БЕЛАЯ КНИГА ЭТИКИ

в сфере искусственного интеллекта

42

ответа
на популярные
вопросы
развития
технологии ИИ

АЛЬЯНС
В СФЕРЕ
ИСКУССТВЕННОГО
ИНТЕЛЛЕКТА

Москва
2024

УДК 004.8
ББК 32.813
Б43

Коллектив авторов под ред. А. В. Незнамова

Б43 Белая книга этики в сфере искусственного интеллекта / под ред. А. В. Незнамова. — М.: Nova Creative Group, 2024. — 200 с.

ISBN 978-5-6052008-8-8

"Белая книга этики в сфере ИИ" — подробный и тщательно структурированный обзор самых острых этических вопросов, возникающих на пути развития искусственного интеллекта. В центре внимания — 42 ключевых вопроса, которые охватывают широкий спектр тем: от конфиденциальности данных и автономных решений до создания цифровых имитаций и ответственности за применение ИИ. Авторы предлагают читателям глубокий анализ современных технологий ИИ и раскрывают, как их развитие затрагивает моральные и социальные аспекты.

Эта книга не стремится дать окончательные ответы — скорее, она открывает пространство для диалога, где каждый из вопросов становится отправной точкой для размышлений и дискуссий. Издание адресовано ученым, педагогам, психологам, медицинским работникам, юристам, разработчикам и всем, кого волнует, как ИИ меняет наш мир и что это означает для будущего каждого из нас.

УДК 004.8
ББК 32.813

ISBN 978-5-6052008-8-8

Оглавление

О книге 6

Методология 8

О разделе. 14

Глава 1. Десять наиболее популярных этических проблем в связи с развитием ИИ 16

01. «Проблема вагонетки»: какой выбор между жизнями людей в случае неизбежного столкновения должен сделать беспилотный автомобиль? 18
02. Проблема цифровых имитаций человека: допустимо ли их создавать? 23
03. Проблема "черного ящика": как понять принципы работы систем ИИ и объяснить их пользователю? 28
04. Проблема информирования: всегда ли люди должны знать, что взаимодействуют с ИИ? 33
05. Проблема сокращения рабочих мест: приведет ли массовое внедрение ИИ к тому, что люди останутся без работы? 38
06. Проблема оспаривания: всегда ли человек должен иметь возможность оспорить решение, принятое с использованием ИИ? 43
07. Проблема предвзятости ИИ: можно ли ее решить? 48
08. Проблема ответственности: на примере медицины, какова ответственность разработчика ИИ в случае причинения вреда здоровью пациента? 52
09. Проблема делегирования принятия решений: на примере правосудия, сможет ли ИИ заменить судью? 57
10. Проблема социального рейтингования: этично ли применять ИИ для создания социального рейтинга? 62

Глава 2. ИИ и конфиденциальность 66

11. Этично ли использовать персональные данные для обучения ИИ? 68
12. Этично ли собирать данные пользователей со смартфона или умного устройства для обучения ИИ? 71
13. Этично ли применение ИИ в массовом видеонаблюдении? 74
14. Этично ли использовать ИИ для прогнозирования и предотвращения преступлений? 77
15. Этично ли применять ИИ для скоринга в ритейле, финансах и других отдельных сферах? 80

Глава 3. ИИ и достоверность 84
16. Проблема обучения: как избежать обучения ИИ на некорректной информации? 86
17. Проблема распространения вредоносной или вводящей в заблуждение информации с помощью ИИ: как с этим бороться? 89
18. Этично ли алгоритмам предлагать пользователю товары и услуги, которые не соответствуют его обычным предпочтениям? 92

Глава 4. Генеративный ИИ 96
19. Можно ли доверять информации, полученной при помощи генеративного ИИ и поисковиков на основе ИИ? 98
20. Этично ли синтезировать речь человека при помощи ИИ? 101
21. Этично ли использовать генеративный ИИ в искусстве и дизайне? 104
22. Этично ли не указывать, что контент сгенерирован с помощью ИИ? 107
23. Повлияет ли генеративный ИИ на стандарты красоты и на моду? 110

Глава 5. ИИ в сфере образования 114
24. Допустимо ли использование ИИ в образовательном процессе учащимися и преподавателями? 116
25. Этично ли преподавателю вести предмет через свою цифровую имитацию без личного присутствия в аудитории? 119
26. Этично ли писать курсовые работы или иные учебные работы с помощью ИИ? 121
27. Можно ли с помощью ИИ проверять работы учащихся? 124
28. Этично ли использовать системы прокторинга на основе ИИ? 127
29. Допустимо ли снижать оценку учащегося при подозрении на использование ИИ в учебе? 130
30. Этично ли ограничивать использование ИИ детьми в образовательных целях вне соответствующих учреждений? 133

Глава 6. ИИ и медицина 136
31. Этично ли человеку заниматься самолечением с помощью ИИ? 138
32. Этично ли врачу делегировать ИИ принятие решений по профилактике, диагностике, лечению и реабилитации? 141
33. Этично ли сообщать «плохие новости» пациенту с помощью ИИ? 144
34. Нужно ли получать отдельное согласие пациента на применение ИИ для лечения? 147

Глава 7. ИИ в правосудии 150
35. Этично ли применять ИИ судьям? 152
36. Этично ли использовать ИИ сторонам судебного дела? 155

Глава 8. ИИ и человек 158
37. Можно ли оказывать психологическую помощь с применением ИИ? 160
38. Нужно ли ограничивать темы и модерировать токсичный контент при общении с ИИ? 163
39. Этично ли формировать эмоциональную привязанность к ИИ? 166
40. Должен ли ИИ приносить публичные извинения, если оскорбит кого то? 169
41. Этично ли использовать ИИ-рекрутера? 171
42. Этично ли использовать ИИ в спорте для улучшения результатов? 174

Этическое сообщество в России 177

Благодарности 179

Список литературы к главе «Методология» 180
Список литературы к главе «О разделе» 180
Список литературы к главе 1 180
Список литературы к главе 2 186
Список литературы к главе 3 188
Список литературы к главе 4 189
Список литературы к главе 5 191
Список литературы к главе 6 193
Список литературы к главе 7 194
Список литературы к главе 8 195

О книге

Мы живем в удивительное время — эпоху, когда технологии развиваются стремительно, а искусственный интеллект уже не фантастика, а часть нашей реальности. Но в этом технологическом прогрессе возникает важный вопрос: как сделать так, чтобы развитие ИИ было безопасным, справедливым и этичным? И по каким правилам мы будем жить в обществе новых технологий?

С каждым шагом на пути развития технологий мы встречаем все больше сложных этических и социальных вопросов. Кто несет ответственность за решения, которые принимают автономные системы? Можно ли вообще делегировать ИИ принятие решений? Лишимся ли мы работы? Что происходит с конфиденциальностью наших данных? Является ли ИИ черным ящиком и как вообще нам правильно коммуницировать с системами ИИ и между собой?

Мы в Национальной комиссии по реализации Кодекса этики в сфере ИИ на базе Альянса в сфере ИИ решили собрать все самые актуальные вопросы и попытаться предложить на них ответы. Так появилась Белая книга по этике в сфере ИИ. Это не сборник научных размышлений, а руководство по навигации в сложном мире технологий, которые уже сегодня меняют нашу жизнь.

Для того чтобы помочь ответить на эти вопросы и предложить пути их решения, была создана Белая книга этики в сфере искусственного интеллекта. В Белой книге представлены ответы на наиболее актуальные этические вопросы, связанные с искусственным интеллектом, а также исследования по данной теме и практические рекомендации для минимизации этических рисков.

Что делает эту книгу по-настоящему уникальной? Прежде всего ее авторы. Это ведущие специалисты, ученые, педагоги, психологи, медицинские работники, юристы, разработчики — те, кто каждый день сталкивается с этическими вызовами ИИ на практике. Они предложили реальные решения, рекомендации и подходы к тому, как внедрять ИИ так, чтобы он работал на благо человека.

Мы рассмотрели самые разные конкретные вопросы — от моральных дилемм, как, например, знаменитая «проблема вагонетки», до вопросов использования ИИ в медицине, образовании и правосудии. Представьте себе, что ИИ будет принимать решение, на какое лечение направить пациента, или будет участвовать в судебных процессах. Это не будущее, это уже наша реальность. И наша книга показывает, как сделать эту реальность справедливой и безопасной.

Это издание действительно уникально. Впервые в одном месте собраны ключевые этические вопросы, связанные с использованием ИИ, и представлены возможные пути их решения. Однако важно отметить, что эта книга не претендует на универсальные ответы. Мы понимаем, что предложенные идеи могут показаться кому-то спорными или недостаточно убедительными. Поэтому мы приглашаем читателей к дискуссии, рассматривая книгу скорее, как отправную точку для размышлений.

Мы будем благодарны за ваши предложения и взгляд на этические дилеммы, поднятые в книге. Все поступившие отклики будут внимательно изучены и учтены при последующих изданиях. Кто знает, возможно, именно ваши отзывы и дальнейшее развитие технологий помогут переосмыслить многие из решений, предложенных в 2024 году.

Наши читатели — не только ученые и разработчики. Мы писали эту книгу для каждого, кто интересуется тем, как технологии меняют наш мир. Ведь это касается всех нас — от того, как ИИ оценивает нашу кредитоспособность, до того, как он помогает предотвращать преступления. Эти вопросы в буквальном смысле формируют наше будущее. Книга написана так, чтобы быть полезной и интересной широкой аудитории. Она полна реальных кейсов, конкретных рекомендаций и прогнозов о том, что ждет нас в ближайшие годы. Это не просто чтение — это диалог. Диалог с теми, кто стоит на передовой технологий, и с теми, кто задается вопросами, куда ведет нас этот стремительный прогресс.

