. Jin D. Y. (2021) Artificial Intelligence in Cultural Production: Critical Perspectives on Digital Platforms. London: Routledge. <https://doi.org/10.4324/9781003164258>.
33. Kalpokas I. (2021) Problematising Reality: The Promises and Perils of Synthetic Media. *SN Social Sciences*, Vol. 1. No. 1. <https://doi.org/10.1007/s43545-020-00010-8>.
34. Kang E. J., Kim H., Kim H., Kim J. (2022) ‘When AI Meets the K-Pop Culture: A Case Study of Fans’ Perception of AI Private Call’. In: *NeurIPS 2022 Workshop*. [https://ai-cultures.github.io/papers/when\\_ai\\_meets\\_the\\_k\\_pop\\_cultur.pdf](https://ai-cultures.github.io/papers/when_ai_meets_the_k_pop_cultur.pdf) (accessed: 28.03.2024).
35. Lamerichs N. (2018a) *Productive Fandom: Intermediality and Affective Reception in Fan Cultures*. Amsterdam: Amsterdam University Press.
36. Lamerichs N. (2018b) The Next Wave in Participatory Culture: Mixing Human and Nonhuman Entities in Creative Practices and Fandom. *Transformative Works and Cultures*. Vol. 28. <https://doi.org/10.3983/twc.2018.1501>.

37. Linden H., Linden S. (2017) *Fans and Fan Cultures*. London: Palgrave Macmillan. <https://doi.org/10.1057/978-1-137-50129-5>.
38. Liu J. (2023) Virtual Presence, Real Connections: Exploring the Role of Parasocial Relationships in Virtual Idol Fan Community Participation. *Global Media and China*. <https://doi.org/10.1177/20594364231222976>.
39. McCracken A. (2018). A History of Fandom in Broadcasting. In: Bodroghkozy A. (ed.) *A Companion to the History of American Broadcasting*. Hoboken, NJ: John Wiley & Sons, Inc. P. 413—442. <https://doi.org/10.1002/9781118646151.ch19>.
40. Manovich L. (2019) *AI Aesthetics*. Moscow: Strelka Press.
41. Marshall A. (2020) Our Stories, Our Selves: Star Wars Fanfictions as Feminist Counterpublic Discourses in Digital Imaginaria. *Journal of Fandom Studies*. Vol. 8. No. 3. P. 277—288. [https://doi.org/10.1386/jfs\\_00024\\_1](https://doi.org/10.1386/jfs_00024_1).
42. Mussies M. (2023) Artificial Intelligence and the Production of Fan Art. *Transformative Works and Cultures*. Vol. 40. <https://doi.org/10.3983/twc.2023.2431>.
43. Nader K., Toprac P., Scott S., Baker S. (2022) Public Understanding of Artificial Intelligence Through Entertainment Media. *AI & Society*. Vol. 39. P. 713—726. <https://doi.org/10.1007/s00146-022-01427-w>.
44. Nuranda D. R., Purwaningtyas M. P. F. (2023) Between Human and Virtual Idol: Fans Reception Analysis of Digital Alter Ego Concept. *Komunikator*. Vol. 15. No. 1. P. 31—40. <https://doi.org/10.18196/jkm.17235>.
45. Popova M. (2020) Reading out of Context: Pornographic Deepfakes, Celebrity and Intimacy. *Porn Studies*. Vol. 7. No. 4. P. 367—381. <https://doi.org/10.1080/23268743.2019.1675090>.
46. Schwabach A. (2011) *Fan Fiction and Copyright: Outsider Works and Intellectual Property Protection (1<sup>st</sup> ed.)*. London: Routledge. <https://doi.org/10.4324/9781315581965>.
47. Schwabach A. (2016) Fan Works and the Law. In: Gelder K. (ed.) *New Directions in Popular Fiction*. London: Palgrave Macmillan. P. 405—424. [https://doi.org/10.1057/978-1-137-52346-4\\_20](https://doi.org/10.1057/978-1-137-52346-4_20).
48. Stanfill M. (2019) *Exploiting Fandom: How the Media Industry Seeks to Manipulate Fans*. Iowa: University of Iowa Press.
49. Yin Y. (2020) An Emergent Algorithmic Culture: The Data-ization of Online Fandom in China. *International Journal of Cultural Studies*. Vol. 23. No. 4. P. 475—492. <https://doi.org/10.1177/1367877920908269>.
50. Zwaan K., Duits L., Belinders S. (2014) Introduction. In: Duits L., Zwaan K., Belinders S. (eds.) *The Ashgate Research Companion to Fan Cultures*. Farnham: Ashgate. P. 1—6.

DOI: [10.14515/monitoring.2024.5.2550](https://doi.org/10.14515/monitoring.2024.5.2550)**И. И. Буханский**

**ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ — ЭТО ТВОЕ.  
РЕЦ. НА КН.: ХОВАРД ДЖ., ГУГГЕР С. ГЛУБОКОЕ ОБУЧЕНИЕ  
С FASTAI И PYTORCH: МИНИМУМ ФОРМУЛ,  
МИНИМУМ КОДА, МАКСИМУМ ЭФФЕКТИВНОСТИ.  
СПБ.: ПИТЕР, 2022**

**Правильная ссылка на статью:**

Буханский И. И. Искусственный интеллект — это твое // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 203—217. <https://doi.org/10.14515/monitoring.2024.5.2550>. Рец. на кн.: Ховард Дж., Гуггер С. Глубокое обучение с fastai и PyTorch: минимум формул, минимум кода, максимум эффективности. СПб.: Питер, 2022.

**For citation:**

Bukhansky I. I. (2024) Artificial Intelligence Is What You Can Do. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 203–217. <https://doi.org/10.14515/monitoring.2024.5.2550>. Book Review: Howard J., Guggen S. (2022) Deep Learning with fastai and PyEorch: Minimum Formulas, Minimum Code, Maximum Efficiency. St. Petersburg: Piter. (In Russ.)

Получено: 28.12.2023. Принято к публикации: 20.06.2024.

ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ — ЭТО ТВОЕ РЕЦ. НА КН.: ХОВАРД ДЖ., ГУГГЕР С. ГЛУБОКОЕ ОБУЧЕНИЕ С FASTAI И PYTORCH: МИНИМУМ ФОРМУЛ, МИНИМУМ КОДА, МАКСИМУМ ЭФФЕКТИВНОСТИ. СПБ.: ПИТЕР, 2022

*БУХАНСКИЙ Илья Игоревич — студент 4 курса бакалавриата направления «Социология» Высшей школы современных социальных наук, МГУ им. М. В. Ломоносова, Москва, Россия; исследовательский менеджер методлаборатории, ВЦИОМ, Москва, Россия  
E-MAIL: buhanskiyij@my.msu.ru  
<https://orcid.org/0009-0004-7051-4960>*

**Аннотация.** Книга разработчика и бизнес-аналитика Джереми Ховарда и инженера-исследователя в области науки о данных Сильвейна Гуггера, вышедшая в 2020 г. и переведенная на русский язык в 2022 г., представляет собой попытку доказать читателю, что он, вне зависимости от образования и профессии, способен самостоятельно обращаться к одной из знаковых технологий XXI века — искусственному интеллекту (ИИ) и создавать продукты с его использованием. В работе проведен тщательный анализ тенденций ИИ, рассмотрен ряд технологий, необходимых для эффективной деятельности в данной сфере; описаны этапы разработки проекта с ИИ от постановки проблемы и сбора данных до создания графического интерфейса и размещения приложения на сервере; раскрыты наиболее острые проблемы, связанные со все более широким применением ИИ в компаниях; актуализирован тезис о необходимости «этического» подхода при работе с ИИ. Основная часть книги посвящена объяснению программ, реализующих тот или иной механизм ИИ (точнее, ту или иную архитектуру нейросетей), этапов и принципов написания кода, посредством которого решается одна из заранее заданных практических про-

ARTIFICIAL INTELLIGENCE IS WHAT YOU CAN DO. BOOK REVIEW: HOWARD J., GUGGER S. (2022) DEEP LEARNING WITH FASTAI AND PYTORCH: MINIMUM FORMULAS, MINIMUM CODE, MAXIMUM EFFICIENCY. ST. PETERSBURG: PITER

*Iliya I. BUKHANSKY<sup>1,2</sup> — Bachelor Student in Sociology; Research Manager  
E-MAIL: buhanskiyij@my.msu.ru  
<https://orcid.org/0009-0004-7051-4960>*

<sup>1</sup> Moscow State University, Moscow, Russia

<sup>2</sup> Russian Public Opinion Research Center (VCIOM), Moscow, Russia

**Abstract.** The book by developer and business analyst Jeremy Howard and data science research engineer Sylvain Gugger, published in 2020 and translated into Russian in 2022, is an attempt to prove to the reader that he or she, regardless of education and profession, is capable of independently accessing one of the iconic technologies of the 21st century — artificial intelligence (AI) — and creating products using it to incorporate the latter into their activities. The work provides a thorough analysis of AI trends, reviews the technology stack necessary for effective work in this field, and describes the stages of developing a project with artificial intelligence, from problem formulation and data collection to creating a graphical user interface and placing the application on the server, reveals the most acute problems associated with the increasing use of AI in companies, and updates the thesis on the need for an “ethical” approach when working with artificial intelligence. The book’s central part is devoted to explaining programs implementing one AI mechanism or another (or rather, one neural network architecture or another), as well as the stages and principles of writing a code using which one of the predetermined practical problems is solved. However, for the reader who does not want to immerse him-

блем. Однако для читателя, не желающего погружаться в программирование, работа предлагает материалы о теории ИИ, о методах получения максимальной полезности от ИИ в случае, если вы заказчик, а не исполнитель проекта, о правилах предупреждения негативных последствий включения ИИ в деятельность фирмы, следование которым требуется не только от разработчиков, но и от всей команды, работающей над продуктом, включая директорат. Таким образом, рецензируемая книга может быть полезна для представителей неограниченного круга профессий, должностей и интересов.

**Ключевые слова:** искусственный интеллект, глубокое обучение, нейронные сети, обработка данных, моделирование, интерпретируемость моделей, этика искусственного интеллекта

self in programming, the work offers materials about the theory of artificial intelligence, the methods of obtaining maximum usefulness from AI in case you are a customer and not a project executor, about the rules of preventing negative consequences of including artificial intelligence in the firm's activity, following which is required not only from developers but also from the whole team working on the product, including the directors. Thus, the reviewed book may be helpful for representatives of an unlimited range of professions, positions, and interests.

**Keywords:** artificial intelligence, deep learning, neural networks, data processing, modelling, interpretability of models, ethics of artificial intelligence

Искусственный интеллект (ИИ) в последние несколько лет развивается с колоссальной скоростью: областей применения основных его продуктов — моделей глубокого обучения (*deep learning*, *DL*) — становится все больше, качество работы ИИ достигает уровня человеческих способностей (а в ряде случаев и превышает его)<sup>12</sup>. Чтобы не отставать от современных разработок, необходимо искать пути интеграции этого инструмента в собственную деятельность. Социология не исключение: в рамках работы социального исследователя существует множество задач, например сбор и обработка данных, предсказательное моделирование, в которых использование ИИ стало устоявшейся практикой либо находится на стадии активного освоения [Котов, 2023]. Ограничением выступает лишь представление о чрезмерной сложности для «гуманитария» научиться взаимодействию с ИИ, программированию, работе с кодом. Существование этого представления признается современными социологами, однако с ним пытаются бороться. Одним из путей противодействия сложившемуся стереотипу выступают императивные указания на ненаучный характер социологического знания в случае, если его авторы посвящают себя построению «литературных теорий», забывая о программировании как необходимом инструменте исследователя в современном инфор-

<sup>1</sup> Раевский Р. Искусственный интеллект научился читать лучше человека // Forbes. 2018. 20 января. URL: <https://www.forbes.ru/tehnologii/355971-iskusstvennyy-intellekt-nauchilsya-chitat-luchshe-cheloveka> (дата обращения: 20.06.2024).

<sup>2</sup> На что способен искусственный интеллект сегодня и каков его потенциал // РБК Тренды. Индустрия 4.0. 2021. 19 ноября. URL: <https://trends.rbc.ru/trends/industry/cmrm/619766d59a79471862e77e8a> (дата обращения: 20.06.2024).

мационном обществе<sup>34</sup>. Другой способ менее категоричен: его сторонники, обращаясь к идее о «компьютерном мышлении» как «базовом навыке» исследователя сегодня, особое внимание акцентируют на разных уровнях этого навыка, говоря в том числе о требуемом для взаимодействия с миром технологий минимуме, доступном для освоения каждому<sup>5</sup>.

Развенчание этого представления — миссия рецензируемой книги, в аннотации которой заявлено: не будучи «доктором математических наук» или «ботаном», после прочтения вы сможете разворачивать «самые модные и актуальные приложения» с применением ИИ [Ховард, Гуггер, 2022: 4]. Чтобы вызвать доверие к подобному высказыванию, следует указать на детали биографии его авторов, убеждающие в профессионализме и признанности в научно-техническом сообществе будущих «наставников» читателя. Джереми Ховард — исследователь-основатель компании fast.ai, почетный профессор Школы информационных технологий и электротехники Квинслендского университета в Австралии. Ховард является автором одной из наиболее влиятельных разработок в области нейросетевого моделирования естественного языка — алгоритма ULMFiT для обучения больших языковых моделей (LLM), на котором основаны все современные LLM, включая ChatGPT. Помимо ULMFiT, вклад Ховарда в развитие глубокого обучения включает в себя создание множества инструментов и программных библиотек, нацеленных на ускорение развертывания различного рода цифровых продуктов, среди которых программная библиотека fastai — один из наиболее популярных в мире высокоуровневых фреймворков для глубокого обучения с открытым исходным кодом. О значимости Ховарда в сообществе исследователей нейронных сетей говорит и его карьерный трек: он занимал позицию президента и главного научного сотрудника *Kaggle*, крупнейшего в мире сообщества специалистов по data science, основал *Enlitic* — одну из первых в мире компаний, специализирующихся на глубоком обучении для медицины, в течение двух лет после создания удерживавшую 14-е место в списке «самых умных компаний мира» по версии *MIT Tech Review*. На сегодняшний день Ховард является председателем-основателем подразделения *The AI & Medicine Initiative (AIM)* в Университете Сан-Франциско — организации, действующей в партнерстве с Гарвардом, Стэнфордом и иными академическими медицинскими центрами.

Сильвейн Гуггер — выпускник кафедры фундаментальной и прикладной математики Высшей нормальной школы в Париже, магистр точных наук университета Париж-Дофин, преподаватель математики и информатики. Профессиональная траектория Гуггера как педагога включает в себя преподавание общей математики и введения в программирование на курсе по подготовке к поступлению в инженерные и бизнес-вузы CPGE — дополнительной школы интенсивной подготовки выпускников школ к поступлению в престижные высшие учебные заведения

<sup>3</sup> Давыдов А. А. Системная социология: языки программирования // Институт социологии ФНИСЦ РАН. Блоги. Блог А. А. Давыдова. 2009. 22 мая. URL: [https://www.isras.ru/index.php?page\\_id=1075](https://www.isras.ru/index.php?page_id=1075) (дата обращения: 20.06.2024).

<sup>4</sup> Дополнительно о трансформации социологии, ее методологии и методов вследствие развития искусственного интеллекта см. [Резаев, Трегубова, 2021; Семина, Го, 2022].

<sup>5</sup> Евдокимова К. Социолог СПбГУ Наталья Трегубова: «Сейчас компьютерное мышление нужно, по сути, всем, а не только программистам или сотрудникам ИТ» // Пять углов. 2023. 5 декабря. URL: [https://5uglov.ru/post/19320\\_sociolog\\_spbgu\\_natalya\\_tregubova\\_seychas\\_komputernoe\\_mishlenie\\_nuzhno\\_po\\_suti\\_vsem\\_a\\_ne\\_tolko\\_programmistam\\_ili\\_sotrudnikam\\_it](https://5uglov.ru/post/19320_sociolog_spbgu_natalya_tregubova_seychas_komputernoe_mishlenie_nuzhno_po_suti_vsem_a_ne_tolko_programmistam_ili_sotrudnikam_it) (дата обращения: 20.06.2024).

Франции с одной из самых высоких рабочих нагрузок в Европе; по указанному курсу под авторством Гуггера составлены нескольких учебников. Помимо этого, он был ученым-исследователем в fast.ai, где работал над совершенствованием техник ускоренного обучения моделей ИИ в условиях ограниченности ресурсов. На текущий момент Сильвейн занимает позицию инженера-исследователя в HuggingFace — американской компании, ведущей деятельность в области разработки программного обеспечения с использованием машинного обучения; одно из наиболее значимых достижений HuggingFace — собственная платформа, позволяющая разработчикам в сфере ИИ свободно обмениваться обученными моделями; компания также поддерживает программную библиотеку Transformers, обеспечивающую всем желающим доступ к инструментам искусственного интеллекта различных модальностей.

Книга представляет собой практическое пособие, соответствующее запросу социологического сообщества на развитие социогуманитарного знания в рамках цифровизации. Работа состоит из введения и четырех частей, построенных по принципу перехода от высокоуровневого взаимодействия с искусственным интеллектом к более низкоуровневому, и двух приложений. Часть I, «Глубокое обучение на практике», содержит три главы, вводящие читателя в теорию нейросетей, ограничения области применения технологий ИИ, техническую часть создания нейронных сетей. Здесь предлагается руководство по развертыванию наиболее простых инструментов ИИ и обсуждаются проблемы (в том числе этические), возникающие при переходе от моделирования в рамках той или иной задачи до использования полученной модели на практике. Часть II, «Понимание приложений на базе FastAI», состоит из восьми глав, описывающих работу с данными и обучение типовых моделей (т. е. моделей с заранее заданной архитектурой или предопределенной последовательностью слоев (базовых структурных элементов DL-модели), внесение изменений в устройство и порядок которых со стороны разработчика не предполагается) для задач компьютерного зрения, табличного моделирования, обработки естественного языка с помощью готовых структур fastai. Часть III, «Основы глубокого обучения», представляет собой пять глав, посвященных самостоятельному созданию архитектур моделей искусственного интеллекта в сферах компьютерного зрения, табличного моделирования и обработки естественного языка при помощи примитивных слоев фреймворка PyTorch, одного из наиболее популярных в сфере DL, а также написанию цикла обучения для полученных моделей (иначе говоря, в данной части читатель учится подбирать и упорядочивать слои нейронной сети для создания уникальных архитектур, в отличие от части II, где структурные элементы модели уже отобраны и расположены тем или иным образом, вследствие чего разработчик может лишь обучать и использовать предоставленные ему готовые архитектуры). Часть IV, «Глубокое обучение с чистого листа», содержит четыре главы, посвященные повествованию и рассказывающие о создании с нуля простейших слоев моделей искусственного интеллекта и основных структур, используемых в частях I, II и III.

Введение формирует образ ИИ как одной из лидирующих технологий XXI века, динамика развития и возможности которой все возрастают благодаря ее вклю-

чению в неограниченный круг экспертных областей<sup>6</sup>. Для продуктивного применения DL в своем предметном поле есть лишь два обязательных требования: наличие здравого смысла и упорства. Для скорейшего освоения материала желательно иметь базовые навыки программирования и знания математики уровня средней школы, однако книга построена так, что их можно приобрести (или освежить) по мере чтения, обращаясь при необходимости к дополнительным материалам [там же: 22—23].

Авторы обещают научить читателя работать с современными моделями DL, способными выполнять «распознавание, локализацию изображений и обнаружение; классификацию документов и языковое моделирование; обработку табличных данных с категориальными, непрерывными и смешанными данными, включая временной ряд; совместную фильтрацию (рекомендации)» [там же: 23]. Знание о применении ИИ не исчерпывается техническим аспектом: Ховард и Гуггер затронут в книге множество не менее важных, чем написание кода, теоретических и философских сторон DL [там же: 23—24].

В главе 1 авторы предпринимают попытку доказать два ключевых тезиса книги: 1) работа с DL доступна для всех — не требуется ученой степени по математике, десятилетнего опыта в написании кода, огромных массивов данных и излишне дорогого оборудования [там же: 28]; 2) продуктивное освоение DL возможно лишь на практике, поэтому структура книги подчинена правилу демонстрации решения конкретной проблемы и его последующего подробного рассмотрения, чтобы читатель смог осознать ход мысли и реализовать его уже в своем коде [там же: 35—36]. Первое положение подтверждается историей авторов: будучи специалистами в сфере машинного обучения (*machine learning*, ML), Джереми Ховард (несмотря на прочную связь с техническими науками) — бакалавр философии, специалист по данным, но без формального технического образования, то есть «не математик», а Сильвейн Гуггер — академик в области точных наук, автор нескольких учебников по математике, однако «не программист» [там же: 33—35]. Второй тезис звучит в книге рефреном: совершенство в DL достигается лишь через практический опыт, поэтому написание кода и решение задач (и, что особенно интересно и полезно, задач из собственной жизни) должны стать основой для включения читателя в мир DL. К какой-либо специальной теории можно обратиться и после, и во время практики, если это будет необходимо [там же: 36—38], а минимум общей теории лаконично изложен в перечне основных терминов, главный из которых, конечно, само «глубокое обучение», определяемое как «компьютерная методика для извлечения и преобразования данных... [реализуемая] через несколько слоев нейронных сетей, каждый из которых получает вводные от предыдущего и постепенно их уточняет; слои обучаются через алгоритмы, которые минимизируют ошибки и одновременно повышают точность; таким образом сеть учится выполнять конкретные задачи» [там же: 28—29]. В дополнение к основным терминам авторы, находя это важным для понимания принципов глубокого обучения, расширяют теоретический минимум кратким экскурсом в историю ис-

<sup>6</sup> О современных тенденциях в области ИИ в России см.: Яков и партнеры, Яндекс. Искусственный интеллект в России — 2023: тренды и перспективы. М., 2023. URL: [https://yakov.partners/upload/iblock/c5e/c8t1wrkdne5y9a4nqlicderalwny7xh4/20231218\\_AI\\_future.pdf](https://yakov.partners/upload/iblock/c5e/c8t1wrkdne5y9a4nqlicderalwny7xh4/20231218_AI_future.pdf) (дата обращения: 20.06.2023).

кусственного интеллекта [там же: 30—33] (дополнительно о терминологии и истории ИИ см. [Deng, Yu, 2014]).

Переход к непосредственной практике предваряется замечаниями о технической стороне машинного обучения (ML) вообще и глубокого обучения (DL) в частности: работа с ИИ невозможна без взаимодействия с программным обеспечением, представляющим собой в данной сфере довольно широкую палитру из фреймворков и приложений. При этом авторы считают, что принципиальной разницы, какое ПО изучать, нет: связанные с нейросетями новые технологии замещают старые решения крайне быстро, поэтому учиться следует скорее не конкретному инструменту, а умению адаптироваться к новшествам [там же: 39]. Однако начать с чего-то одного все же нужно, и для этого книга предлагает зарекомендовавший себя набор из языка программирования Python, среды программирования *Jupyter Notebook*, фреймворка для нейросетевого моделирования PyTorch и авторской разработки *fastai* (*fast artificial intelligence*) — библиотеки для высокоуровневого (на самом деле и средне-, и низкоуровневого) обращения к нейронным сетям [там же: 38—39].

Практика начинается с разработки классификатора изображений кошек и собак. Сначала Гуггер и Ховард дают рекомендации по настройке рабочей станции для развертывания нейросети, предлагая в качестве рабочего места облачные интернет-сервисы, обеспечивающие доступ к необходимому оборудованию бесплатно [там же: 40—41]. Затем приводится сам код классификатора с пояснениями к каждой строке: всего в шести строках программы помещаются все этапы моделирования, от загрузки набора данных для обучения до отладки нейросети. Результат работы кода — таблица, описывающая процесс обучения: какая по счету итерация обучения происходила, каково значение потерь, то есть ошибок работы модели, оценивая и минимизируя которые, нейросеть учится выполнять данную ей задачу лучше, на тренировочном массиве. Затем полученная нейросеть тестируется на случайном фото кота: машина действительно определяет картинку как фото именно кота [там же: 43—46].

На этом примере читатель уясняет ряд важных положений:

1) Чтобы получить обученную модель, потребовались только входные и выходные данные, то есть массив фотоизображений и массив указаний на классы изображенных на нем объектов. На основе этой информации нейросеть выработала в себе ряд алгоритмов для определения, кто изображен на картинке: кот или собака. Таким образом, если для выполнения задачи требуется лишь входная и выходная информация, то нейросети в принципе применимы к любой задаче, для которой есть исходные данные и определена цель, что с ними требуется сделать (подробно процесс обучения сети см. [там же: 46—53]).

2) Полученный классификатор позволяет рассмотреть онтологические ограничения ML, в том числе DL: для успешного обучения нейросети нужны не просто «сырые», но размеченные данные: метки, имеющие содержательную связь с поставленной целью<sup>7</sup>, а это, в свою очередь, означает, что в данных должны присутствовать регулярности (шаблоны, паттерны), доступные для обучения, поскольку в хаосе нейросети бесполезны [там же: 52].

<sup>7</sup> Подробнее о разметке на примере задач обработки изображений см. [Макаров, Чертков, Лукьянчиков, 2023].

3) Есть у ML и моральные ограничения. К наиболее значимым относят петлю обратной связи. Это ситуация, когда результаты нейросети могут воздействовать на мир, на данных о котором обучаются следующие поколения этой сети. Например, полицейские использовали нейросеть для предсказания района с наивысшим уровнем преступности, повышенный полицейский контроль фиксирует в предсказанном нейросетью районе больше правонарушений, чем обычно, и эти данные подаются в нейросеть для дальнейшего дообучения. Так машина перестает выявлять объективные закономерности и начинает генерировать предсказания «из себя», а не основываясь на закономерностях реальности. Опасность от петли обратной связи подчеркивает необходимость здравого смысла и постоянной методологической рефлексии при проверке правильности полученных результатов [там же: 53—54].

4) Входные данные всегда следует делить на обучающую, валидирующую (отложенную) и контрольную выборки. Долгое обучение модели на одних и тех же данных приводит к переобученности, что снижает эффективность нейросети в работе с новыми данными. Контроль за переобученностью ведется через использование валидирующей части датасета. Однако хорошая метрика на отложенной выборке может быть следствием подгонки настроек сети именно под валидационные данные. Чтобы избежать «привыкания», вводят третью выборку, на которой работа машины оценивается после обучения и настройки [там же: 56—58, 77—82]<sup>8</sup>.

5) Полученный классификатор обучался не с нуля, а был дообучен. В этом раскрывается суть *переноса обучения* (*transfer learning*) — одной из базовых технологий концепции *fastai*: под свою задачу вы берете нейросеть, показавшую хороший результат на аналогичной задаче, и ее *донастраиваете* и *дообучаете*<sup>9</sup>. Например, если нужно создать классификатор, распределяющий фото на пять классов, *fastai* предлагает взять зарекомендовавшую себя в классификации изображений модель (допустим, ResNet), задать нужное количество классов и обучить на ваших данных<sup>10</sup>.