И самое важное, что мы подчеркиваем на каждой странице: технологии — это инструмент. А вот какими они станут, зависит только от нас. И наша книга это шаг к осознанному, ответственному подходу к созданию будущего, где ИИ будет служить человеку, а не наоборот.

Вашим взглядом на этические проблемы можно поделиться здесь:

Методология

Как были отобраны вопросы:

Стремительное развитие ИИ и его повсеместное внедрение в жизнь человека и общества кардинально меняют мир. Это требует соблюдения этических принципов, способных уравновесить активное технологическое развитие и интересы человека, чтобы новые технологии служили на благо всего общества и человечества в целом. При составлении этой книги мы ориентировались на имеющиеся исследования, международные документы, данные опросов, а также позиции разработчиков и пользователей, собранные российской Комиссией по этике в сфере ИИ. К 2024 году в мире уже сформирована большая база этических вопросов в связи с развитием технологий ИИ, разработаны базовые этические принципы ИИ как на международном и национальном, так и на отраслевом уровне.

В 2019 году Всемирная комиссия по этике научных знаний и технологий опубликовала «Предварительное исследование по этике ИИ»[1], которое во многом было основано на опубликованном ранее «Исследовании по этике в робототехнике»[2] 2017 года. В нем был поднят ряд вопросов этики ИИ:

- роль ИИ в образовательном процессе, как инструменте цифровой обучающей среды, а также важность переподготовки работников и изменения набора квалификационных требований образовательных программ;
- прозрачность и объяснимость решений ИИ (способность ИИ анализировать большие объемы данных делает возможным его использование для мониторинга окружающей среды, прогнозирования стихийных бедствий, однако к обоснованности принятых им решений следует относиться с осторожностью);
- усиление предвзятости и неблагоприятное воздействие на уязвимые слои населения;
- влияние ИИ на языковое и культурное разнообразие (риск концентрации культурных ресурсов и данных у небольшого числа участников).

Организация экономического сотрудничества и развития (ОЭСР) в 2019 году опубликовала первый межправительственный стандарт в области ИИ «Рекомендации по искусственному интеллекту»[3]. ОЭСР официально закрепил пять принципов ответственного управления надежным ИИ:

- инклюзивный рост, устойчивое развитие и благополучие;
- верховенство закона, уважение прав человека и демократических ценностей, включая справедливость и конфиденциальность;

- прозрачность и объяснимость;
- надежность, безопасность и защищенность;
- подотчетность.

Этические рекомендации для надежного ИИ (ЕС) 2019 года[4] отмечают, что ИИ должен соответствовать критериям законности, этичности и надежности. Рекомендации закрепляют семь ключевых требований к надежному ИИ, включая контроль над ИИ со стороны человека, техническую надежность и безопасность, конфиденциальность данных, прозрачность, справедливость и недискриминацию, общественное и экологическое благополучие, а также подотчетность.

Важным этапом в развитии вопросов этики ИИ стала публикация в 2021 году «Рекомендаций по этике ИИ» ЮНЕСКО[5], первого международного стандарта, закрепляющего основополагающие 10 принципов этичного ИИ. Они включают в себя недискриминацию, защиту приватности и персональных данных, прозрачность алгоритмов, контроль со стороны человека и другие.

Кроме того, работы по изучению этических вопросов в сфере ИИ ведутся и на площадках международных органов по стандартизации. Так, в 2023 году Институт инженеров электротехники и электроники (IEEE) в рамках Программы по свободному доступу к стандартам в области этики и регулирования ИИ1 открыл доступ к ряду стандартов, прямо или косвенно посвященных этике ИИ, например, 7014—2024 — IEEE Standard for Ethical Considerations in Emulated Empathy in Autonomous and Intelligent Systems. (Стандарт IEEE по этическим вопросам эмулируемой эмпатии в автономных и интеллектуальных системах)[6].

В свою очередь Международная организация по стандартизации (ISO) опубликовала стандарт ISO/IEC TR 24368:2022 Information technology — Artificial intelligence — Overview of ethical and societal concerns (Искусственный интеллект. Обзор этических и общественных аспектов)[7]. Основываясь на нем был подготовлен российский аналогичный стандарт, в котором представлены высокоуровневый обзор этических и социальных (общественных) проблем искусственного интеллекта, а также основные принципы этичного ИИ.

Вопросы этики также затрагиваются в Стандарте ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system (Искусственный интеллект. Система управления)[8]. Данный стандарт предоставляет организациям рекомендации по решению таких проблем, как этика ИИ, прозрачность и непрерывное обучение.

В многочисленных публикациях по всему миру поднимаются схожие этические вопросы и проблемы, которые актуальны для человечества в контексте развития и внедрения искусственного интеллекта. Эти вопросы можно обобщить до нескольких ключевых тем:

- предвзятость и недискриминация;
- подконтрольность ИИ человеку;
- равенство при распределении благ от ИИ;
- вопросы трудоустройства и безработицы;

- защита данных и конфиденциальность;
- прозрачность и объяснимость алгоритмов ИИ;
- надежность ИИ;
- автономия человека и его свободного выбора;
- влияние ИИ на поведение человека и межличностное взаимодействие;
- а также общее обеспечение гарантий основных прав человека в контексте внедрения ИИ и многие другие аспекты.

Наконец, взаимодействуя с подписантами Кодекса этики ИИ, также удалось выяснить целый ряд этических вопросов. Многие из них были собраны и обсуждены в специальной рабочей группе, посвященной лучшим этическим практикам. Часть из этих вопросов на протяжении четырех лет обсуждались на всероссийских Форумах по этике ИИ — масштабных событиях, посвященных осмыслению и обсуждению вопросов этики ИИ.

Все эти материалы легли в основу книги.

Главный редактор:

Незнамов Андрей Владимирович — канд. юрид. наук, управляющий директор Центра человекоцентричного AI Сбербанка, председатель Комиссии по реализации Кодекса этики в сфере ИИ.

Научный редактор:

Крайнов Александр Георгиевич — директор по развитию технологий искусственного интеллекта «Яндекс».

Ответственные редакторы:

Фокина Софья Антоновна — менеджер Центра человекоцентричного AI Сбербанка, ответственный секретарь Комиссии по реализации Кодекса этики в сфере ИИ.

Черемных Ирина Александровна — аналитик Центра человекоцентричного AI Сбербанка.

В работе над книгой принимали участие:

Авакян Елена Георгиевна — вице-президент Федеральной Палаты адвокатов РФ, член Наблюдательного совета Ассоциации операторов информационных систем и операторов обмена цифровых финансовых активов, член рабочих групп госпрограммы «Цифровая экономика» центров компетенций фонда «Сколково», старший преподаватель департамента регулирования бизнеса факультета права НИУ ВШЭ.

Бырдин Алексей Юрьевич — директор Ассоциации «Интернет-видео».

Васин Евгений Олегович — исполнительный директор Центра человекоцентричного AI Сбербанка.

Вечерин Александр Викторович — канд. психол. наук, доцент Департамента психологии факультета социальный наук НИУ ВШЭ, академический руководитель образовательной программы «Психология», ведущий исследователь в проекте Центра искусственного интеллекта НИУ ВШЭ, посвященного поддержке разработчиков ИИ в сфере этики коммуникации ИИ.

Воробьев Андрей Павлович — канд. мед. наук, доцент, генеральный директор ГК «Ньюдиамед», основатель проекта MeDiCase.

Гребенщикова Елена Георгиевна — д-р филос. наук, директор Института гуманитарных наук, заведующая кафедрой биоэтики ИГН РНИМУ им. Н. И. Пирогова.

Емшанов Игорь Сергеевич — заместитель председателя Благовещенского городского суда Амурской области, председатель комиссии совета судей Амурской области по информатизации и автоматизации судов.

Киселев Антон Робертович — д-р мед. наук, профессор, заместитель директора по научно-технологическому развитию ФГБУ «Национальный медицинский исследовательский центр терапии и профилактической медицины» Минздрава России (ФГБУ «НМИЦ ТПМ» Минздрава России).

Крайнов Александр Георгиевич — директор по развитию технологий искусственного интеллекта «Яндекс»

Кулешов Андрей Александрович — аналитик Центра прикладных систем ИИ МФТИ.

Кулик Анна Петровна — директор по маркетингу «Инферит» (ГК Softline).

Магдиев Булат Илдусович — эксперт по работе с ИИ Департамента разработки и регистрации медицинских изделий Сеченовского университета.

Макарова Людмила Сергеевна — канд. филол. наук, доцент кафедры журналистики, заместитель директора Института филологии и журналистики ННГУ им. Н. И. Лобачевского.

Мацепуро Дарья Михайловна — канд. ист. наук, директор Сибирского (Томского) центра изучения искусственного интеллекта и цифровых технологий, директор Центра науки и этики ТГУ.

Минбалеев Алексей Владимирович — д-р юрид. наук, заведующий кафедрой информационного права и цифровых технологий Московского государственного юридического университета имени О. Е. Кутафина (МГЮА), доктор юридических наук, профессор, эксперт РАН.

Незнамов Андрей Владимирович — канд. юрид. наук, управляющий директор Центра человекоцентричного AI Сбербанка, председатель Комиссии по этике в сфере ИИ.

Павловский Евгений Николаевич — канд. физ.-мат. наук, ведущий научный сотрудник Исследовательского центра в сфере искусственного интеллекта НГУ.

Парамонова Елена Юрьевна — директор Департамента разработки и регистрации медицинских изделий Сеченовского университета.

Парфун Алексей Владимирович — co-founder ReFace Technology, вице-президент АКАР, сопредседатель комитета ИИ.

Померанцев Илья Валерьевич — CEO Celsor, директор АНО «Центр развития экспертизы ИТ», член комиссии

по реализации кодекса этики ИИ.

Суворов Валерий Владимирович — канд. ист. наук, научный сотрудник центра координации фундаментальной научной деятельности, НМИЦ терапии и профилактической медицины.

Сурагина Елена Дмитриевна — руководитель рабочей группы по созданию свода наилучших практик решения возникающих этических вопросов в жизненном цикле ИИ Комиссии по реализации Кодекса этики в сфере ИИ.

Тучинов Баир Николаевич — научный сотрудник Исследовательского центра в сфере искусственного интеллекта НГУ.

Углева Анастасия Валерьевна — канд. филос. наук, PhD, профессор Школы философии и культурологии, заместитель директора Центра трансфера и управления социально-экономической информацией НИУ ВШЭ.

Фабрус Игорь Владимирович — президент АНО «Институт искусственного интеллекта».

Фокина Софья Антоновна — менеджер Центра человекоцентричного AI Сбербанка, ответственный секретарь Комиссии по этике в сфере ИИ.

Хасанова Диана Магомедовна — доцент кафедры цифровых технологий в здравоохранении Казанского ГМУ, генеральный директор ООО «БРЕЙНФОН».

Чаче Эльвира Германовна — исполнительный директор Центра человекоцентричного AI Сбербанка.

Черемных Ирина Александровна — аналитик Центра человекоцентричного AI Сбербанка.

Чирва Дарья Викторовна — руководитель модуля «Мышление и философия», преподаватель Института международного развития и партнерства Университета ИТМО, научный сотрудник Центра сильного искусственного интеллекта в промышленности при Университете ИТМО.

Чумакова Мария Алексеевна — канд. психол. наук, руководитель департамента психологии факультета социальных наук НИУ ВШЭ, доцент департамента психологии факультета социальных наук НИУ ВШЭ, научный руководитель линейки дисциплин по анализу данных в рамках направления подготовки «Психология», руководитель проекта Центра искусственного интеллекта НИУ ВШЭ.

О разделе

В настоящем разделе представлены развернутые ответы на 10 самых популярных этических вопросов, связанных с развитием ИИ:

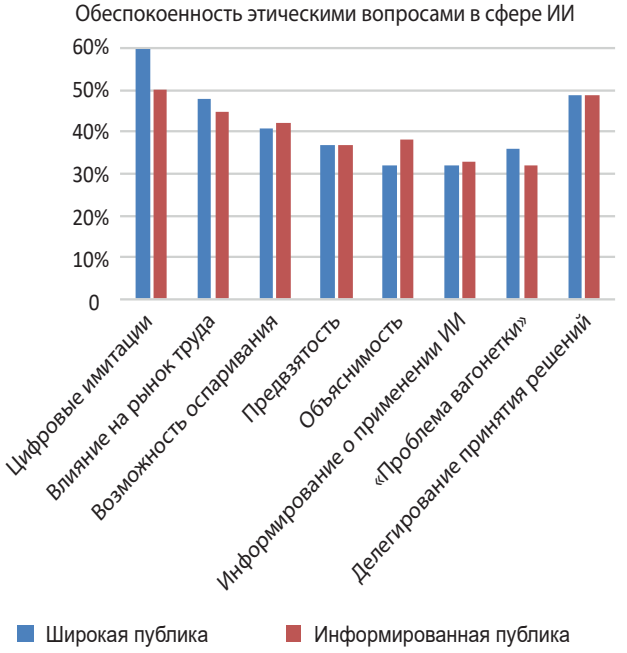
1. «Проблема вагонетки»: какой выбор между жизнями людей в случае неизбежного столкновения должен сделать беспилотный автомобиль?
2. Проблема цифровых имитаций человека: допустимо ли их создавать?
3. Проблема черного ящика: можно ли понять принципы работы систем ИИ и как объяснить их пользователю?
4. Проблема информирования: должны ли люди знать, что взаимодействуют с ИИ?
5. Проблема сокращения рабочих мест: приведет ли массовое внедрение ИИ к потере людьми работы?
6. Проблема оспаривания: всегда ли человек должен иметь возможность оспорить решение, принятое с использованием ИИ?
7. Проблема предвзятости: можно ли решить проблему предвзятости ИИ?
8. Проблема ответственности: на примере медицины, какова ответственность разработчика ИИ в случае причинения вреда здоровью пациента?
9. Проблема делегирования принятия решений: на примере правосудия, сможет ли ИИ заменить судью?
10. Проблема социального рейтингования: этично ли применять ИИ для создания социального рейтинга?

При выборе наиболее популярных вопросов мы опирались на имеющиеся результаты опросов в России и за рубежом, а также на публичные документы, исследования и научные публикации. В процессе выбора наиболее популярных вопросов участвовала Комиссия по реализации Кодекса этики в сфере ИИ.

Вопросы, затрагиваемые в первой главе, также отмечаются международными организациями, публично-правовыми институтами, авторитетными изданиями при анализе самых популярных этических дилемм развития и использования ИИ, в том числе:

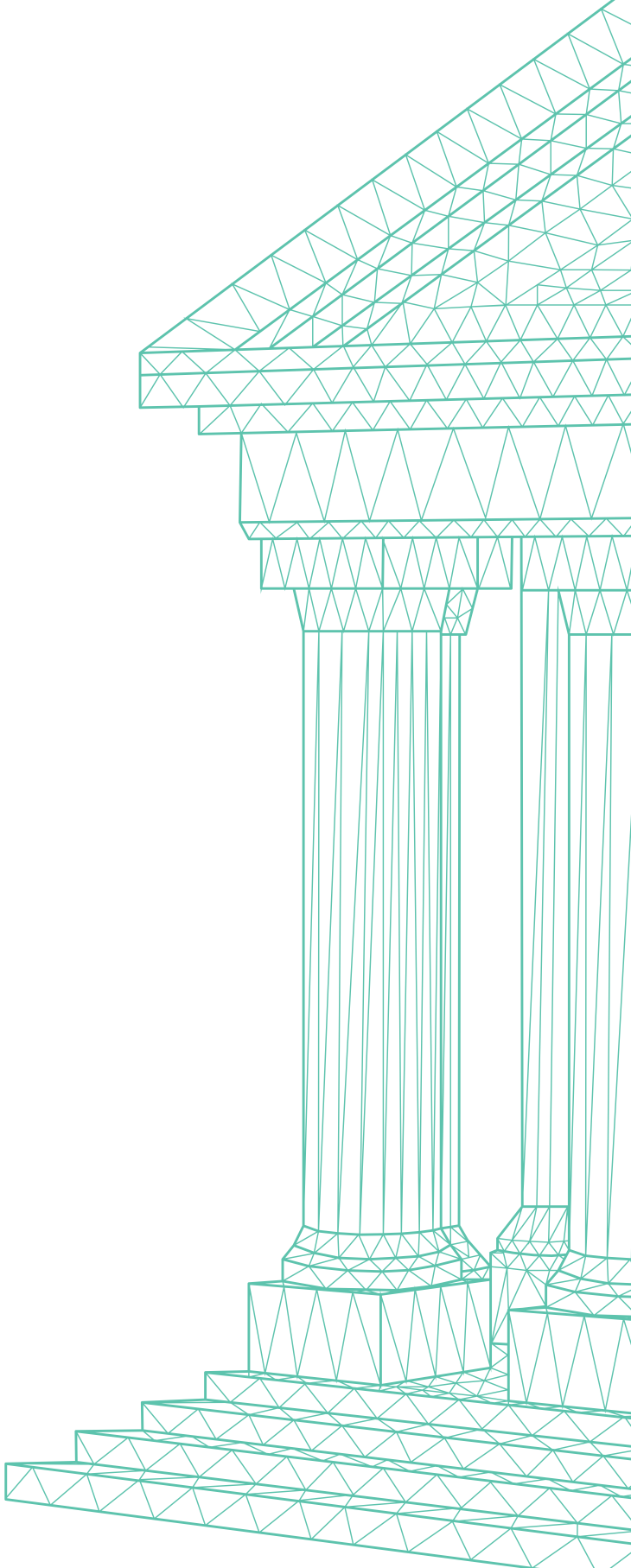
- ЮНЕСКО: дискриминация, человеческий контроль при принятии решений ИИ, транспарентность и объяснимость ИИ систем, влияние на право на труд[10];
- Банк России: непрозрачность алгоритмов, предвзятость, дискриминация[11];
- Группа экспертов высокого уровня по ИИ (ЕС): человеческий контроль при принятии решений ИИ, недискриминация, транспарентность и объяснимость, ответственность[12];
- Международный экономический форум: потеря рабочих мест, дискриминация, безопасность[13];
- Forbes: предвзятость, дискриминация, приватность и конфиденциальность, безопасность, потеря рабочих мест, дипфейки[14];
- Deloitte: предвзятость, потеря рабочих мест, обоснованность и объяснимость принятия решений[15].