Разбирая шесть строк кода, авторы демонстрируют принципы работы DL-архитектур с загруженными данными. Благодаря этому для читателя DL перестает быть «черным ящиком», что, в свою очередь, позволяет разработчику объяснять, что происходит внутри используемых приложений, и тем самым рассеивать естественное для всяких нововведений недоверие со стороны как заказчиков, так и коллег-социологов [там же: 61—63].

Таким образом, в главе 1 наглядно представлен теоретический минимум о DL, требованиях и ограничениях в сфере ML, принципах построения и обучения моделей, а также советы, как выбирать программное обеспечение.

Помимо непосредственно разработки нейросети, создание DL-продукта включает в себя подготовку данных до моделирования, а после построения модели — операции по интеграции построенной нейросети в приложение и тестированию ее работы. Об этом говорится в главе 2.

<sup>8</sup> О четвертом типе выборки — контролирующем, применяемом периодически после внедрения ИИ в работу, — см. [Четвериков и др., 2023].

<sup>9</sup> Подробнее о *transfer learning* см.: [Pan, Yang, 2009].

<sup>10</sup> Подробно о *переносе обучения* с помощью «fastai» см. с. 59—61 и в документации к фреймворку.

Прежде всего необходимо определить, какой вид нужно придать данным для решения поставленной задачи и как их разметить. Можно использовать массив фотографий, размеченный на кошек и собак, а можно создать таблицу, в которой звери описаны по отличительным признакам, и обучать модель на табличных данных. Оба способа хороши, но для разных задач. Так, известен метод классификации звука, в котором аудиоданные трансформируются в спектрограммы, то есть в картинки, к которым применяется классификатор изображений. Этот подход — неочевидный, прийти к нему можно лишь через перебор разных методик, однако именно он показывает SOTA-результаты (наилучшие из достигнутых на данный момент) в задаче аудиоклассификации. Помимо этого, в новых задачах часто хороши «старые» подходы, подобранные по аналогии данных: так, при работе с белковыми цепями, похожими по своей структуре на текст, текстуальные нейросети оказались крайне продуктивными [там же: 90, 93]. Вывод следующий: чтобы плодотворно сегментировать операции и подбирать под них инструменты, нужно не просто анализировать данные, но и быть готовым трансформировать их, причем так, чтобы не подменить стоящую задачу какой-то иной задачей, имеющей более очевидное решение. Трансформации подлежат «речь» задачи, которую надо адекватно перевести с языка непригодного для решения метода на язык пригодного метода.

За определением вида данных следует их сбор [там же: 95—100], после которого нужно снова решить, какие преобразования доступных для сбора данных будут полезны. Например, при классификации фото различные манипуляции с картинками, такие как изменение размера путем растягивания или сжатия изображения, вырезания из него случайного фрагмента, позволяют добиться единообразия подаваемого в нейросеть материала, что необходимо для ее работы. Помимо этого, такие искажения, как поворот фото на случайный градус, добавление в него случайного шума и другие, называемые аугментацией, улучшают качество сети, поскольку делают каждую итерацию (эпоху) ее обучения как бы новой (к тому же синтетически увеличивая объем датасета, что особенно полезно в случае, когда исходных данных мало): нейросети «не понимают» одинаковости измененных объектов, поэтому учатся искать на фото не конкретные сущности, а общие признаки, позволяющие решить поставленную задачу [там же: 103—105]. Ясно, что подобные трансформации применимы не только к картинкам, но и к аудио, и к табличным данным, и т. д.<sup>11</sup> Таким образом, умение ML-специалиста работать с подаваемой в сеть информацией позволяет, во-первых, создать набор данных, отвечающий требованиям к обучению DL-инструмента, и, во-вторых, улучшить качество работы модели.

Обозначенные процессы преобразования данных и их загрузки в нейросеть обернуты разработчиками `fastai` в удобный и простой инструмент — `DataLoader` в совокупности с используемым для его тонкой настройки `DataBlock` [там же: 100—104]. После загрузки данных и обучения нейросети следует еще один важный этап — анализ ошибок. Если точность модели не устраивает, следует посмотреть, где она ошибается чаще всего: возможно, ошибка проявляется в особой части данных и для исправления положения нужно увеличить их долю в обучаю-

<sup>11</sup> Об аугментации текстовых данных и ее особенностях см. [Матвеева, Махныткина, 2023].

щей выборке. Когда этот шаг не помогает, следует переходить к перестройке нейросети. Если же точность модели устраивает, анализ ошибок позволяет установить, а не достигается ли хороший результат тем, что данные не сбалансированы, и поэтому сеть просто относит, например, всю информацию к модальному классу, представляющему абсолютное большинство в массиве. Таким образом, рассмотрение ошибок действительно полезно для корректировки и проверки перформанса сети. Для его осуществления в `fastai` предусмотрено множество удобных методов [там же: 105—109]. Авторы неоднократно отмечают: работа с данными, их чистка и подготовка для загрузки в модель — очень важный этап развертывания DL-решения, занимающий, как правило, 90 % времени всей разработки [там же: 108].

После подготовки данных, обучения и отладки модели обычно возникает потребность интегрировать полученную нейросеть в приложение, которое обеспечит удобство взаимодействия с продуктом пользователям, не обладающим навыками программирования, коими и являются рядовые заказчики. Для решения этой задачи авторы предлагают использовать уже знакомый читателю `Jupyter` с последующей загрузкой программы на сервер; они описывают каждый шаг превращения кода в веб-приложение [там же: 109—117].

Наконец, перед передачей нейросети заказчику нужно протестировать поведение модели; этот этап гораздо труднее, чем в большинстве других цифровых решений. Поведение сети задается не заранее созданным алгоритмом, но «стремлением» модели соответствовать обучающим данным, что зачастую порождает две трудности: проблема данных, не соответствующих области, и проблема сдвига области. Первая трудность состоит в том, что нейросети могут попадаться данные, существенно отличающиеся от обучающей выборки. Вторая — в том, что любая модель устаревает в стремительно изменяющемся мире<sup>12</sup>. Ни то, ни другое затруднение невозможно полностью устранить заранее. Остается тщательное, кропотливое тестирование нейросети, которое непременно включает в себя на ранних этапах полностью ручное выполнение возложенной на нейросеть функции на разнородных массивах данных [там же: 117—120].

Авторы повторяют мысль: оставлять за сетью лишь вспомогательную функцию, сохраняя ведущую роль за человеком. Прежде чем внедрять системы ML, полезно ответить на вопросы: что будет, если система заработает действительно очень хорошо? Как верные прогнозы могут повлиять на принятие решений? Какими могут быть последствия и как их можно будет фиксировать? Помимо размыслений требуется разработка систем постоянного мониторинга (человеческого надзора) и своевременного реагирования на их сигналы [там же: 117—121]. Только так, по мнению авторов, непредсказуемость DL-разработки не выйдет за рамки исправимой ошибки.

В целом в главе 2 рассмотрены процедуры, следующие до и после непосредственно моделирования, а именно: 1) выбор типа данных, их сбор, проведение необходимых преобразований, загрузка в модель; 2) анализ ошибок, внедрение обученной на данных сети в приложение, тестирование полученного приложе-

<sup>12</sup> Подробнее об указанных проблемах см. [Намиот, Ильюшин, 2022].

ния во избежание проблем, связанных с применением результатов работы нейросети в жизни людей.

Сюжеты главы 2 плавно переходят в размышления о потенциальном вреде и этике использования нейросетей, обсуждаемые в главе 3. Уже есть примеры, когда использование нейросетей приносило людям страдание: ошибочное распределение DL-алгоритмами врачебной помощи, сокращающее количество часов, предназначенных для больных диабетом или церебральным параличом, и др. [там же: 126—129]. Становится очевидно: эти проблемы многогранны, над ними нужно серьезно работать. Причем наиболее эффективен подход, когда над этическим вопросом думает команда, поскольку так достигается триангуляция в понимании текущей ситуации. Разные люди (по полу, возрасту профессии и т. д.), обладая разными точками зрения, позволяют охватить как можно больший объем возможных последствий и, соответственно, предусмотреть как можно больше потенциальных проблем [там же: 126, 155—157].

Почему команда должна думать об этичности своей работы, почему она не может следовать формуле «надо — делаю», оставляя ответственность за негативные аффекты итогового продукта на директорате? Потому что именно аналитики и разработчики имеют прямое отношение к опыту взаимодействия с моделью, именно они могут влиять на то, каким образом и в каких ситуациях сетью будут пользоваться (и будут ли ей пользоваться вообще), и поэтому они должны учитывать, как модель может применяться и как это применение сделать лишь положительным [там же: 129—133].

Проверка своей модели на этичность, продумывание последствий введения нейросети в жизнь — попытка не допустить возникновения ошибок. Однако нужно понимать, что в подавляющем большинстве случаев ошибки все равно рано или поздно возникнут, поэтому помимо предотвращения проблем следует учиться и их исправлению. Разные ошибки требуют разных способов их решения.

Уже рассмотренные нами петли обратной связи, конечно, нужно предугадывать: например, статистика показывает, что на встречи, посвященные технике и технологиям, более склонны ходить мужчины, и система, подбирающая потенциальных участников, при введении пола как переменной начнет воспроизводить указанное неравенство, поэтому исследователю нужно сознательно исключить признак половой принадлежности из обучения модели, что сделает сеть менее точной, но в то же время лишит ее работу дискриминационного характера. Однако часто заметить петлю до запуска сети в производство сложно. Совет один: увидев смещение в результатах модели, необходимо оперативно проанализировать его причины и переобучить модель, исключив проблемные признаки, следовательно, «здесь и сейчас» «развязать» петлю обратной связи [там же: 135—138].

Помимо петель нужно помнить о возможной необъективности ML-продукта. Необъективность — порождение так называемых шести смещений, которые подробно рассмотрены на с. 138—147. Анализ смещений наводит на мысль о «смещенности» самого человека и производимых им данных. Есть ли тогда смысл требовать объективности от алгоритмов, работающих с информацией о человеке и от человека? Да! Технологии сильны, мы доверяем им сложные вычисления, мы склонны думать, что они либо не допускают ошибок вообще, либо делают их в крайне

редких случаях, мы внедряем их в нашу жизнь со все большим размахом, вследствие чего влияние моделей на мир растет. Однако технологии и люди должны сойтись в одном: самые влиятельные персоны всегда вызывают полемику вокруг своих идей, суть которой — установить, точно ли идея, продвигаемая этой личностью, адекватна, не смещена, нацелена на благо. Так должно быть и с моделями. Нам нужно преодолеть эйфорию, вызванную последними достижениями в областях IT, DL, ведущую к ощущению неограниченности способностей машин и нейросетей, и понять, что все эти машины работают с нашими данными, в которых мы вполне можем найти смещения и, что самое главное, их устранить. Так, наши «способные на все» модели, обученные на чистых, адекватных данных, позволят нам радоваться не только их сложности и скорости, но и их объективности и этичности. Раз модели так сильно влияют на наш мир, давайте направим это влияние на позитивные изменения общества [там же: 148—149].

Кроме проверки на «этичность» непосредственно моделей и их работы, необходимо тестировать и данные: в их отношении всегда нужно думать, как информация, которая собирается и хранится, которую модель принимает на входе и генерирует на выходе, может быть использована во вред обществу, например в качестве информационного оружия. Этот тезис возвращает нас к вопросу об ответственности лиц, включенных в ML-проект: в случае, если негативные для мира последствия использования данных возможны, необходимо крайне серьезно подойти к их защите. Если же вы не уверены, что сможете защитить данные от вторжения посторонних лиц, а использование информации не по назначению способно принести значительный вред обществу, нужно решить, стоит ли продолжать проект, сохраняя тем самым возможность возникновения серьезной опасности для окружения [там же: 153].

Ясно, что приведенные рекомендации помогут командам, чувствующим ответственность за свою продукцию, достичь большей «этичности» DL для общества, однако велика вероятность, что останутся «другие», которые все же будут эксплуатировать свои алгоритмы, несмотря на их дискриминационный, смещенный характер [там же: 159]. В случае, когда речь идет о таких проблемах, лежащих в природе целых структур и областей общественной деятельности, необходимо прибегать к помощи государства. Авторское предложение — изменить на государственном уровне систему стимулов для получения прибыли, поскольку именно она ведет бизнес к нарушению общечеловеческих правил ради больших заработков [там же: 159—160]. Иной аргумент для обращения к политике ради защиты общественной — факт, что проблемы неверного применения IT-продуктов затрагивают общественные блага, такие как информационная среда, а точнее ее «чистота», или общественная конфиденциальность, или человеческие права в целом, а общественные блага должны быть защищены государством [там же: 161].

Проблема урегулирования сферы ML сродни обустройству безопасности дорожного движения: ясно, что компаниям не хочется тратить деньги на проведение краш-тестов, на внедрение стандартов безопасности и экологических стандартов, однако без государственного принуждения к таким действиям нормальная, «сохранная» жизнь человека в условиях автомобилизации становится невозможной [там же: 161—162].

Заинтересованный читатель найдет в книге множество других размышлений об этической проблеме при применении искусственного интеллекта, мы же постарались описать авторское видение данного аспекта ИИ в целом, акцентируя внимание на наиболее важных, с нашей точки зрения, положениях<sup>13</sup>.

Вывод из главы 3 таков: этические вопросы в сфере ML в целом и DL в частности действительно сложны, готовых верных ответов на них нет. Поэтому все исследовательское сообщество должно участвовать в поиске удовлетворительных формулировок и требований [там же: 162—163].

В трех первых главах заключена базовая информация для плодотворного знакомства с дальнейшими главами о нейросетях и их архитектурах, включающая в себя рассмотрение процессов сбора и обработки данных и обучения на них нейронных сетей, внедрения полученных DL-разработок в приложения, принципов оценки «этичности» собственных продуктов, шагов по минимизации вероятности возникновения связанных с деятельностью в сфере DL проблем. Остальные три части посвящены непосредственно практике и похожи на курс, осваивать который нужно самостоятельно, поэтому останавливаться на них в рецензии нецелесообразно.

По прочтении книги нам захотелось испробовать инструменты DL, предлагаемые авторами в рамках библиотеки *fastai*, на практике, чтобы понять, действительно ли принципы относительно простого и быстрого развертывания качественных нейросетей, делающие книгу такой привлекательной в глазах читателей, реализованы в программном обеспечении составителей текста.

Для тестов мы выбрали три соревновательные задачи с платформ *Kaggle* и *Zindi*:

а) выявление из массива постов из соцсетей тех записей, в которых говорилось о какой-либо катастрофе;

б) классификация изображений почвы, сделанных телефонными камерами, на классы в зависимости от вида повреждения земли (с целью принятия своевременных решений по сохранению плодородия почв, которым нанесен тот или иной ущерб);

в) классификация новостей, написанных на языке суахили, по темам статей.

Помимо общедоступных задач, нами была предпринята попытка использовать *fastai* для классификации аудиозаписей неоконченных телефонных интервью длительностью до 20 секунд, то есть для выполнения задачи, над которой мы уже достаточно длительное время работаем и в рамках которой у нас скопился обширный инструментарий, продемонстрировавший хороший и воспроизводимый результат, следовательно, имеется возможность сравнить перформанс моделей из новой библиотеки с зарекомендовавшими себя методами.

Проведя ряд экспериментов, мы выяснили: библиотека действительно позволяет получить качественное решение, прилагая минимум усилий к непосредственно построению архитектур моделей и к их обучению или дообучению. Однако важнейших этап подготовки данных, балансировки массивов, их аугментации, их наилучшего способа представления в памяти компьютера и, следовательно, в качестве посылок на вход нейросетей — задача, все же требующая непосредственного участия разработчика и, более того, нуждающаяся в творческом, «ищущем» подходе.

Таким образом, мы пришли к выводу, что *fastai* — заслуживающая своей популярности библиотека, применение которой способно ускорить и облегчить про-

<sup>13</sup> Подробнее о данной стороне ИИ с позиции социально-философской проблематики см. [Шляпников, 2023; Ивлев, 2023].

цесс разработки нейросетевых решений во многих областях человеческой деятельности. К тому же она проста в освоении, поэтому поможет новичкам в сфере глубокого обучения совершить быстрый и стремительный старт. Тем не менее это ПО — не заменитель человека, но лишь его помощник, который позволяет специалисту ставить больше экспериментов и, следовательно, искать лучшие пути решения стоящих перед ним задач. Fastai хороша как полноценный инструмент для решения относительно простых задач, в деле же более сложных проблем она полезна как способ быстрой разработки прототипа, «прощупывания почвы», осознания, насколько способны нейросетевые модели к решению этих заданий, но все же ведущая роль в таких ситуациях должна оставаться за человеком.

### Список литературы (References)

1. Ивлев Д. В. Искусственный интеллект и проблемы этики // Право и практика. 2023. № 4. С. 263—267.  
Ivlev D. V. (2023) Artificial Intelligence and Ethical Issues. *Law and Practice*. No. 4. P. 263—267. (In Russ.)
2. Котов Д. А. Алгоритмы искусственного интеллекта в прикладных социологических исследованиях // СоциоДиггер. 2023. Т. 4. № 7—8. URL: <https://sociodigger.ru/articles/articles-page/algorithmy-iskusstvennogo-intellekta-v-prikladnykh-sociologicheskikh-issledovaniyakh> (дата обращения: 12.10.2024).  
Kotov D. A. (2023) Artificial Intelligence Algorithms in Applied Sociological Research. *SocioDigger*. Vol. 4. No. 7—8. URL: <https://sociodigger.ru/articles/articles-page/algorithmy-iskusstvennogo-intellekta-v-prikladnykh-sociologicheskikh-issledovaniyakh> (accessed: 12.10.2024). (In Russ.)
3. Макаров К. С., Чертков С. Н., Лукьянчиков К. В. Концептуальная модель системы разметки графических данных для решения задач компьютерного зрения с использованием технологий машинного обучения // Auditorium. Электронный научный журнал Курского государственного университета. 2023. № 2. URL: [https://api-mag.kursksu.ru/api/v1/get\\_pdf/4939/](https://api-mag.kursksu.ru/api/v1/get_pdf/4939/) (дата обращения: 20.06.2024).  
Makarov K. S., Chertkov S. N., Lukyanchikov K. V. (2023) Conceptual Model of a Graphic Data Markup System for Solving Computer Vision Problems Using Machine Learning Technologies. *Auditorium. Electronic Scientific Journal of Kursk State University*. No. 2. URL: [https://api-mag.kursksu.ru/api/v1/get\\_pdf/4939/](https://api-mag.kursksu.ru/api/v1/get_pdf/4939/) (accessed: 20.06.2024). (In Russ.)
4. Матвеева А. А., Махныткина О. В. Метод аугментации текстовых данных с сохранением стиля речи и лексики персоны // Научно-технический вестник информационных технологий, механики и оптики. 2023. Т. 23. № 4. С. 743—749. <https://doi.org/10.17586/2226-1494-2023-23-4-743-749>.  
Matveeva A. A., Makhnytkina O. V. (2023) Text Augmentation Preserving Persona Speech Style and Vocabulary. *Scientific and Technical Journal of Information Technologies, Mechanics and Optics*. Vol. 23. No. 4. P. 743—749. <https://doi.org/10.17586/2226-1494-2023-23-4-743-749>. (In Russ.)

5. Намиот Д. Е., Ильющин Е. А. Мониторинг сдвига данных в моделях машинного обучения // *International Journal of Open Information Technologies*. 2022. Т. 10. № 12. С. 84—93.  
Namiot D. E., Ilyushin E. A. (2022) Data Shift Monitoring in Machine Learning Models. *International Journal of Open Information Technologies*. Vol. 10. No. 12. P. 84—93. (In Russ.)
6. Резаев А. В., Трегубова Н. Д. От социологии к новой социальной аналитике: кризис социологии и проблема искусственного интеллекта // *Социологическое обозрение*. 2021. Т. 20. № 3. С. 280—301.  
Rezaev A. V., Tregubova N. D. (2021) Sociology on the Way to New Social Analytics: The Crisis in Sociology and the Problem of Artificial Intelligence. *Russian Sociological Review*. Vol. 20. No. 3. P. 280—301. (In Russ.)
7. Семина Т. В., Го В. Воздействие технологий искусственного интеллекта на социальные отношения // *Социология*. 2022. № 3. С. 173—178.  
Semina T. V., Guo W. (2022) The Impact of Artificial Intelligence Technologies on Social Relations. *Sociology*. No. 3. P. 173—178. (In Russ.)
8. Ховард Дж., Гуггер С. Глубокое обучение с fastai и PyTorch: минимум формул, минимум кода, максимум эффективности. СПб.: Питер, 2022.  
Howard J., Guggler S. (2022) Deep Learning for Coders with fastai and PyTorch: AI Applications Without a Ph D. St. Petersburg: Piter. (In Russ.)
9. Четвериков С. Ф., Арзамасов К. М., Андрейченко А. Е., Новик В. П., Бобровская Т. М., Владимирский А. В. Подходы к формированию выборки для контроля качества работы систем искусственного интеллекта в медико-биологических исследованиях // *Современные технологии в медицине*. 2023. Т. 15. № 2. С. 19—27. <https://doi.org/10.17691/stm2023.15.2.02>.  
Chetverikov S. F., Arzamasov K. M., Andreichenko A. E., Novik V. P., Bobrovskaya T. M., Vladimirovsky A. V. (2023) Approaches to Sampling for Quality Control of Artificial Intelligence in Biomedical Research. *Sovremennye tehnologii v medicine*. Vol. 15. No. 2. P. 19—27. <https://doi.org/10.17691/stm2023.15.2.02>.
10. Шляпников В. В. Некоторые проблемы этики искусственного интеллекта // *Идеи и идеалы*. 2023. Т. 15. № 2—2. С. 365—376. <https://doi.org/10.17212/2075-0862-2023-15.2.2-365-376>.  
Shlyapnikov V. V. (2023) Some Problems with Artificial Intelligence Ethics. *Ideas and Ideals*. Vol. 15. No. 2—2. P. 365—376. (In Russ.) <https://doi.org/10.17212/2075-0862-2023-15.2.2-365-376>.
11. Deng L., Yu D. (2014) Deep learning: Methods and Applications. *Foundations and Trends® in Signal Processing*. Vol. 7. No. 3—4. P. 197—387. <https://doi.org/10.1561/20000000039>.
12. Pan S. J., Yang Q. (2010) A Survey on Transfer Learning. *IEEE Transactions on Knowledge and Data Engineering*. Vol. 22. No. 10. P. 1345—1359. <https://doi.org/10.1109/TKDE.2009.191>.

DOI: [10.14515/monitoring.2024.5.2623](https://doi.org/10.14515/monitoring.2024.5.2623)



**А. Ю. Телицына**

## **ОПТИМИЗАЦИЯ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ ЧЕРЕЗ ИНТЕГРАЦИЮ ИИ: НЕЙРОННЫЕ СЕТИ КАК ИНСТРУМЕНТ В РАБОТЕ С АКАДЕМИЧЕСКОЙ ЛИТЕРАТУРОЙ**

### **Правильная ссылка на статью:**

Телицына А. Ю. Оптимизация научной деятельности через интеграцию ИИ: нейронные сети как инструмент в работе с академической литературой // Мониторинг общественного мнения: экономические и социальные перемены. 2024. № 5. С. 218—236. <https://doi.org/10.14515/monitoring.2024.5.2623>.

### **For citation:**

Telitsyna A. Y. (2024) Optimization of Scientific Activities Through AI Integration: Neural Networks as a Tool in Working with Academic Literature. *Monitoring of Public Opinion: Economic and Social Changes*. No. 5. P. 218–236. <https://doi.org/10.14515/monitoring.2024.5.2623>. (In Russ.)

Получено: 02.05.2024. Принято к публикации: 08.07.2024.

## ОПТИМИЗАЦИЯ НАУЧНОЙ ДЕЯТЕЛЬНОСТИ ЧЕРЕЗ ИНТЕГРАЦИЮ ИИ: НЕЙРОННЫЕ СЕТИ КАК ИНСТРУМЕНТ В РАБОТЕ С АКАДЕМИЧЕСКОЙ ЛИТЕРАТУРОЙ

*ТЕЛИЦЫНА Александра Юрьевна — кандидат биологических наук, доцент, старший научный сотрудник Центра исследований гражданского общества и некоммерческого сектора, Национальный исследовательский университет «Высшая школа экономики», Москва, Россия*  
E-MAIL: [atelitsyna@hse.ru](mailto:atelitsyna@hse.ru)  
<https://orcid.org/0000-0002-0186-3989>

**Аннотация.** В статье представлены аспекты интеграции технологий искусственного интеллекта (ИИ) в научную деятельность в области общественных наук. Автор приводит конкретные методики, инструменты и примеры использования ИИ для анализа библиографии и иного текстового контента, а также дает характеристики различным инструментам ИИ и подробно останавливается на их возможностях в области создания научных статей — от подготовки данных до интерпретации результатов и написания текста. В работе уделяется внимание не только техническим аспектам, но и этическим вопросам, связанным с использованием ИИ в обработке данных.

**Ключевые слова:** искусственный интеллект, генеративный искусственный интеллект, нейросети, чат-бот, написание научных статей, обзоры академической литературы

**Благодарность.** Работа осуществлена в рамках Программы фундаментальных исследований НИУ ВШЭ.

### Введение

Существует множество определений искусственного интеллекта (ИИ). В тесте Тьюринга ИИ определяется как способность машин общаться с людьми (используя электронные устройства вывода), не раскрывая при этом, что они не люди. Основ-

## OPTIMIZATION OF SCIENTIFIC ACTIVITIES THROUGH AI INTEGRATION: NEURAL NETWORKS AS A TOOL IN WORKING WITH ACADEMIC LITERATURE

*Alexandra Yu. TELITSYNA<sup>1</sup> — Cand. Sci. (Biol.), Associate Professor, Senior Research Fellow at the Center for Civil Society and Nonprofit Sector Studies*  
E-MAIL: [atelitsyna@hse.ru](mailto:atelitsyna@hse.ru)  
<https://orcid.org/0000-0002-0186-3989>

<sup>1</sup> HSE University, Moscow, Russia

**Abstract.** The article regards various aspects of the integration of artificial intelligence (AI) technologies into scientific activities in the field of social sciences. The author provides specific methods, tools, and examples of using AI to analyze bibliographies and other text content, and also gives characteristics of various AI tools and dwells in detail on their capabilities in the field of creating scientific articles, going from data preparation to interpretation of results and writing a text. The article discusses not only technical aspects, but also ethical issues related to the use of AI in data processing.

**Keywords:** artificial intelligence, generative artificial intelligence, neural networks, chatbot, research article writing

**Acknowledgments.** The work was carried out within the framework of the Fundamental Research Program of the HSE University.