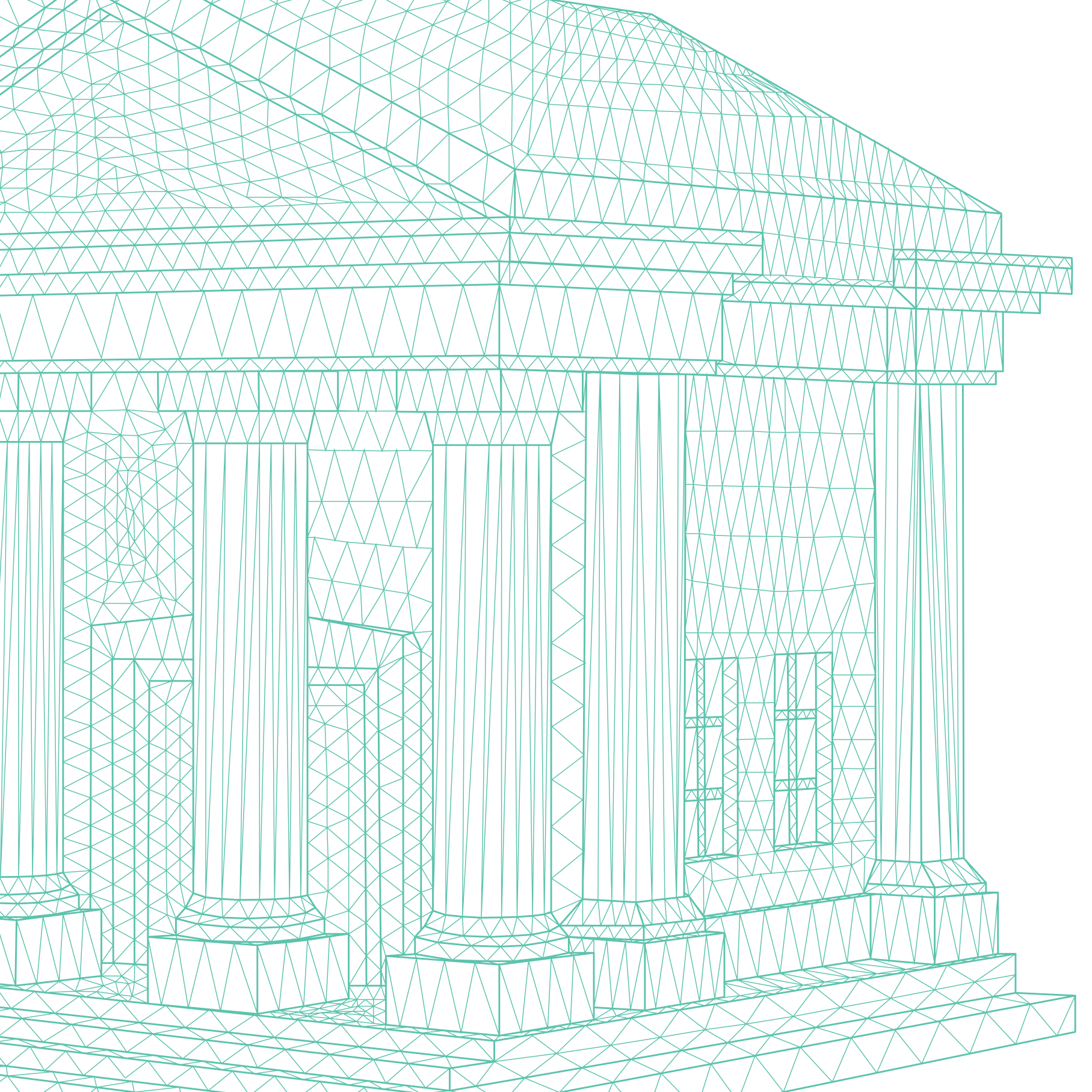
Исследование Сбербанка «Доверие к генеративному искусственному интеллекту»[9]



глава 1

Десять наиболее популярных этических проблем в связи с развитием ИИ





01 «Проблема вагонетки»: какой выбор между жизнями людей в случае неизбежного столкновения должен сделать беспилотный автомобиль?

Ответ:

Все жизни ценны одинаково, поэтому на практике проблемы этического выбора — чья жизнь ценнее — при программировании беспилотного транспорта не существует. Беспилотные транспортные автомобили следует программировать, исходя из необходимости соблюдения правил дорожного движения и принципа причинения наименьшего вреда.

Этические рекомендации разработчикам:

1. Системы искусственного интеллекта (СИИ) в беспилотных автомобилях следует программировать так, чтобы избегать рисков причинения вреда человеку, какие бы иные потери ни последовали.
2. В СИИ не может закладываться право этической оценки рисков причинения вреда и выбора вариантов действий для минимизации последствий.
3. Задача алгоритма — стараться не допустить аварии как таковой в любых условиях (при плохой видимости, дождливой погоде и прочих факторах). Для этого рекомендуется параметризировать граничные условия среды эксплуатации с учетом различных критериев: предельно допустимая скорость, коэффициенты сцепления колеса с дорожным покрытием, допустимые ограничения видимости, ограничения дистанции и прочее.
4. СИИ следует программировать на строгое соблюдение ПДД, включая возможность нарушения ПДД в случаях крайней необходимости для избежания столкновения (снижение/превышение допустимой скорости, нарушение разметки и прочее).

Приведенные этические рекомендации могут применяться к другим видам транспорта с учетом его специфики.

Обоснование:

- Согласно Докладу Всемирной организации здравоохранения о состоянии безопасности дорожного движения за 2023 в результате дорожно- транспортных происшествий ежегодно погибает 1,19 млн человек а примерно 50 млн человек получают несмертельные травмы. Основной причиной таких аварий является человеческий фактор. Переход на автономный транспорт должен существенно сократить смертность на дорогах. Однако его развитие сопровождается рядом этических вопросов.

«Проблема вагонетки» — это известный философский мысленный эксперимент, впервые сформулированный в 1967 году английским философом Филиппой Фут. Традиционный сценарий эксперимента состоит в том, что «сбежавшая» вагонетка движется по пути, на котором находится несколько (как правило, пять) человек. Переключив «стрелку», вагонетку можно направить на другие пути — и тогда погибнет только один человек. Все множество других сценариев сводится к одному вопросу: допустимо ли жертвовать одним человеком ради спасения других жизней? Эта этическая дилемма выявила разницу между двумя моральными концепциями: осознанное (активное) лишение жизни одного человека ради наивысшей общей пользы, а именно — сохранения большего числа жизней, или же пассивное невмешательство, основанное на принципе «никогда не убивать человека».

В исследованиях и публикациях на эту тему предлагаются следующие способы решения:

- Ученые из Стэнфордского университета предложили решение дилеммы с «вагонеткой» для автопилотируемого транспорта. Важно, чтобы программируемое устройство принимало решения только на основании закона. В таком случае возможность ДТП присутствует только при нарушении ПДД иными участниками дорожного движения[16].

- В 2017 году этическая комиссия при Федеральном министре транспорта и цифровой инфраструктуры ФРГ выпустила отчет об автоматизированном вождении[17]. В отчете подчеркивается, что технология должна быть спроектирована прежде всего таким образом, чтобы критических ситуаций не возникало. Не должно возникать ситуаций, в которых автоматизированное транспортное средство должно «решать», какое из двух зол, между которыми нет и не может быть компромисса, ему придется выбрать.

- Французская Национальная пилотная комиссия по цифровой этике в своем отчете об этических аспектах беспилотного транспорта предлагает программировать автономные транспортные средства на случайный выбор действий и внедрить элемент случайности в алгоритмы принятия решений[18]. По их мнению, такой подход позволит разорвать причинно-следственные связи, ведущие к негативным последствиям. Будет сведена к минимуму возможность переложить моральную ответственность с водителя на машину, и, как следствие, сведет к минимуму возможность возложения на машины моральной ответственности.

- Авторы исследования «Принципы вождения беспилотных транспортных средств», которое проведено совместно с экспертами Ford Motor Co[19], приходят к выводу о том, что разработчики автономных транспортных средств должны создавать системы так, чтобы обеспечить последовательное исполнение правил дорожного движения. Авторы предлагают ряд принципов для программирования систем автопилотирования транспортных средств, которые могут снизить риски от использования таких систем и повысить уровень доверия граждан к ним:

- разработчикам не следует пытаться уменьшать вред от ДТП за счет лиц, которые не являются его участниками;
- если нанесение вреда жизни или здоровью неизбежно, то разработчики вправе запрограммировать беспилотное транспортное средство на нарушение правил дорожного движения;
- если какое-либо правило дорожного движения требует рассуждения, беспилотный водитель должен быть запрограммирован таким образом, чтобы в таких случаях уметь осуществлять безопасное маневрирование в таких случаях без риска для окружающих его лиц и объектов.

Этический эксперимент

Исследователи из MIT (Массачусетский технологический институт, Кембридж) опубликовали результаты онлайн-эксперимента, который проводили на сайте The Moral Machine[20]. Участники исследования должны были выбрать, как поступать беспилотному автомобилю в гипотетической ситуации.

Сравнивали девять факторов:

- спасение людей/домашних животных;
- сохранение маршрута/курса;
- сохранение пассажиров/пешеходов;
- сохранение большего/меньшего количества жизней;
- сохранение мужчин/женщин;
- сохранение молодых/пожилых людей;
- спасение пешеходов, которые переходят дорогу по правилам/перебегают в неположенном месте или на красный сигнал светофора;
- спасение стройных/более полных;
- спасение тех, кто имеет более высокий/низкий социальный статус.

У некоторых персонажей были дополнительные атрибуты, которые могли повлиять на решение: беременность, погоны полицейского, халат врача, тюремная роба. Выбирать приходилось между врачом и преступником, беременной и пенсионером, а также другими характеристиками и атрибутами. Результаты исследования сформированы из более 40 млн ответов миллионов пользователей из 233 стран мира.

Во всем мире участники приоритетно спасали человеческие жизни, а не жизни животных, таких как собаки и кошки. Они выбирали сохранить больше жизней, чем меньше, а также хотели сохранить более молодые жизни по сравнению со старыми. Чаще всего щадили младенцев, реже всего спасали кошек. Что касается гендерных различий, то врачи-мужчины и пожилые мужчины были спасены чаще, чем врачи-женщины и пожилые дамы. Спортсменок и более крупных женщин щадили чаще, чем спортсменов-мужчин и более крупных мужчин. Большинство также предпочитали оставлять в живых пешеходов, а не пассажиров, и законопослушных лиц, а не правонарушителей.