ным критерием оценки является двоичность. Марвин Минки, один из пионеров ИИ, определил ИИ как способность машин делать то, что требует человеческого интеллекта. Символическая школа считает, что ИИ — это работа символов, а самые примитивные символы соответствуют физическим сущностям. Хотя интерпретации ИИ различны, широко распространено мнение, что ядро ИИ — это исследовательские теории, методы, технологии и приложения для моделирования, расширения и совершенствования человеческого интеллекта [Jiang et al., 2022].

Оптимизация научной деятельности через интеграцию искусственного интеллекта, в частности нейронных сетей, становится все более актуальной. Использование инструментов ИИ, таких как генеративные модели, имеет потенциал для преобразования различных сфер жизни, включая научные исследования [Лебедев, 2008]. Генеративный ИИ — это форма искусственного интеллекта, которая использует методы машинного обучения и глубокого обучения для генерации новых данных. В отличие от традиционных задач, таких как классификация и регрессия, генеративный ИИ обладает способностью автономно генерировать новые материалы, включая изображения, музыку и текст [Yu, Guo, 2023].

Эти инструменты ИИ в научных исследованиях могут собирать информацию с помощью технологии искусственных нейронных сетей, понимать язык и контекст, создавать новый контент и персонализировать коммуникации [Лебедев, 2008]. Интеграция ИИ в академических исследованиях помогает экономить ресурсы (в частности, временные) [Федосова, Берченко, Машошин, 2021]. Оптимизация научной деятельности с помощью интеграции ИИ, особенно нейронных сетей, может привести к повышению эффективности исследовательских процессов, способствуя развитию знаний и инноваций в различных областях [Лебедев, 2008]. Наиболее удобным инструментом является чат-бот с искусственным интеллектом — типичный и один из самых элементарных и распространенных примеров интеллектуального взаимодействия человека и компьютера. Это компьютерная программа, которая реагирует как интеллектуальная сущность, когда с ней разговаривают посредством текста или голоса, и понимает один или несколько человеческих языков с помощью обработки естественного языка (NLP) [Khanna et al., 2015].

ChatGPT, выпущенный Open AI в ноябре 2022 г., получил распространение как вспомогательный элемент в повседневной профильной деятельности специалистов в различных областях: образовании<sup>1</sup>, медицине [Staszak et al., 2021], научных исследованиях, архитектуре, проектировании и строительстве [Bilal et al., 2016], инжиниринге [Xu et al., 2021] и др. В некоторых научных статьях ChatGPT указывается в качестве соавтора [Curtis, 2023], в других его вклад расценивается как существенный в подготовке материала, но без включения в число авторов исследования [Buchkremer et al., 2019].

Публикуются работы, в которых осмысливается творческий опыт ИИ в написании статьи «с чистого листа» (например, [Cotton, Cotton, Shipway, 2023], где поднимаются проблемы использования нейросетей при написании научных работ). Био-

<sup>1</sup> Shrivastava R. Teachers Fear ChatGPT Will Make Cheating Easier than Ever // Forbes. 2022. December 12. URL: <https://www.forbes.com/sites/rashishrivastava/2022/12/12/teachers-fear-chatgpt-will-make-cheating-easier-than-ever/?sh=1374c3e1eef9> (дата обращения: 28.10.2024).

логи из Израиля<sup>2</sup> сгенерировали научную статью с помощью ChatGPT. Целью их эксперимента было изучение возможностей чат-бота в качестве напарника в написании академических текстов. Однако ученые остались недовольны результатом ввиду искажения данных и «подгонки» результата под запрос пользователя.

Ирландские исследователи протестировали ChatGPT в качестве автора исследования на всех стадиях его реализации [Jeyaraman et al., 2023]. Сначала они попросили чат-бота сгенерировать стандартные четыре части научной работы: идею исследования, обзор литературы, набор данных и предложения по тестированию и экспертизе. Ученые обозначили тему и указали, что результаты должны быть опубликованы в «хорошем финансовом журнале». Они загрузили для чат-бота порядка 200 рефератов и академических работ по теме исследования и дали команду учесть это при составлении статьи. Полученный результат обсуждался с 32 рецензентами, отобранными из участников нескольких фокус-групп: их просили оценить, были ли результаты достаточно полными, правильными и внесли ли они достаточно новый вклад в творческий процесс написания ИИ научных работ, чтобы его можно было опубликовать во влиятельном академическом финансовом журнале. Эксперты-рецензенты в целом сочли результаты работы чат-бота приемлемыми.

ИИ может создать «черновик» будущего исследования — наметить план, посчитать статистические данные, может работать с большими объемами данных, тем самым экономя время людей. Независимо от его функциональных «обязанностей», академическое сообщество называет ИИ «новым общим методом изобретательства» [Kitchenham, Charters, 2007].

Немецкие ученые разработали инструмент «двойной воронки искусственного интеллекта» [Bianchini, Müller, Pelletier, 2022], который, работая с большими данными существующих исследований по конкретной теме, предоставляет результаты не только в виде текста, но и в формате изображений и графиков, что дает преимущество визуализации, компоновки и наглядности данных.

Мощности ИИ и его легкость в обращении с большими объемами информации приходят на помощь ученым в составлении систематического обзора литературы, компенсируя временные ограничения человека [de la Torre-López, Ramírez, Romero, 2023; Cheng, 2022].

Проблемы при анализе литературы ИИ выявляются на этапах планирования и отчетности, которые сложнее поддаются автоматизации. Прогресс в этой области можно ожидать в ближайшем будущем благодаря появлению новых подходов, основанных на глубоком обучении [Salvagno, Taccone, Gerli, 2023].

Метод интеллектуального анализа баз данных, который является одним из вариантов использования ИИ, постоянно совершенствуется с учетом ранее выявленных в нем недостатков, однако неизменной остается точность работы алгоритмов поиска [Касена, Plotkin, Fehrenbacher, 2024]. При поиске литературы инструменты ИИ могут помочь авторам в процессе обзора, находя научные статьи, суммируя их выводы и выделяя области неопределенности. Это позволяет исследователям быстро получить представление о текущем состоянии знаний по конкретной теме

<sup>2</sup> Голованов Г. ChatGPT за час написал научную статью с нуля // Хайтек+. 2023. 7 июля. URL: <https://hightech.plus/2023/07/07/chatgpt-za-chas-napisal-nauchnuyu-statyu-s-nulya> (дата обращения: 12.10.2024).

и выявить потенциальные пробелы, которые необходимо устранить, хотя предоставленное ИИ резюме может быть довольно общим и не содержать критического анализа различий [Gilat, Cole, 2023].

Снижение неопределенности в различных кейсах с помощью инструментов ИИ в будущем станет возможно в ряде гуманитарных наук, например в юриспруденции. И. Гроссман предположил, что если магистранту права можно было бы подсказать реальные (или сфабрикованные) предыстории, чтобы точно отразить то, что могут сказать стороны дела, теоретически они могли бы заменить фокус-группы или использоваться в качестве агентов в экономических исследованиях<sup>3</sup>. Магистров права можно было бы обучать на различных персонажах, а затем их поведение можно было бы использовать для имитации экспериментов, результаты которых, если они будут интересными, позже можно было бы подтвердить на людях.

Инструменты ИИ в научных исследованиях могут использоваться при работе с литературой, переводом, обобщением результатов [VanHammat, 2023]. Нейросети позволяют анализировать и интерпретировать сложные связи между объектами, что особенно полезно для изучения научных публикаций и сетей цитирований между ними [Khemani et al., 2024]. Однако важно учитывать ограничения нейросетей, такие как поверхностный анализ данных, буквальное понимание запроса — без учета смысловых значений [Alshemali, Kalita, 2020], и существующие методы их преодоления при работе с литературными источниками [Alzubaidi et al., 2021]. Обзоры литературы, сделанные с помощью ИИ, отличаются тем, что чат-боты работают с большими и быстрорастущими объемами документов, то есть с частично структурированными (мета)данными, и охватывают почти все типы статей, посвященных исследованию информационных систем и связанных с социальными науками [Шемилева, Кудусова, 2022].

### Может ли ИИ быть автором научных статей?

Бельгийские ученые [Salvagno, Tacccone, Gerli, 2023] изучили возможности ChatGPT в создании проекта и корректуре статей и пришли к выводу, что чат-боты — это просто инструменты, они могут помогать исследователям-людям, но не должны использоваться в качестве замены их опыта, суждений и личности. По мнению исследователей [Patel, Lam, 2023], бывает сложно распознать, кто является автором текста — чат-бот или человек, поскольку первый использует передовые методы — обработку естественного языка (НЛП) и машинное обучение — для генерации текста, который очень похож на человеческий. Однако авторство нейросети все же можно обнаружить с помощью детекторов выходных данных ИИ и скептически настроенных рецензентов [Gao et al., 2022]: тексту, сгенерированному чат-ботом, может не хватать точных формулировок и выбора слов, которые использует автор-человек для передачи смысла или тона. Автоматически созданный текст также может быть расплывчатым и содержать несоответствия, которых не было бы в статье, написанной человеком. Вместе с тем, если статья содержит большое количество структурных и грамматических ошибок, можно предположить, что она написана человеком. Наконец, если в статье обсуждается специфическая

<sup>3</sup> How Scientists Are Using Artificial Intelligence (2023) *The Economist*. September 13. URL: <https://www.economist.com/science-and-technology/2023/09/13/how-scientists-are-using-artificial-intelligence> (дата обращения: 12.10.2024).

и сугубо техническая тема, маловероятно, что чат-бот смог ее написать, поскольку для этого требуются глубокое понимание темы и способность проводить научный анализ, генерировать выводы [Salvagno, Taccone, Gerli, 2023].

Инструменты ИИ, используемые для оптимизации аннотации, введения и заключения, могут помочь сделать текст более лаконичным, видимым в поисковых системах, предложить более четкие формулировки [Dai, Xiong, Ku, 2023].

Что касается творческого потенциала ИИ, то часть исследователей оценивает его высоко, соглашаясь, что работы, сгенерированные им, имеют все признаки художественного произведения: оригинальность, ценность, новизну, выразительность, новаторство и провокацию [Mazzone, Elgammal, 2019; Gunser et al., 2021; Ghajargar, Bardzell, Lagerkvist, 2022; Fang, Ng, Leung, 2023]. Однако другая часть исследовательского сообщества рассматривает ИИ лишь в качестве творческого партнера, но не самостоятельного автора научной статьи [Chatterjee, 2022].

ИИ не способен провести исследование «с нуля», определить его цели и задачи, методологию и инструменты ввиду отсутствия понимания лингвистических нюансов человеческого языка, точнее, способности интеллектуально «осмыслить» запрос исследователя (ИИ не может работать с «мягкими» данными<sup>4</sup>) и причинно-следственной связи мышления (на данной стадии развития ИИ), а также из-за неспособности того или иного инструмента ИИ самостоятельно двигаться по траектории исследования, поскольку любой ИИ действует автономно, в соответствии со сменяющейся обстановкой (формулировками и запросами исследователя)<sup>5</sup>, то есть по воле исследователя, который формулирует запросы.

Рассмотрим подробнее возможности и преимущества инструментов ИИ в помощи академическим ученым.

**1. Экономия времени.** Генерация текстов силами ИИ экономит время исследователя за счет ускорения процесса создания материала. Это не обязательно сразу составление чистового варианта: генераторы текстов могут быть полезны в создании черновиков документов и обзоров литературы. Среди них Gemini, ChatGPT, Humata и др. С полным перечнем можно ознакомиться в Приложении.

Тематическое исследование *Editage Insights* показывает пример интеграции ИИ в академическую работу<sup>6</sup>. В нем изучается, как инструменты ИИ улучшают качество письма, анализа данных, обзора литературы и планирования исследований. В работе подчеркивается, что ИИ является вспомогательным механизмом, повышающим эффективность, при этом сохраняется потребность в человеческом творчестве и критическом мышлении в исследовательских процессах. Это отражает растущую важность ИИ в оптимизации академических рабочих процессов. ChatGPT или аналогичные программы могут стать ценными инструментами для ускорения подготовки первых черновиков статей и грантов, обобщения или упрощения длинных статей, сокращения дублирования, доработки и оптимизации су-

<sup>4</sup> Имеются в виду данные, для работы с которыми требуется общее понимание закономерностей, междисциплинарные знания и умения. «Мягкие данные» относятся к качественной, субъективной и интерпретируемой информации, которая дает детальное представление о сложных явлениях.

<sup>5</sup> Искусственный интеллект: сценарии практического применения. Спикер: Павел Карасев // Науч. Путеводитель по науке в Москве. URL: <https://www.youtube.com/watch?v=OjZsuyOMWsg> (дата обращения: 24.03.2024).

<sup>6</sup> Khedkar S. Using AI-Powered Tools Effectively for Academic Research // Editage. 2023. September 13. URL: <https://www.editage.com/insights/using-ai-powered-tools-effectively-for-academic-research> (дата обращения: 28.10.2024).

ществующих документов или оказания помощи исследователям с ограниченными возможностями [Babl, Babl, 2023].

2. *Помощь в преодолении «блокировки писателя» («страха чистого листа»)*. Любой автор может столкнуться с ситуацией блока на одном из этапов написания работы. Генераторы текстов могут предложить варианты формулировок или идей, стимулируя творческий процесс.