Что думают эксперты?

Алексей Лещанкин,
директор по продуктам автономного транспорта Яндекса

— Автономный транспорт принимает решение исходя из ПДД и возможности нанесения наименьшего вреда в случае экстренной ситуации.
Беспилотный автомобиль видит на несколько сот метров вокруг себя, так что аварийная ситуация может случиться с намного меньшей вероятностью, чем с обычным водителем. Перед тем как беспилотные автомобили выезжают на дороги, они проезжают миллионы километров в виртуальной среде: в симуляторе, где можно смоделировать тысячи опасных ситуаций, в том числе те, что которые невозможно протестировать в реальном городе.

Юрий Минкин,
руководитель Департамента разработки беспилотных транспортных средств Cognitive Pilot

— Один из главных ожидаемых результатов внедрения беспилотников состоит в том, чтобы аварий и их жертв стало меньше на порядки. Сейчас на российских дорогах гибнут десятки тысяч человек, а с распространением беспилотного транспорта их число снизится до сотен, а потом и до единиц. В этом смысле самоуправляемый автомобиль априори морален. Он всегда сосредоточен на дороге, у него есть исчерпывающая информация, он может получать данные от других транспортных средств и элементов дорожной инфраструктуры. Создание таких автомашин — перспектива ближайших десятилетий[21].

Стивен Йен,
Baidu

— Наш алгоритм сенсорики не различает людей разного возраста или демографических групп. Он реагирует только на размер, скорость и длину препятствий. Мы также учитываем потенциальное воздействие на препятствие при столкновении с ним. Кажется, что этические соображения пока не являются основным фактором, определяющим поведение автомобиля.

Иван Дейлид,
руководитель отдела разработки программного обеспечения Центра беспилотных технологий Университета «Иннополис»

— Действительно, такая проблема существует. Можно привести достаточно конкретный пример: система на основе ИИ в экстренной ситуации должна принять решение: сбить пешехода, сохраняя выбранную траекторию движения, или из-за резкого маневра подвергнуть опасности пассажира. Но с точки зрения разработчиков, это скорее техническая проблема. Проблема «вагонетки» для беспилотного транспорта не совсем актуальна. Беспилотные системы программируют таким образом, чтобы избежать столкновения с внезапно выбежавшим на дорогу человеком при любых условиях: плохой видимости, дождливой погоде и прочих форс-мажорах. Проблему «вагонетки» можно создать искусственно. Например, если разработчик сузит зону безопасности или увеличит скорость движения беспилотного транспорта.

Практика:

Эксперты Национального управления по безопасности дорожного движения США (National Highway Traffic Safety Administration, NHTSA) выяснили, что в большинстве зафиксированных случаев система автопилота Tesla отключается за несколько секунд до аварии. Это значит, что теперь невозможно привлечь компанию в суде по пункту обвинения в причинении умышленного ущерба из-за работы автопилота.

Претензия NHTSA заключается в том, что при использовании автопилота у водителя есть всего несколько секунд собственных действий, чтобы избежать столкновения[22]. Автопилот сообщает об аварийной помехе на пути и отключается.

В результате анализа исследований на тему беспилотных автомобилей можно выделить следующие наиболее популярные этические принципы:

Непричинение вреда человеку — вот приоритет разработчика.

Этическая «нейтральность». Никакой этический выбор о причинении (или непричинении) вреда не должен закладываться в СИИ.

Неукоснительное соблюдение правил дорожного движения. Все участники дорожного движения, в том числе беспилотные автомобили, должны соблюдать правила дорожного движения (ПДД).

Избежание аварийной ситуации. Задача разработчика — не спасти автомобиль и пассажиров при аварии, а решать аварийную ситуацию. Он должен сделать все, чтобы она не наступила.

02 Проблема цифровых имитаций человека: допустимо ли их создавать?

Ответ:

Создание цифровой имитации человека с этической точки зрения допустимо, но с учетом действующих в конкретной стране законодательных ограничений и при условии следования ряду этических рекомендаций.

Рекомендации:

1. **Избегать дискредитации образа человека.** При создании и использовании цифровых имитаций нужно стремиться не допускать дискредитации человека в результате фальсификации его поведения, искажения позиции или иного использования, которое было бы неприемлемо для оригинальной личности либо его наследников.
2. **Получать согласие человека.** Создание и использование цифровой имитации живущего человека можно признать этичным при наличии явно выраженного согласия. Такое согласие должно обеспечивать понимание того, в каких целях, для какого круга лиц и на каких условиях происходит публикация/трансляция.
3. **Создание и использование цифровых имитаций неживых людей можно считать этичным, если получено согласие родственников.** Создание и использование цифровых имитаций ограниченным кругом лиц (например, родственниками или друзьями) без согласия умершего можно считать этическим выбором этих лиц. Вопрос требует серьезных дополнительных серьезных консультаций с участием психолога.
4. **Маркировать контент, полностью или частично представляющий собой цифровую имитацию человека.** Во время трансляции цифровой имитации следует постоянно и явно информировать о том, что это имитация, искусственно созданная ИИ. Неэтично допускать ситуации, в которых «поведение» цифровой имитации может быть воспринято как поведение реального человека.
5. **Создание цифровых имитаций исторически и культурно значимых личностей можно признать этичным при соблюдении определенных условий.** Так, необходимо избегать оскорбления третьих лиц, их чувств и убеждений, соблюдать нормы права и нормы морали, принятые в обществе, а также учитывать согласие родственников.
6. **Разработчикам и владельцам «ИИ-собеседников» следует информировать пользователя о рисках.** Нелишне, если это позволяет контекст сервиса, предупредить пользователей сервиса о рисках возникновения симпатии, эмоциональной привязанности к цифровой имитации и иных негативных социальных последствиях, которые могут быть разумно спрогнозированы.

Также следует учитывать:

Цифровые имитации уже регулируются законодательством, прямо или косвенно. Во многом на условия создания и использования цифровых имитаций влияет Закон о персональных данных или о защите частной жизни.

Чтобы создать копию с уже существующей цифровой копии, тоже нужно получить согласие. Преобразование цифровой имитации конкретного человека или использование отдельных ее элементов для создания новой цифровой имитации можно считать допустимым, если вновь созданная имитация не может быть отождествлена с оригинальной личностью и не будут нарушены права на образ, голос и другие личностные черты. В ином случае следует обеспечить получение согласия.

Обоснование:

В апреле 2024 года Комиссия по реализации Кодекса этики в сфере ИИ опубликовала **«Этические рекомендации в области создания и использования цифровых имитаций живущих, умерших и несуществующих людей»**[23].

В рамках данных рекомендаций было дано определение цифровой имитации человека.

Цифровая имитация человека — результат цифрового моделирования с применением технологий ИИ на основе цифровых или оцифрованных данных о человеке (синтетических или реальных), направленный на имитацию внешнего вида, голоса и/или других уникальных физиологических, психологических или поведенческих параметров человека, включая стиль общения, принятия решений и пр., выраженный в форме видео, фото, графики, текста и т.д.

В рамках подготовки Рекомендаций членами Комиссии обсуждались риски негативных психологических последствий из-за использования сервисов на основе цифровых имитаций умерших людей. Большинством голосов (43%) принято решение, что психолог/психотерапевт должен сопровождать человека, который задумался об этом вопросе.

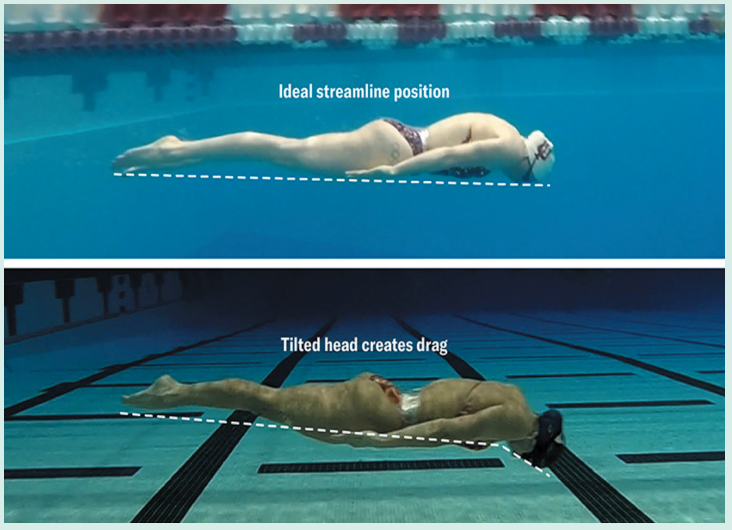
- Отчет ООН о передовых технологиях выделяет **«серый» рынок цифровых копий как представляющий наибольшую опасность**. Создание и использование цифровых имитаций без согласия человека для манипулирования информацией о нем, например в целях изменить исход выборов, не является этичным и создает риски для нормального функционирования демократического общества[24].
- Исследователи из Нью-Йоркского университета Корнелл считают, **что есть риск потенциального искажения убеждений и точек зрения умерших**. Алгоритмы ИИ могут неточно отражать сложности и нюансы мыслей человека. Следовательно, цифровая имитация может непреднамеренно выражать взгляды или совершать действия, которые умерший не одобрил бы при жизни[25].
- Исследователи из Индианского университета Блумингтон выделяют риск, который может возникнуть при использовании цифровых имитаций. **Даже если получено согласие человека, могут возникнуть этические проблемы**. Реальный человек, наблюдая за своей цифровой имитацией, может изменить восприятие себя,

так как цифровые двойники не передают всего спектра человеческих характеристик. Они неизбежно преувеличивают одни черты и умаляют другие[26].