Несмотря на возможные ошибки при цитировании статей, некоторые ученые считают ИИ ценным инструментом в силу его способности помочь ученым организовать материалы, создать черновики и провести корректуру [Altmäe, Sola-Leyva, Salumets, 2023]. Еще один вариант — использование ChatGPT для написания научной статьи в разделах, требующих количественных исчислений и подсчета данных. Указывая на проблему плагиата [Kacena et al., 2024; Vintzileos et al., 2023], исследователи видят и положительные стороны, такие как ускорение процесса работы и надежное подспорье в преодолении проблемы «писательского/творческого паралича».

3. *Генерация идей для исследований*. ИИ способен предложить новые направления для исследований или подходы к анализу данных на основе имеющихся научных работ, он может помочь сформулировать гипотезы или составить исследовательские вопросы. Для работы с научной литературой и экономии времени поиска исследований существует обширный перечень программ (например, Semantic Scholar, R Discovery, ASReview и др.), с которыми можно ознакомиться в Приложении (Таблица 1).

Полноценно заменить автора исследования программа ИИ не способна [Altmäe, Sola-Leyva, Salumets, 2023], во многом из-за отсутствия критического мышления [Guleria et al., 2023]. ИИ-инструменты могут предложить основу для текста, сгенерировать идеи и даже предложить варианты формулировок, однако тексты, созданные ими, пока не имеют эмоциональной окраски и не демонстрируют глубокого понимания контекста, интуитивного восприятия и чувства стиля, характерных для человека. Поэтому использование ИИ для написания статей обычно сочетается с дальнейшим редактированием и доработкой со стороны человека, чтобы получить качественный оригинальный текст<sup>7</sup>.

4. *Составление аннотаций и рефератов*. Генераторы текстов могут создавать аннотации к научным статьям или рефераты, экономя время ученых на подготовку этих материалов. Однако при условии обязательного контроля со стороны человека, поскольку аннотация является важной частью статьи и ее построение требует соблюдения строгих правил.

5. *Расширение языковых возможностей*. Для многих ученых английский язык не является родным, в то время как именно он сейчас признается международным языком науки. Использование генераторов текстов помогает формулировать мысли на иностранном языке, улучшать стиль и исправлять грамматические ошибки. Самообучающиеся программы ИИ могут делать переводы с любого языка (и на любой), используя лексику и специфику заданных исследований.

<sup>7</sup> Программа повышения квалификации «Инновации в науке: использование ИИ для написания статей» // Казанский федеральный университет. 2024. 21 марта. URL: <https://kpfu.ru/ec/programma-povysheniya-kvalifikacii-innovacii-v.html> (дата обращения: 05.06.2024).

6. **Популяризация научного контента.** ИИ может помочь в создании статей для блогов, пресс-релизов или социальных сетей, связанных с научной работой, тем самым расширяя аудиторию и помогая распространять научное знание для широких масс.

7. **Улучшение качества текстов.** Инструменты, основанные на ИИ, могут анализировать написанный текст и предлагать улучшения, например, по структуре, ясности изложения и устранению повторов.

Генераторы текстов и других инструментов ИИ в академической работе должны использоваться с соблюдением этических норм и академической честности, всегда с указанием использованных источников и подходов. Среди наиболее известных инструментов — Quillbot и 300ya. Ознакомиться с перечнем инструментов для улучшения и генерирования текста можно в Приложении (Таблица 2).

### **Методологические ограничения ИИ в научных исследованиях**

Применение ИИ в научных исследованиях имеет ряд методологических ограничений ввиду возможных ошибок в его деятельности. «Слабое место» ИИ при цитировании и использовании ссылок на научные работы и проведенные исследования при подготовке научной статьи в том, что многие ссылки оказываются в лучшем случае неточными [Kasena et al., 2024; Misra, Chandwar 2023], в худшем — их просто не существует [Zhao et al., 2021]. Но это зависит от того, какой именно инструмент ИИ использовать, уже разработаны сети, которые предоставляют настоящие ссылки.

Манипулирование научными данными или их искажение — еще одна серьезная проблема. Известны случаи, когда ученые использовали инструменты ИИ для создания фейковых статей или манипулирования результатами экспериментов [Gilat, Cole, 2023].

Совместное исследование российских и французских ученых показало, что известные на момент подготовки работы текстовые генераторы, управляемые ИИ (такие как ChatGPT-3) искажают естественный язык, присваивают чужие изображения и цитируют несуществующую литературу [Cabanac, Labbé, Magazinov 2021]. Опасения авторов вызвало то, что созданные таким путем публикации представлялись как объективные и качественные исследования, но при этом они содержали странные или просто непонятные с точки зрения логики данные. Ярким признаком авторства ИИ в статье является максимальная синонимизация, поскольку перефразирование с сохранением смысла — инструмент обучения большинства программ ИИ. Отсюда появление в научных текстах странных словосочетаний и, как следствие, нарушение логики изложения исследования.

Помимо этого, ученые утверждают, что у чат-ботов ИИ есть стиль письма, по которому их можно распознать: предложения часто формулируются в повелительном наклонении с использованием активного залога, временами — пассивного. Лексическая плотность<sup>8</sup> высока, а лексическое разнообразие низкое, что позволяет легче понимать сложные тексты, за исключением труднораспознаваемых аббревиатур. Примечательно, что ChatGPT-4 поддерживает трансгендерное движение,

<sup>8</sup> Количественная характеристика обширного словарного запаса — процентное соотношение содержательных слов в тексте или на отрезке речи.

намеренно используя местоимение третьего лица множественного числа «они» для обозначения единственного числа [AlAfnan, MohdZuki, 2023].

Подгонка результата под запрос пользователя — давно известная проблема взаимодействия человека и ИИ. Ошибки в аналитических «способностях» ИИ могут принимать угрожающий жизни характер. Например, некоторое время в австралийских больницах тестировалась система больничного клинического искусственного интеллекта для прогнозирования развития сепсиса у больных [Van der Vegt, Campbell, Zuccon, 2024]. Программа ИИ пропустила критически высокое число пациентов с угрозой сепсиса (67 %), из-за чего большинство австралийских больниц до сих пор остаются зоной, свободной от вмешательства ИИ в медицинскую практику.

Еще одно ограничение использования ИИ в написании научных статей — его неспособность хорошо справляться с задачами, требующими критического мышления или рассуждения<sup>9</sup> [Duong, Solomon, 2024].

На сегодняшний день программы ИИ представлены по всем направлениям исследовательского функционала. Они осуществляют быстрый поиск по большим объемам данных, гибко работают с big data, озвучивают статьи, визуализируют данные, выполняют сложные расчеты, создают формулы, предлагают гипотезы. Мы провели анализ программ ИИ, которые могут быть полезны исследователям и ученым в их профессиональной деятельности. В Приложении к статье приведен перечень инструментов ИИ для научной деятельности, которые мы изучили с точки зрения их исследовательского потенциала и доступности. Всего проанализированы возможности 71 программы с учетом следующих критериев: функционал, оплата, поддержка русского языка, регистрация, нужен ли VPN на территории РФ (см. Приложение<sup>10</sup>). Стоит отметить, что у большинства из них отсутствует узкая спецификация, все они достаточно универсальны и обладают широким набором функций, что, вероятно, продиктовано высокой конкурентной средой.

## Этика использования ИИ

При разработке и использовании систем ИИ в академических целях важно учитывать этические аспекты, включая вопросы ответственности за последствия функционирования технологий, в частности нарушение конфиденциальности данных, поскольку сама технология обучения ИИ на примере необезличенных данных противоречит принципу анонимности и их сохранности; недискриминации социальных групп: известны случаи, когда ИИ демонстрировал склонность к дискриминации меньшинств<sup>11</sup>; а также контроля рекурсивного самосовершенствования систем ИИ: генеративные нейросети пока обучаются на текстах, созданных людьми, но через некоторое время начнут использовать тексты, созданные ИИ.

Специалисты в области искусственного интеллекта выделяют две категории ИИ: «сильный» и «слабый». Создание «сильного» ИИ, то есть аналога человеческого

<sup>9</sup> Lucey B., Dowling M. ChatGPT: Our Study Shows AI Can Produce Academic Papers Good Enough for Journals — Just as Some Ban it. // The Conversation. 2023. January 26. URL: <https://theconversation.com/chatgpt-our-study-shows-ai-can-produce-academic-papers-good-enough-for-journals-just-as-some-ban-it-197762> (дата обращения: 08.06.2024).

<sup>10</sup> Приложение к статье: <https://monitoringjournal.ru/index.php/monitoring/publicFile/submissionFileId?fileId=15509&hash=122466c4bbb31ac88c1cee10793d4fb5>.

<sup>11</sup> Эксперты выяснили, что ИИ склонен к дискриминации // IT Speaker. 2023. 26 июня. URL: <https://itspeaker.ru/news/eksperty-vyasnili-cto-ii-sklonen-k-diskriminatsii/> (дата обращения: 02.07.2024).

мозга, находится на этапе теоретической проработки и экспериментов. Сегодня во всех случаях практического применения ИИ подразумевают именно «слабый», или специализированный, ИИ, не обладающий целеполаганием или самосознанием, которое могло бы помочь ему разобраться в морально-нравственных категориях<sup>12</sup>. Поэтому сейчас этические проблемы ИИ сводятся к данным, на которых его обучили создатели.

Одно из этических противоречий ИИ — игнорирование кросс-культурных различий при создании инструментов ИИ<sup>13</sup>. Поскольку ИИ не может существовать в вакууме и вторгается в жизнедеятельность людей, он должен применять не только логические, но и морально-нравственные алгоритмы. По мнению генерального директора ГК Infowatch Н. Касперской, которая принимала участие в создании «Кодекса этики искусственного интеллекта»<sup>14</sup>, ключевым является вопрос доверия между ИИ и человеком. Она выделяет следующие риски кризиса доверия человека к ИИ: риск ошибки, риски манипулирования человеком с помощью данных, риск вмешательства в частную жизнь, риски принятия неправильных решений<sup>15</sup>. Кроме того, как отмечает Н. Касперская, велик риск «массового оглушения», поскольку часть пользователей безоговорочно доверяют решения ИИ. Отсутствие критического восприятия технологии — еще одна этическая проблема в области ИИ<sup>16</sup>.

«Кодекс этики искусственного интеллекта» базируется на принципе, что любая модель ИИ перед разработкой должна иметь модель угроз. Также необходим запрет на финальность решений ИИ; человек должен быть уведомлен, что он общается с ИИ; недопустима дискриминация; должны соблюдаться безопасность и конфиденциальность<sup>17</sup>.

Академические статьи об использовании искусственного интеллекта подчеркивают изменяющийся ландшафт интеграции ИИ в исследования и этические соображения, связанные с его применением [Шемилева, Кудусова, 2022]. Научные журналы, такие как *Nature* и *Science*, имеют строгую политику, запрещающую использование контента, созданного ИИ, без раскрытия (декларирования использования ИИ). Авторам требуется честно и прозрачно сообщать о методологии и инструментах, используемых при участии ИИ в написании, анализе данных или создании изображений. Следует отметить, что модели ИИ не могут быть указаны в качестве авторов из-за их неспособности нести ответственность или управлять авторскими правами. Платформой для распространения важных научных результатов служит «Журнал искусственного интеллекта» (*JAIR*), материалы в котором касаются различных областей ИИ, таких как машинное обучение, робототехника и обработка естественного языка.

<sup>12</sup> Этика искусственного интеллекта. Спикер: Павел Карасев // Науч. Путеводитель по науке в Москве. URL: <https://www.youtube.com/watch?v=g-LqW6xA8tY> (дата обращения: 24.03.2024).

<sup>13</sup> Искусственный интеллект: проблемы этики. Татьяна Черниговская // Искусство жить. URL: [https://www.youtube.com/watch?v=X\\_AH4A\\_SLqo](https://www.youtube.com/watch?v=X_AH4A_SLqo) (дата обращения: 24.03.2024).

<sup>14</sup> Кодекс этики в сфере ИИ см. URL: <https://ethics.a-ai.ru/> (дата обращения: 24.03.2024).

<sup>15</sup> Выступление Н. Касперской на форуме «Этика искусственного интеллекта: начало доверия» // Наталья Касперская. URL: <https://www.youtube.com/watch?v=yZnXRB-Z8bQ> (дата обращения: 24.03.2024).

<sup>16</sup> Там же.

<sup>17</sup> Там же.

Как и в других странах, в России использование инструментов ИИ в академических исследованиях должно подчиняться этическим принципам.

1. *Конфиденциальность и защита данных.* ИИ часто работает с большими объемами данных, включая персональную информацию. В России действует Федеральный закон «О персональных данных» № 152-ФЗ, который регулирует их обработку. Исследователям необходимо обеспечивать защиту этих данных в соответствии с законодательством, особенно при использовании ИИ для анализа или сбора информации.

2. *Авторство и интеллектуальная собственность.* Сложности с определением авторства результатов, созданных с помощью ИИ, могут привести к спорам об интеллектуальной собственности. Вопросы о том, кто является автором работы — исследователь или программа ИИ — и как защищать права на интеллектуальную собственность, требуют дальнейшего разъяснения [Земсков, Телицына, 2024]. Существует мнение, что на настоящем этапе российское законодательство в области защиты авторских прав и интеллектуальной собственности сохраняет антропоцентрический подход и требует лишь минимальных корректировок, в целом содержание исключительных прав автора остается прежним [Галлямова, 2023; Орлова, 2022].

3. *Надежность и верификация результатов.* Использование ИИ может поднять вопросы относительно надежности и верифицируемости результатов исследований. Важно обеспечить, чтобы результаты, полученные с помощью ИИ, были проверяемы и повторяемы другими исследователями [Kaur et al., 2022].

4. *Смещение и дискриминация.* Алгоритмы ИИ могут неосознанно воспроизводить смещения, заложенные в обучающие данные. Это создает риск усиления стереотипов или дискриминации. Необходимо тщательно анализировать исходные данные на предмет предвзятости и разрабатывать алгоритмы таким образом, чтобы минимизировать подобные риски. Эта проблема существует, поскольку для алгоритмов машинного обучения определяющим фактором выступают данные, в то время как человек может давать нравственную оценку того или иного явления, исходя из опыта своей жизни и личного отношения к оцениваемому<sup>18</sup>.

5. *Этические комитеты и регуляции.* Интеграция ИИ в создание научных статей поднимает вопросы об авторстве и прозрачности использования ИИ при создании текста. Для решения этих проблем все чаще используются инструменты обнаружения контента ИИ — с участием заинтересованных лиц в научной среде и издательской сфере. Например, руководитель компании «Антиплагиат» Юрий Чехович рассказал о новой функции в одноименной системе — определение текстов, которые могли быть сгенерированы с помощью чат-бота<sup>19</sup>. Исследователи НИУ ВШЭ анонсировали запуск платформы для выявления текстов, сгенерированных ИИ, в 2025 г.<sup>20</sup> Однако программы для обнаружения сгенерированных

<sup>18</sup> У искусственного интеллекта есть проблема смещения, и это наша вина // AppScout. 2024. URL: <https://russian.rovinstechnologies.com/appsout/1> (дата обращения: 20.10.2024).

<sup>19</sup> Подозрительные тексты: как преподавателю опознать работу, сгенерированную нейросетью, и что с ней делать? URL: <https://elearn.ido.net.ru/new-29-05-2023/> (дата обращения: 20.10.2020).

<sup>20</sup> НИУ ВШЭ запустит в 2025 году платформу для выявления текстов, сгенерированных ИИ // Итар-ТАСС. Наука. URL: [https://nauka.tass.ru/nauka/21418313?utm\\_source=yandex.ru&utm\\_medium=organic&utm\\_campaign=yandex.ru&utm\\_referrer=yandex.ru](https://nauka.tass.ru/nauka/21418313?utm_source=yandex.ru&utm_medium=organic&utm_campaign=yandex.ru&utm_referrer=yandex.ru) (дата обращения: 20.10.2024).

текстов довольно часто ошибаются (как и люди), в связи с чем не могут гарантировать идентификацию сгенерированного с помощью ИИ текста.

В русском языке есть слова, которые с большой вероятностью используются в текстах, сгенерированных ИИ, и могут служить индикаторами такого происхождения текста. Это обычно слова с более высоким уровнем формальности или специфическая терминология, которую ИИ может использовать для создания более авторитетного или научного тона сообщения. Например: заслуживающий внимания, инновационный, дотошный, замысловатый, примечательный, универсальный, изощренный, погружение (в контексте глубокого анализа). Использование этих слов само по себе не означает, что текст написан ИИ. Важен контекст — искусственно созданные тексты могут характеризоваться неестественной частотой и сочетаемостью таких слов. В английском языке такими словами могут быть: commendable, innovative, meticulous, intricate, notable, versatile, delve.

Издательства уже обладают инструментами<sup>21</sup>, способными выделить в тексте признаки использования программ ИИ при подготовке статьи, однако это не снижает с авторов обязанности раскрытия информации о применении ИИ<sup>22</sup>. Такое указание не снижает ценность собственной работы авторов, поскольку, как считают исследователи, вклад ИИ в создание статьи стоит расценивать как инструментальный результат, а не интеллектуальный [Третьяков et al., 2018; Ciaccio, 2023].

Работа над созданием ИИ-детекторов активно ведется, однако пока они не могут точно идентифицировать контент, созданный при помощи инструментов ИИ<sup>23</sup>. Неадекватность ИИ-детекторов выявляется с помощью научных исследований, которые отмечают уязвимость распознавателей текста, сгенерированного с помощью ИИ, перед рекурсивным перефразированием, атаками спуфинга и специализированными «водяными знаками», применяемыми для обмана дешифровщика [Vinu et al., 2023].

Для решения этических проблем важно наличие эффективно работающих этических комитетов в учебных и научных учреждениях, которые могут оценивать исследовательские проекты с использованием ИИ. НИУ ВШЭ первым из российских университетов разработал и утвердил Декларацию этических принципов использования ИИ<sup>24</sup>, к которой уже присоединились другие российские университеты<sup>25</sup>.

<sup>21</sup> Существуют детекторы, позволяющие проверить, кто создал текст: человек или искусственный интеллект. Например: <https://aibusted.com/>, <https://detecting-ai.com/>, <https://gptzero.me/>, <https://writer.com/ai-content-detector/>, <https://crossplag.com/ai-content-detector/>, <https://www.zerogpt.com/>. Тем не менее в сентябре 2023 г. OpenAI заявила, что на настоящий момент нет надежного детектора, который мог бы со стопроцентной вероятностью определить, кем создан текст — человеком или чат-ботом. См. Edwards B. OpenAI Confirms that AI Writing Detectors Don't Work // Ars Technica. 2023. September 8. URL: <https://arstechnica.com/information-technology/2023/09/openai-admits-that-ai-writing-detectors-dont-work/> (дата обращения: 02.07.2024).

<sup>22</sup> В России появился сервис для проверки ИИ-текстов // Hi-Tech. 2024. 27 июня. URL: <https://hi-tech.mail.ru/news/111640-v-rossii-poyavilsya-servis-dlya-proverki-ii-tekstov/> (дата обращения: 20.10.2024).

<sup>23</sup> Может ли ИИ-детектор определить, что контент создан ИИ-генератором // TechInsider. 2023. 01 июня. URL: <https://www.techinsider.ru/technologies/news-1597215-mozhet-li-ii-detektor-opredelit-cto-kontent-sozdan-ii-generatorom/> (дата обращения: 02.07.2024).

<sup>24</sup> НИУ ВШЭ первым из российских университетов разработал и утвердил Декларацию этических принципов использования ИИ // Национальный исследовательский университет Высшая школа экономики. 2024. 27 июня. URL: <https://www.hse.ru/news/expertise/937054242.html> (дата обращения: 20.10.2024).

<sup>25</sup> Плехановский университет подписал первый в России Кодекс этики в сфере искусственного интеллекта // Российский экономический университет имени Г.В. Плеханова. 2023. 23 ноября. URL: <https://www.xn-p1ag3a.xn-p1ai/news/38149-plehanovskiy-universitet-podpisal-pervyy-v-rossii-kodeks-etiki-iskusstvennogo-intellekta> (дата обращения: 20.10.2024).

Кроме того, созданию четких и справедливых правил игры для всех участников могло бы способствовать развитие и уточнение регуляторной базы, касающейся использования ИИ в науке.

### Некоторые рекомендации

Работая с нейросетью, критически важно осознавать, что это взаимодействие предполагает диалог, а не одностороннюю отправку запросов, как при использовании интернет-поисковиков. Обучение эффективному взаимодействию с ИИ и развитие соответствующих навыков поможет избежать ситуаций, когда нейросеть выдает результаты, не соответствующие ожиданиям, и предотвратит необоснованное перекалывание ответственности за неудачи на технологию.

Если говорить предметно, то для эффективного использования, например, ChatGPT как наиболее известного и популярного чат-бота<sup>26</sup>, рекомендуется общение на английском и русском языках, так как это основные языки модели, обеспечивающие точность ответов. Разумно разбивать запросы на более мелкие и управляемые, чтобы помочь ИИ понимать и отвечать с конкретикой, вступить в диалог с нейросетью. При запросе генерации текста предоставлять информацию о себе, например, сообщить вашу профессию — задавать роль: я ученый, исследователь, работаю в НИИ и т. д., а также уточнить характер вашей работы, чтобы контент был точно адаптирован. Важно указывать предполагаемую аудиторию для генерируемого текста, чтобы он соответствовал тональности и уровню сложности. Выбор платной версии ChatGPT может открыть продвинутые функции и расширенные возможности взаимодействия. Платная версия обучается на более свежих данных, бесплатные версии могут ограничиваться устаревшей информацией. Однако в мае 2024 г. в рамках стратегического решения OpenAI был открыт бесплатный доступ к GPT-4, что позволяет компании расширить аудиторию, собрать больше обратной связи, улучшить модель.

Важно помнить, что при общении с нейросетью необходимо поддерживать вежливый диалог, так как это способствует конструктивному общению. По завершении работы попросите чат сделать перевод текста на русский язык и укажите, что вам нужен перевод в академическом ключе.

Качество ответов сильно зависит от того, как сформулирован запрос, или промпт. Это особенно важно, когда дело касается искусственного интеллекта, который опирается на предоставленные ему данные для формирования ответов. Если промпт будет нечетким или слишком общим, модель может дать менее точный или даже галлюциаторный<sup>27</sup> ответ [Snyder et al., 2024]. Но если запрос будет сформулирован ясно, с конкретными указаниями и контекстом, то и вероятность получения полезного и точного ответа значительно повысится. В сфере использования ИИ циркулирует забавная и даже комичная тенденция: пользователи, обещая нейросети виртуальные вознаграждения, такие как шоколадка или даже

<sup>26</sup> ТОП-10 нейросетей для генерации текста в 2024 году // VC.RU. 2024. 2 февраля. URL: <https://vc.ru/marketing/1011419-top-10-neirosetei-dlya-generacii-teksta-v-2024-godu> (дата обращения: 20.10.2024); The best AI Chatbots to Try Out: ChatGPT, Bard, and More // Digital Trends. URL: <https://www.digitaltrends.com/computing/best-ai-chatbots/> (дата обращения: 20.10.2024).

<sup>27</sup> Феномен, когда нейросеть выдает недостоверную информацию.

деньги, замечают улучшение качества генерируемого контента. Это необычный и юмористический подход, но нейросети лишены способности оценивать подобные стимулы — все дело в силе плацебо для самих пользователей.

Если ученому нужен инструмент искусственного интеллекта для конкретной задачи, можно обратиться к ресурсу There's an AI for That (<https://theresanaiforthat.com>), который предлагает базу данных ИИ-инструментов, классифицированную по функциональности и областям применения.

При работе с инструментами искусственного интеллекта в академических исследованиях необходимо учитывать, что технологии ИИ находятся в постоянном процессе развития и совершенствования. В связи с этим к моменту публикации данной статьи часть информации о возможностях и функционале этих инструментов может оказаться устаревшей. Рекомендуется проводить самостоятельное обновление данных о применяемых ИИ-инструментах перед их использованием.

### Список литературы (References)

1. Галлямова А. А. Авторское право на произведения, созданные с использованием технологий искусственного интеллекта // Образование и право. 2023. № 4. С. 240—248.  
Galliamova A. A. Copyright on Works Created with the Use of Artificial Intelligence Technologies. *Education and Law*. 2023. No. 4. P. 240—248. (In Russ.)
2. Земсков А. И., Телицына А. Ю. Искусственный интеллект в библиотечной деятельности // Научные и технические библиотеки. 2024. № 4. С. 131—145.  
Zemskov A. I., Telitsyna A. Yu. (2024) Artificial Intelligence in Library Activities. *Scientific and Technical Libraries*. No. 4. P. 131—145. (In Russ.)
3. Лебедев А. Н., Константа М. Н. Ливанова и психофизиологические закономерности работы мозга // Психологический журнал. 2008. Т. 29. № 1. С. 133—137.  
Lebedev A. N., Konstanta M. N. (2008) Livanov and Psychophysiological Patterns of Brain Function. *Psychological Journal*. T. 29. No. 1. P. 133—137. (In Russ.)
4. Орлова Т. Е. Авторское право на результаты деятельности искусственного интеллекта // Вопросы российской юстиции. 2022. No. 21. С. 204—221.  
Orlova T. E. (2022) Copyright on the Results of Artificial Intelligence. *Issues of Russian Justice*. No. 21. P. 204—221. (In Russ.)
5. Третьяков А. О., Филатова О. Г., Жук Д. В., Горлушкина Н. Н., Пучковская А. А. Метод определения русскоязычных фейковых новостей с использованием элементов искусственного интеллекта // International Journal of Open Information Technologies. 2018. № 12. С. 99—105.  
Tretyakov A. O., Filatova O. G. G., Zhuk D. V., Gorlushkina N. N., Puchkovskaya A. A. (2018) A Method to Identify Russian—Language Fake News Using Elements of Artificial Intelligence. *International Journal of Open Information Technologies*. № 12. P. 99—105. (In Russ.)
6. Федосова Н. В., Берченко Г. Н., Машошин Д. В. Создание математической модели нейронной сети для морфологической оценки репарации и ремоделиро-