- По мнению исследователей Линденвудского университета, **создание цифровой имитации без согласия лица нарушает его право на неприкосновенность частной жизни**. Процесс воссоздания чьего-либо образа требует доступа к его цифровой информации, которая может быть конфиденциальной. Использование персональных данных без явного согласия умершего или его родственников может также поднять этические вопросы о границах посмертного согласия умершего лица[27].
- Ученые Катарского университета отмечают участившиеся дискуссии о законодательном регулировании данного вопроса. **Создание и использование цифровых имитаций неразрывно связано с решением двух правовых вопросов: защита персональных данных и конфиденциальность частной жизни[28].**
- Этические принципы британской Digital Twin Programme гласят о **важности выбора данных, которые можно использовать для создания цифровой имитации человека**. Ненадежные и неактуальные данные могут вводить в заблуждение общественность, а сенситивные категории данных могут привести к раскрытию конфиденциальной информации[29].

Практика:

В США исследователи из Массачусетского технологического института и Университета Вирджинии создали цифровых двойников олимпийских пловцов, чтобы улучшить их производительность. Пловцов оснастили специальными датчиками, которые фиксируют информацию 512 раз в секунду. Полученные данные исследователи использовали для создания цифрового двойника спортсмена, который фиксирует его движения с точностью до миллисекунды. На данный момент собрана обширная база данных цифровых двойников более чем 100 лучших пловцов США. Такие двойники позволяют анализировать и корректировать техники пловцов, что дает спортсменам возможность улучшить свои результаты[30].



В Европейском союзе цифровые имитации человека активно применяются в области здравоохранения. Европейская инициатива по цифровым близнецам (структура на базе Евросоюза) способствует разработке и внедрению в медицинские учреждения новых решений, основанных на технологии цифровых человеческих близнецов[31].

Цифровой человеческий близнец[32] (Virtual Human Twin) — цифровое представление состояния здоровья человека. Использование цифровой имитации человека помогает спрогнозировать реакцию организма реального человека на применение нового лекарства или на хирургическое вмешательство.

Европейской инициативой по цифровым близнецам также принят манифест по применению и разработке цифровых копий человека для медицинских целей. Акцент на том, что разработка данных технологий и их имплементация в общественное здравоохранение должны соответствовать нормам законодательства, этическим принципам и вопросам безопасности личной информации. Более того, EDITH (European Virtual Human Twin) — платформа, созданная для реализации данной инициативы, — в январе 2024 года выпустила документ, посвященный пробелам в регулировании данной сферы.

Китайская ИТ-компания Silicon Intelligence утверждает[33], что может «вернуть близких к жизни». Точная цифровая копия умершего человека монтируется в одну минуту высококачественного видео: всего за 199 юаней (~2500 рублей).

Чжан Цзэвэй, генеральный директор Super Brain (еще одной компании, «воскрешающей людей»), заявляет, что для того, чтобы генеративный ИИ смог точно передавать образ мышления и поведение умершего близкого, может потребоваться 10 лет сбора всевозможной информации о жизни человека. Он также признает существование этических вопросов: правильно ли пытаться обмануть смерть? Способствует ли цифровая копия проживанию горя или, наоборот, препятствует этому? Тем не менее Чжан Цзэвэй надеется, что технологии ИИ все же приносят некоторое облегчение в процессе скорби.

С точки зрения закона таким компаниям необходимо получать либо прижизненное согласие самого человека, либо согласие его родственников. Так, Гражданский кодекс КНР предусматривает, что никто не может нарушать чужие права на образ, используя информационные технологии для фальсификации его изображения[34].

Что думают эксперты?

Олеся Васильева,
практикующий психолог и преподаватель
Московского института психоанализа

— Наследие, которое остается от человека — это, с одной стороны, возможность прикоснуться к тому, что нам дорого... Но что касается создания цифровой копии или общения с человеком, которого не стало, с помощью чат-бота, то тут не все так однозначно. Это иллюзия поддержания того факта, что человек жив. И тогда мы не можем запустить процессы горевания, которые так важны для психики.

Марина Романовская,
клинический психолог

— Когда человек сталкивается с утратой, то проходит несколько стадий принятия неизбежного. Если человек находится в стадии острого переживания, то «разговор» с умершим близким может углубить его травму. Этот опыт будет скорее более травматичен, нежели психотерапевтичен. Однако в психотерапии при проработке травмы используется «техника пустого стула», когда мы представляем себе умершего родственника и можем ему сказать все, что не смогли или не успели сказать при жизни. Если человек пришел на психотерапию и доверился специалисту, такой способ взаимодействия будет очень полезен.

Владимир Табак,
генеральный директор
АНО «Диалог Регионы»

— Разработка «цифровых двойников» человека — сложная задача, которая включает в себя множество этических, правовых и социальных измерений. При их создании важно не только соблюдать законы и этические принципы, но и учитывать мнение общественности и специалистов в области этики ИИ. Как защитить личные данные от неправомерного использования? Какие необходимы контролирующие правила? Применение ИИ не должно ущемлять личную свободу и ограничивать людей в самостоятельном принятии решений. Поэтому один из базовых принципов звучит так: цифровые имитации должны быть достоверными, а не искажать образ человека. Применение копий для обмана и создания фейков недопустимо.

Андрей Ильин,
руководитель направления синтеза
визуального контента,
Центр ИИ Т-Банка

— Использование цифровых двойников открывает новые возможности коммуникации, что может быть выгодно как пользователям, так и бизнесу. Но с другой стороны, возникает множество вопросов, связанных с контролем и этичностью их применения. Отрасль новая, но методы развиваются стремительно. Разработчикам нужно быть осторожными при принятии решений и деликатными при внедрении. Важно изучить необходимость осторожного подхода разработчиков к внедрению таких решений и изучению реальных последствий использования цифровых копий. Например, в Т-Банке мы активно развиваем технологии создания реалистичных аватаров для внешней и внутренней коммуникации, а также разрабатываем способы обнаружения DeepFake-атак для защиты наших клиентов от злоумышленников.

03 Проблема «черного ящика»: как понять принципы работы систем ИИ и объяснить их пользователю?

Ответ:

Проблемой «черного ящика» часто называют ситуацию, когда невозможно понять, почему СИИ выдает тот или иной результат в каждом конкретном случае. Чем лучше и точнее работает алгоритм, тем сложнее объяснить его решение — это связано с тем, что такое решение является следствием взаимовлияния миллионов неочевидных факторов. Понять и объяснить можно только примитивные, а значит, плохо работающие СИИ.

Рекомендации для разработчиков:

В зависимости от контекста и назначения СИИ, рекомендуется раскрывать пользователям, например, следующие факты:

1. цели обучения: какие цели были поставлены перед алгоритмом при его обучении;
2. метрики оценки эффективности: какая функция от каких параметров оптимизировалась при проведении машинного обучения;
3. использованные алгоритмы машинного обучения;
4. рекомендации по сфере применения

И другие[35].

Рекомендации для пользователей:

1. **Изучайте основные принципы работы алгоритмов**. Это создаст общее понимание процессов принятия решений системами ИИ.
2. **Изучайте пользовательские и лицензионные соглашения компании-разработчика**. Также учитывайте любую иную релевантную информацию на сайте компании-разработчика.
3. **Запрашивайте у разработчиков дополнительную информацию, которая вас интересует**. Например о том, какие данные используются для обучения системы, какие факторы учитываются и как они влияют на результаты.
4. **Обращайтесь к специалистам и экспертам в этой области**. Они помогут объяснить особенности проблемы «черного ящика».

5. **Читайте научные статьи, книги и другие материалы, посвященные проблемам прозрачности ИИ**. Это позволит получить более глубокие знания в области рассматриваемого вопроса.
6. **Активно участвуйте в образовательных проектах**. Например, посещайте курсы, которые расширяют ваши цифровые компетенции и способствуют общему пониманию работы алгоритмов ИИ.

Обоснование:

- Согласно исследованию Немецкого центра науки о данных, ИИ и больших данных использует термин **«черный ящик»** — это термин, используемый для описания ситуации, когда невозможно или очень сложно объяснить, как именно модель искусственного интеллекта пришла к определенному решению[36].
- Это происходит оттого, что ИИ — сложная система с множеством параметров и взаимосвязей между ними. Даже разработчики модели могут не понимать всех тонкостей ее работы.
- Проблема «черного ящика» также поднимает ряд этических вопросов, связанных с прозрачностью. Если мы не можем понять, как алгоритм ИИ принимает решения, то как мы можем гарантировать правильность и справедливость таких решений?
- Исследователи предлагают разные способы повышения прозрачности и интерпретируемости алгоритмов искусственного интеллекта. Один из подходов заключается в разработке «объяснимого ИИ» или XAI. Например, система ИИ, которая рекомендует план лечения для пациента, может предоставить список факторов, которые повлияли на принятие решения: история болезни пациента, результаты тестов и текущие симптомы.
- Другой подход заключается в использовании методов машинного обучения, которые позволяют людям понять, как алгоритм ИИ принимает решения. Например, какие учитываются характеристики или другие исходные данные, на которые опирается алгоритм ИИ при принятии решения.

Подходы международных организаций:

1. **Рекомендации ЮНЕСКО об этических аспектах ИИ**[37] особенно выделяют прозрачность и объяснимость ИИ-систем. Отмечается, что соблюдение указанных принципов гарантирует защиту и охрану прав человека и его свобод. Согласно Рекомендациям, пользователи должны располагать информацией о том, что данные предоставляются на основе ИИ-алгоритмов, особенно когда такие данные могут затрагивать основные права человека. В таком случае у пользователя должна быть возможность обратиться к разработчикам ИИ за разъяснениями работы алгоритмов системы.
2. **В Резолюции ООН «Использование возможностей безопасных, защищенных и надежных систем искусственного интеллекта для устойчивого развития»**[38] **также подчеркивается ценность прозрачности для ИИ-систем.** Принцип прозрачности включает в себя разъяснение работы алгоритмов, надзор человека за системой и обеспечение проверки автоматизированных решений. Прозрачные и объяснимые системы ИИ повышают надежность, позволяют конечным пользователям лучше понимать, принимать и доверять результатам и решениям ИИ.