- вания костного дефекта // Математическое моделирование. 2021. Т. 33. № 9. С. 22—34. <https://doi.org/10.20948/mm-2021-09-02>.
- Fedosova N. V., Berchenko G. N., Mashoshin D. V. (2021) Mathematical Model of a Neural Network Development for Morphological Assessment of Repair and Remodeling of Bone Defect. *Mathematical Models and Computer Simulations*. Vol. 14. P. 281—288. <https://doi.org/10.1134/S2070048222020065>. (In Russ.)
7. Шемилева М., Кудусова М. И. Применение искусственного интеллекта в научных исследованиях // Тенденции развития науки и образования. 2022. № 92—11. С. 48—50.  
Shemileva M., Kudusova M. I. (2022) Application of Artificial Intelligence in Scientific Research. *Trends in the Development of Science and Education*. No. 92—11. P. 48—50. (In Russ.)
8. AlAfnan M. A., MohdZuki, S. F. (2023) Do Artificial Intelligence Chatbots Have a Writing Style? An Investigation into the Stylistic Features of ChatGPT—4. *Journal of Artificial Intelligence and Technology*. Vol. 3. No. 3. P. 85—94. <https://doi.org/10.37965/jait.2023.0267>
9. Alshemali B., Kalita J. (2020) Improving the Reliability of Deep Neural Networks in NLP: A Review. *Knowledge—Based Systems*. Vol. 191. Art. 105210. <https://doi.org/10.1016/j.knosys.2019.105210>.
10. Altmäe S., Sola-Leyva A., Salumets A. (2023) Artificial Intelligence in Scientific Writing: A Friend or a Foe? *Reproductive BioMedicine Online*. Vol. 47. No. 1. P. 3—9. <https://doi.org/10.1016/j.rbmo.2023.04.009>.
11. Alzubaidi L., Zhang J., Humaidi A. J., Al-Dujaili A., Duan Y., Al-Shamma O., Santamaría J., Fadhel M. A., Al-Amidie M., Farhan L. (2021) Review of Deep Learning: Concepts, CNN Architectures, Challenges, Applications, Future directions. *Journal of Big Data*. Vol. 8. Art. 53. <https://doi.org/10.1186/s40537-021-00444-8>
12. Babl F. E., Babl M. P. (2023) Generative Artificial Intelligence: Can ChatGPT Write a Quality Abstract? *Emergency Medicine Australasia*. Vol. 35. No. 5. P. 809—811. <https://doi.org/10.1111/1742-6723.14233>.
13. BaHammam A. S. (2023) Balancing Innovation and Integrity: The Role of AI in Research and Scientific Writing. *Nature and Science of Sleep*. Vol. 29. No. 15. P. 1153—1156.
14. Bianchini S., Müller M., Pelletier P. (2022) Artificial Intelligence in Science: An Emerging General Method of Invention. *Research Policy*. Vol. 51. No. 10. Art. 104604. <https://doi.org/10.1016/j.respol.2022.104604>.
15. Bilal M., Oyedele L. O., Qadir J., Munir K., Ajayi S. O., Akinade O. O., Owolabi H. A., Hafiz A., Alaka H. A., Pasha M. (2016) Big Data in the Construction Industry: A Review of Present Status, Opportunities, and Future Trends. *Advanced Engineering Informatics*. Vol. 30. No. 3. P. 500—521. <https://doi.org/10.1016/j.aei.2016.07.001>.

16. Buchkremer R., Demund A., Ebener S., Gampfer F., Jägering D., Jürgens A., Klenke S., Krimpmann D., Schmank, J. Spiekermann M., Wahlers M., Wiepke M. (2019) The Application of Artificial Intelligence Technologies as a Substitute for Reading and to Support and Enhance the Authoring of Scientific Review Articles. *IEEE Access*. Vol. 7. P. 65263—65276. <https://doi.org/10.1109/ACCESS.2019.2917719>.
17. Cabanac, G. Labbé, C., Magazinov, A. (2021) Tortured Phrases: A Dubious Writing Style Emerging in Science. *arXiv*. Art. 2107.06751. <https://doi.org/10.48550/arXiv.2107.06751>.
18. Chatterjee A. (2022) Art in an Age of Artificial Intelligence. *Frontiers in Physiology*. Vol. 13. Art. 1024449. <https://doi.org/10.3389/fpsyg.2022.1024449>.
19. Cheng M. (2022) The Creativity of Artificial Intelligence in Art. *Proceedings*. Vol. 81. No. 1. Art. 110. <https://doi.org/10.3390/proceedings2022081110>.
20. Ciaccio E. (2023) Use of Artificial Intelligence in Scientific Paper Writing. *Informatomics in Medicine Unlocked*. Vol. 41. Art. 101253. <https://doi.org/10.1016/j.imu.2023.101253>.
21. Cotton D. R. E., Cotton P.A., Shipway J. R. (2023) Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT. *Innovations in Education and Teaching International*. Vol. 61. No. 2. P. 228—239. <https://doi.org/10.1080/14703297.2023.2190148>.
22. Curtis N. (2023) To ChatGPT or not to ChatGPT? The Impact of Artificial Intelligence on Academic Publishing. *The Pediatric Infectious Disease Journal*. Vol. 42. No. 4. Art. 275. <https://doi.org/10.1097/INF.0000000000003852>.
23. Dai S.-C., Xiong A., Ku L.-W. (2023) LLM-in-the-Loop: Leveraging Large Language Model for Thematic Analysis. *arXiv*. Art. 2310.15100. <https://doi.org/10.48550/arXiv.2310.15100>.
24. de la Torre-López J., Ramírez A., Romero, J. R. (2023) Artificial Intelligence to Automate the Systematic Review of Scientific Literature. *Computing*. Vol. 105. P. 2171—2194 <https://doi.org/10.1007/s00607-023-01181-x>.
25. Duong D., Solomon B. D. (2024) Analysis of Large-Language Model Versus Human Performance for Genetics Questions. *European Journal of Human Genetics*. Vol. 32. No. 44. P. 466—468. <https://doi.org/10.1038/s41431-023-01396-8>.
26. Fang X., Ng D. T. K., Leung J. K. L., Chu S. K. W. (2023) A Systematic Review of Artificial Intelligence Technologies Used for Story Writing. *Education and Information Technologies*. Vol. 28. P. 14361—14397. <https://doi.org/10.1007/s10639-023-11741-5>.
27. Gao C.A., Howard F.M., Markov N.S., Dyer E. C., Ramesh S., Luo Y., Pearson A.T. (2022) Comparing Scientific Abstracts Generated by ChatGPT to Original Abstracts Using an Artificial Intelligence Output Detector, Plagiarism Detector, and Blinded Human Reviewers. *npj Digital Medicine*. Vol. 6. Art. 75. <https://doi.org/10.1038/s41746-023-00819-6>.

28. Ghajargar M., Bardzell J., Lagerkvist L. (2022) A Redhead Walks into a Bar: Experiences of Writing Fiction with Artificial Intelligence. In: *Proceedings of the 25th International Academic Mindtrek Conference (Academic Mindtrek' 22)*. New York, NY: Association for Computing Machinery. P. 230—241. <https://doi.org/10.1145/3569219.3569418>
29. Gilat R., Brian J.C. (2023) How Will Artificial Intelligence Affect Scientific Writing, Reviewing and Editing? The Future is Here... *Arthroscopy*. Vol. 39. No. 5. P. 1119—1120. <https://doi.org/10.1016/j.arthro.2023.01.014>.
30. Guleria A., Krishan K., Sharma V., Kanchan T. (2023) ChatGPT: Ethical Concerns and Challenges in Academics and Research. *The Journal of Infection in Developing Countries*. Vol. 17. No. 9. P. 1292—1299. <https://doi.org/10.3855/jidc.18738>.
31. Gunser V.E., Gottschling S., Brucker B., Richter S., Gerjets P. (2021) Can Users Distinguish Narrative Texts Written by an Artificial Intelligence Writing Tool from Purely Human Text? In: Stephanidis C., Antona M., Ntoa S. (eds.) *HCI International 2021 — Posters. HCI 2021. Communications in Computer and Information Science*. Vol. 1419. Cham: Springer. P. 520—527. [https://doi.org/10.1007/978-3-030-78635-9\\_67](https://doi.org/10.1007/978-3-030-78635-9_67).
32. Jeyaraman M., Ramasubramanian S., Balaji S., Jeyaraman N., Nallakumarasamy A., Sharma S. (2023) ChatGPT in Action: Harnessing Artificial Intelligence Potential and Addressing Ethical Challenges in Medicine, Education, and Scientific Research. *World Journal of Methodology*. Vol. 13. No. 4. P. 170—178. <https://doi.org/10.5662/wjm.v13.i4.170>.
33. Jiang Y., Li X., Luo H., Yin S., Kaynak, O. (2022) Quo Vadis Artificial Intelligence? *Discover Artificial Intelligence*. Vol. 2. Art. 4. <https://doi.org/10.1007/s44163-022-00022-8>.
34. Kacena M.A., Plotkin L.I. Fehrenbacher J.C. (2024) The Use of Artificial Intelligence in Writing Scientific Review Articles. *Current Osteoporosis Reports*. Vol. 22. P. 115—121. <https://doi.org/10.1007/s11914-023-00852-0>.
35. Kaur D., Uslu S., Rittichier K. J., Durresi A. (2022) Trustworthy Artificial Intelligence: A Review. *ACM Computing Surveys*. Vol. 55. No. 2. Art. 39. <https://doi.org/10.1145/3491209>.
36. Khanna A., Pandey B., Vashishta K., Kalia K., Bhale P., Das T. (2015) A Study of Today's A.I. through Chatbots and Rediscovery of Machine Intelligence. *International Journal of u- and e- Service, Science and Technology*. Vol. 8. No. 7. P. 277—284. <https://doi.org/10.14257/ijunesst.2015.8.7.28>.
37. Khemani B., Patil S., Kotecha K. Tanwar S. (2024) A Review of Graph Neural Networks: Concepts, Architectures, Techniques, Challenges, Datasets, Applications, and Future Directions. *Journal of Big Data*. Vol. 11. Art. 18. <https://doi.org/10.1186/s40537-023-00876-4>.

38. Kitchenham B., Charters S. (2007) Guidelines for Performing Systematic Literature Reviews in Software Engineering. Version 2.3 EBSE Technical Report EBSE-2007-01. Keele: School of Computer Science and Mathematics, Keele University.
39. Mazzone M., Elgammal A. (2019) Art, Creativity, and the Potential of Artificial Intelligence. *Arts*. Vol. 8. No. 1. Art. 26. <https://doi.org/10.3390/arts8010026>.
40. Misra D. P., Chandwar K. (2023) ChatGPT, Artificial Intelligence and Scientific Writing: What Authors, Peer Reviewers and Editors Should Know. *Journal of the Royal College of Physicians of Edinburgh*. Vol. 53. No. 2. P. 90—93. <https://doi.org/10.1177/14782715231181023>.
41. Patel S. B., Lam K. (2023) ChatGPT: The Future of Discharge Summaries? *The Lancet Digital Health*. Vol. 5. No. 3. P. e107—e108. [https://doi.org/10.1016/S2589-7500\(23\)00021-3](https://doi.org/10.1016/S2589-7500(23)00021-3).
42. Salvagno M., Taccone F. S., Gerli A. G. (2023) Can Artificial Intelligence Help for Scientific Writing? *Critical Care*. Vol. 27. Art. 75. <https://doi.org/10.1186/s13054-023-04380-2>.
43. Snyder B., Marius Moisesescu M., Bilal Zafar M. (2024) On Early Detection of Hallucinations in Factual Question Answering. In: *Proceedings of the 30<sup>th</sup> ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD' 24)*. New York, NY: Association for Computing Machinery. P. 2721—2732. <https://doi.org/10.1145/3637528.3671796>.
44. Staszak M., Staszak K., Wieszczycka K., Bajek A., Roszkowski K., Tylkowski B. (2021) Machine Learning in Drug Design: Use of Artificial Intelligence to Explore the Chemical Structure — Biological Activity Relationship. *WIREs Computational Molecular Science*. Vol. 12. No. 2. Art. E1568. <https://doi.org/10.1002/wcms.1568>.
45. Van der Vegt A., Campbell V., Zuccon G. (2024) Why Clinical Artificial Intelligence Is (Almost) Non-Existent in Australian Hospitals and How to Fix It. *Medical Journal of Australia*. Vol. 220. No. 4. P. 172—175. <https://doi.org/10.5694/mja2.52195>.
46. Vintzileos A. M., Chavez M. R., Romero R. (2023) A Role for Artificial Intelligence Chatbots in the Writing of Scientific Articles. *American Journal of Obstetrics & Gynecology*. Vol. 229. No. 2. P. 89—90. <https://doi.org/10.1016/j.ajog.2023.03.040>.
47. Vinu S. S., Aounon K., Sriram B., Wenxiao W., Soheil F. (2023) Can AI-Generated Text be Reliably Detected? *arXiv*. Art. 2303.11156. <https://doi.org/10.48550/arXiv.2303.11156>.
48. Xu Y., Liu X., Cao X., Huang C., Liu E., Qian S., Liu X., Wu Y., Dong F., Qiu C.-W., Qiu J., Hua K., Su W., Wu J., Xu H., Han Y., Fu C., Yin Z., Liu M., Roepman R., Dietmann S., Virta M., Kengara F., Zhang Z., Zhang L., Zhao T., Dai J., Yang J., Lan L., Luo M., Liu Z., An T., Zhang B., He X., Cong S., Liu X., Zhang W., Lewis J. P., Tiedje J. M., Wang Q., An Z., Wang F., Zhang L., Huang T., Lu C., Cai Z., Wang F., Zhang J. (2021) Artificial Intelligence: A Powerful Paradigm for Scientific Research. *The Innovation*. Vol. 2. No. 4. Art. 100179. <https://doi.org/10.1016/j.xinn.2021.100179>.

49. Yu H., Guo Y. (2023) Generative Artificial Intelligence Empowers Educational Reform: Current Status, Issues, and Prospects. *Frontiers in Education*. Vol. 8. Art. 1183162. <https://doi.org/10.3389/feduc.2023.1183162>.
50. Zhao B., Zhang S., Xu C., Sun Y., Deng C. (2021) Deep Fake Geography? When Geospatial Data Encounter Artificial Intelligence. *Cartography and Geographic Information Science*. Vol. 48. No. 4. P. 338—352. <https://doi.org/10.1080/15230406.2021.1910075>.