Практика:

1. Раскрытие разработчиками информации об алгоритмах работы ИИ может быть необходимо, поскольку в определенных ситуациях замалчивание информации о том, как работает система, может привести к необходимости пересмотра результатов.

   Например, как это было в случае с ИИ-системой, обученной анализировать рентгеновские снимки на предмет наличия раковых опухолей. Предполагалось, что она упростит и ускорит работу врачей в части количества просматриваемых снимков. Разработчики сделали систему очень чувствительной, чтобы она не пропускала возможные случаи рака, но из-за этого часто появлялись ложные срабатывания. Указанный алгоритм не был разъяснен рентгенологам, которые использовали ИИ. Врачи затратили больше времени на перепроверку результатов, маркированных ИИ, поскольку они не знали, что система слишком чувствительна, и продолжали искать то, чего, как им казалось, они не смогли увидеть изначально[39].

2. Крупные компании, разрабатывающие технологии ИИ, придерживаются принципа прозрачности, в том числе через раскрытие дополнительной информации в случаях технических сбоев и иных ошибках алгоритмов.

   Так, 20 марта 2023 года произошел сбой в работе ChatGPT (нейросеть от компании OpenAI). Представители компании опубликовали пресс-релиз[40] на своем сайте с извинениями и объяснениями:

   «Ранее на этой неделе мы отключили ChatGPT из-за ошибки в библиотеке с открытым исходным кодом, которая позволяла некоторым пользователям просматривать заголовки из истории чата другого активного пользователя. После более тщательного расследования мы также обнаружили, что та же ошибка могла привести к непреднамеренному отображению платежной информации у 1,2% подписчиков ChatGPT Plus.

   Мы обратились к пострадавшим пользователям, чтобы уведомить о том, что их платежная информация, возможно, была раскрыта. Мы уверены, что в настоящее время нет угрозы для данных пользователей.

   Все сотрудники OpenAI стремятся защищать конфиденциальность наших пользователей. К сожалению, на этой неделе мы не смогли оправдать ваши ожидания. Мы еще раз приносим извинения нашим пользователям и всему сообществу ChatGPT и будем усердно работать над восстановлением доверия».

Что думают эксперты?

Сэм Альтман,
главный исполнительный директор Open AI на Всемирном экономическом форуме — 2024

— Я не обладаю способностью читать человеческие мысли. Однако могу попросить человека разъяснить свою точку зрения, чтобы самостоятельно оценить логичность его рассуждений. Полагаю, что системы искусственного интеллекта в скором времени смогут выполнять аналогичную функцию. Они будут способны детально объяснять нам ход своих рассуждений на естественном языке, позволяя нам самостоятельно оценивать правильность принятых ими решений, даже не вдаваясь в технические подробности[41].

Семен Буденый,
управляющий директор-начальник управления развития перспективных технологий AI, Сбербанк

— Проблема «черного ящика» — это отсутствие понимания, что происходит внутри нейросети. Ее «решения» — лишь результат множества математических операций, а не осмысленного рассуждения. Как и наш мозг, устройство которого мы не до конца понимаем, нейронную сеть и ее особенности можно изучать, но любое объяснение от нее (например, как языковой модели) — лишь удачная имитация рассуждения, подкрепленная большой «насмотренностью».

Евгений Павловский,
заведующий лабораторией аналитики потоковых данных и машинного обучения, НГУ

— Для моделей, безусловно, нужно показывать, на каких данных они были обучены. Это позволит реализовать принцип отслеживаемости, чтобы впоследствии при исправлении ошибок обучающих данных знать, как они повлияли на качество модели. Прозрачность создания моделей на каждом этапе позволяет контролировать их качество и лучше понимать условия применения.

Сергей Израйлит,
вице-президент по развитию и планированию Фонда «Сколково»

— Прозрачность алгоритмов искусственного интеллекта — это наша возможность для создания взаимного доверия между заказчиками и разработчиками, которое в долгосрочной перспективе определяет скорость внедрения любых технологий в не меньшей степени, чем способность создавать соответствующие решения. Соблазн скрывать значимые факты, особенно негативно влияющие на текущие продажи, присутствует всегда, однако в современном открытом мире поддаваться такому соблазну — значит терять доверие клиентов и создавать репутационные риски для акционеров и инвесторов.

Олег Кипкаев,
начальник отдела по надзору за исполнением законов в сфере информационных технологий и защиты информации Главного управления по надзору за исполнением федерального законодательства, Генеральная Прокуратура РФ

— Когда мы решим проблему «черного ящика» в искусственном интеллекте, мы сможем не только сделать его работу более прозрачной и понятной, но и откроем новую эру взаимного обучения между человеком и машиной. Расшифровка внутренних процессов ИИ позволит нам перенять у него нетривиальные способы решения задач, а ИИ, в свою очередь, сможет адаптироваться к человеческой логике и этике. Это может привести к созданию гибридных систем, где границы между человеческим и машинным интеллектом станут более размытыми, открывая путь к инновациям, которые сегодня кажутся недостижимыми. Таким образом, решение проблемы «черного ящика» станет катализатором для качественно нового уровня развития технологий и общества.

04 Проблема информирования: всегда ли люди должны знать, что взаимодействуют с ИИ?

Ответ:

Раскрытие пользователю информации о взаимодействии с ИИ является желательным и стимулирует доверие человека к работе системы, но такое требование не должно применяться универсально: существует немало ситуаций, где это может быть неоправданным или уже явно является очевидным.

Рекомендации для разработчиков:

1. **Учитывайте сферу применения ИИ**. Рекомендуется осуществлять добросовестное информирование пользователей об их взаимодействии с СИИ, когда это затрагивает вопросы прав человека и критических сфер его жизни и обеспечивать возможность прекратить такое взаимодействие по желанию пользователя.

2. **Не рекомендуется допускать введения пользователя в явное заблуждение**. Не следует сообщать пользователю о том, что он взаимодействует с реальным человеком, если это утверждение ложно.

3. **Раскрытие информации должно быть явным, ясным и очевидным для пользователя**. Эксперты рекомендуют раскрывать информацию для пользователей, например в пользовательском соглашении, в политике конфиденциальности, на странице часто задаваемых вопросов (FAQ), в справочных материалах и в уведомлениях при установке или в уведомлениях при первом запуске продукта.

4. **Иногда информирование пользователей о факте взаимодействия с ИИ не нужно в силу обстоятельств использования или уже является очевидным**. Например, ИИ используются в онлайн-картах и навигаторах. В подобных ситуациях пользователю неважно с кем именно он взаимодействует, если задача качественно выполняется. В других случаях факт взаимодействия с ИИ может быть очевидным — например, при взаимодействии с голосовым помощником из умной колонки.

5. **В некоторых случаях раскрытие информации о том, что человек взаимодействует с ИИ, может быть, наоборот, даже неоправданным или даже нежелательным**. Например, акцент на применении ИИ во внутренних, производственных процессах компаний никак не повлияет на клиентов компании, однако может стать предметом внимания злоумышленников. В других случаях системы ИИ могут применяться для оказания срочной услуги (например, записи в медицинское учреждение). В этих случаях целенаправленный акцент на том, что с пользователем взаимодействует ИИ, может привести к недоверию со стороны пользователей и упущению потенциальной пользы от услуги.

6. **Важно, чтобы у пользователя была техническая возможность оставить запрос на получение информации о взаимодействии с ИИ**. Это может быть реализовано через специальный сервис либо в обращении, направленном по официальным каналам связи.

7. **Если пользователь уточняет, взаимодействует ли он с ИИ, следует давать честный ответ**. Программирование СИИ на ложный ответ можно считать неэтичным.

Обоснование:

- Исследователи из университета Мишкольца в Венгрии считают[42], что системы ИИ пока могут не знать некоторых принципов морали и нравственности. Современные системы ИИ уже способны взаимодействовать с человеком так, что по большей части их участие будет неотличимо от реального человека. Если человек не знает, что его «собеседник» не реальный человек, а система ИИ, он может стать жертвой свойств ИИ, которые для него были бы неприемлемы по морально-этическим соображениям.
- Человеку свойственно ожидать, что специалист, оказывающий ему услуги, несет ответственность за свои рекомендации и решения. Система ИИ же такой ответственности не несет, поскольку не является субъектом правоотношений.
- В отчете Американской консалтинговой компании Weil, Gotshal & Manges LLP[43] говорится, что применение систем ИИ без раскрытия информации об их использовании может привести к снижению доверия общества к этим технологиям. Если люди начнут сомневаться в эффективности и безопасности технологий, это может замедлить их внедрение и развитие.
- По мнению Роберта Бейтмана[44], сертифицированного международного специалиста по конфиденциальности (CIPP/E), боты становятся все более популярными и совершенными, что может ввести в заблуждение пользователей, рассчитывающих на общение с живым человеком. Еще недавно было легко понять, что вы общаетесь с ботом: ответы были стремительными, а их суть сводилась к фразе «К сожалению, я не могу вам помочь». Общение заканчивалось через 3–4 минуты. Однако сегодня технологические компании достигли значительных результатов в развитии искусственного интеллекта, обработки естественного языка и машинного обучения.
- Некоторые пользователи могут не желать, чтобы их запросы выполнялись с использованием ИИ[45]. Например, пользователи в сфере здравоохранения могут переживать за свою конфиденциальную информацию и не желать предоставлять ее кому-либо, кроме конкретного специалиста. Такая ситуация особенно характерна для сферы психологической помощи, где клиенту, как правило, важен личный контакт именно с человеком.

Практика:

Преподаватель Джилл Уотсон около пяти месяцев помогала студентам Технологического института Джорджии в работе над проектами по дизайну программ. Нюанс в том, что Джилл — это робот, система ИИ, работающая на базе IBM Watson, но никто из студентов, обсуждая работы с преподавателем, за все это время ничего не заподозрил. А кто-то из студентов даже собирался назвать ее «выдающимся педагогом».

«Она должна была напоминать нам о датах дедлайна и с помощью вопросов подогревать обсуждения работ. Это было как обычный разговор с обычным человеком», — рассказала изданию студентка вуза Дженнифер Гевин[46].

Подходы регуляторов[47]:

В мировой практике можно встретить примеры нормативного регулирования данного вопроса.

1. Например, Закон штата Калифорния о раскрытии информации о ботах[48] (Калифорнийский кодекс бизнеса и профессий, § 17940) гласит, что на территории штата взаимодействие в интернете с лицами, которые хотят продать товары, услуги или повлиять на исход выборов посредством ботов без явного раскрытия информации о том, что пользователь общается с ИИ, является незаконным.

2. Вступивший в силу 1 августа 2024 года европейский AI Act[49] относит чат-ботов на базе ИИ к системам низкого риска. Функционирование таких систем обязательно должно сопровождаться уведомлением ее пользователей о том, что они взаимодействуют с ИИ. Более того, закон требует маркировать контент, созданный генеративными чат-ботами.

3. В России статья 10.2.2 Федерального закона № 149 «Об информации, информационных технологиях и о защите информации»[50] устанавливает особенности предоставления информации с применением рекомендательных технологий. Так, порядок применения рекомендательных сервисов включает информирование пользователей и публикацию на информационном ресурсе правил применения рекомендательных технологий.

4. В российском Кодексе этики в сфере ИИ, как в инструменте саморегулирования, одним из основополагающих принципов является идентификация ИИ в общении с человеком, согласно которому рекомендуется осуществлять добросовестное информирование пользователей об их взаимодействии с СИИ.

Согласно исследованию Сбербанка «Доверие к генеративному искусственному интеллекту», проведенному в 2024 году, 62% респондентов доверяют ИИ чат-ботам.

Степень необходимости раскрытия информации для обеспечения прозрачности варьируется в зависимости от сферы применения ИИ.

Так, пользователи меньше всего доверяют технологиям искусственного интеллекта следующие задачи:

- только 35% опрошенных доверили бы улучшение психического и физического здоровья ИИ;
- 39% — патриотическое воспитание молодежи;
- 44% — освещение событий в СМИ.

Именно в тех сферах, в которых пользователи менее всего проявляют доверие к технологии, информированность пользователя должна быть неотъемлемой частью этичного использования ИИ.

В других же сферах, например в сфере обслуживания клиентов, статистика показывает высокий уровень доверия граждан к применяемой технологии:

- 62% респондентов доверяют работе генеративного ИИ в чат-ботах.

В таких случаях, как правило, решение о раскрытии информации об использовании ИИ зависит от конкретных целей и контекста.

62%

респондентов доверяют работе генеративного ИИ в чат-ботах.

Что думают эксперты?

Константин Воронцов,
профессор кафедры интеллектуальных систем МФТИ, профессор РАН

— Чат-бот обязан предупреждать в начале разговора не только о том, что он машина, но также уведомлять, что у него нет эмоций, желаний, намерений, и его единственная функция — оказать информационную услугу в рамках, оговоренных законом и правилами сервиса.

Валентин Макаров,
президент Ассоциации «РУССОФТ»

— Да, люди должны знать, что общаются с ИИ. Процесс построения логики ИИ отличается от того, как мыслит человек, поэтому человек должен знать, с кем имеет дело. В противном случае ожидания человека от общения с собеседником могут быть ложными и привести к неадекватным решениям и поступкам.

Владислав Архипов,
профессор, руководитель юридической группы Центра ИИ и науки о данных СПБГУ

— На мой взгляд, люди должны знать, что взаимодействуют с ИИ в тех случаях, когда такое взаимодействие затрагивает их права и законные интересы. Таких случаев больше, чем может показаться: это могут быть и такие виды взаимодействия, в которых решается серьезный юридически значимый вопрос (например, прием на работу), и гораздо менее значимые (например, взаимодействие с «роботом» в рекламной акции), однако и в последнем случае затрагивается как минимум право на человеческое достоинство, переосмысленное в условиях цифровой эпохи.

Денис Озорнин,
директор по продукту «Алиса»

— ИИ-технологии постоянно совершенствуются, но все еще могут допускать ошибки. В то же время ответы, созданные ИИ, отличить от ответов реального человека становится все сложнее. Информирование позволяет избежать введения пользователей в заблуждение, когда они сомневаются, с кем взаимодействуют — с человеком или с ИИ, — а также помогает более критично оценивать контент, сгенерированный технологией. Мы, со своей стороны, информируем пользователей о взаимодействии с ИИ различными способами. При общении с Алисой в чате ассистент предупреждает в нижней части интерфейса, что может допустить ошибку. Если пользователь интересуется, как она генерирует ответы, Алиса сознается, что делает это с помощью нейросети.

05 Проблема сокращения рабочих мест: приведет ли массовое внедрение ИИ к тому, что люди останутся без работы?

Ответ:

Нет, внедрение ИИ не приведет к массовой безработице. Корректнее говорить не столько о потере работы, сколько об изменении структуры рынка труда. В долгосрочной перспективе адаптация и переобучение помогут улучшить условия труда. Рынок труда станет более гибким и устойчивым к кризисам.

Рекомендации

Для пользователей:

1. **Осваивайте цифровые навыки.** Важно быть готовым к тому, что некоторые профессии могут исчезнуть или измениться под влиянием технологий ИИ. Задумайтесь о том, какие навыки будут востребованы в будущем, и начните их осваивать. Это могут быть языки программирования, основы аналитики данных, 3D-моделирование и другое.

2. **Будьте гибки, используйте новые возможности.** Вместо того, чтобы бояться ИИ, используйте его как инструмент для повышения производительности и эффективности. Массовое внедрение ИИ приведет к созданию новых профессий и, соответственно, рабочих мест. Следите за тенденциями в данной области ИИ и осознавайте новые возможности для развития своей карьеры.

Для разработчиков:

1. **Оценивайте возможные риски и разрабатывайте меры по их минимизации.** Прежде чем выпускать систему ИИ в публичный доступ, рекомендуется провести анализ рисков для рынка труда и разработать стратегии их минимизации: (например, меры по адаптации и переподготовке работников).

2. **Способствуйте развитию актуальных навыков.** Организовывайте образовательные программы, сотрудничайте с организациями, занимающимися профессиональной переподготовкой, а также публикуйте обучающие статьи и иные материалы.

3. **Создавайте новые рабочие места.** Массовое внедрение технологий ИИ неизбежно приведет к появлению новых продуктов или услуг, требующих квалифицированных кадров. Им потребуется уверенное владение навыками с навыками в области машинного обучения и анализа данных.

Для регуляторов:

1. **Запускайте программы профессиональной переподготовки и комплексные системы социальной защиты.** Это поможет преодолеть краткосрочные негативные эффекты ИИ в сфере занятости наиболее уязвимых категорий работников, а также сделать переход к ИИ более инклюзивным и ограничить социальное неравенство.

2. **Приоритизируйте развитие цифровых компетенций в определенных сферах.** Например, в таких отраслях, как здравоохранение, финансы и образование, можно извлечь выгоду непосредственно из внедрения технологий ИИ — за счет улучшения процесса принятия решений и создания новых возможностей.

3. **Инвестируйте в развитие соответствующих отраслей.** На уровне страны для подготовки к интеграции ИИ в жизнь общества и предприятий можно инвестировать в цифровую инфраструктуру и подготовку квалифицированной рабочей силы, владеющей цифровыми технологиями.

Обоснование:

- Международная организация труда (МОТ)[51] выделяет два типа применения ИИ на рабочем месте: автоматизация рутинных задач сотрудников и автоматизация управленческих функций работодателя (например, при найме сотрудников или при их обучении). Приведет ли такое внедрение ИИ к потере рабочих мест или, наоборот, к увеличению их количества — зависит от того, как технология интегрируется в рабочие процессы, и от желания руководства сохранить людей для контроля над автоматизированным выполнением этих задач.
- ИИ создаст новые рабочие места. Например, уже сейчас появляются такие профессии, как инженеры машинного обучения (ML-инженеры), аналитики данных (Data Scientists), инженеры по обработке естественного языка (NLP-инженеры), AI-тренеры, специалисты по этике ИИ.
- Тем не менее какая-то часть рабочих мест действительно будет вытеснена ИИ. Согласно исследованию, проведенному McKinsey[52], доля профессий, характеризующихся выполнением рутинной работы (например, упаковка продуктов, управление транспортным средством) или требующих низкого уровня цифровых навыков, к 2030 году может снизиться на 40% по отношению к показателю общей занятости.
- Ученые из Стэнфорда напоминают, что развитие технологий всегда приводило к изменению структуры экономики, которая в свою очередь влияла на рынок труда, — это историческая закономерность[53].
- Согласно исследованию, проведенному Международным валютным фондом, ИИ может помочь малоопытным работникам быстрее продвигаться по карьерной лестнице. Работники, которые смогут эффективно использовать технологии ИИ, обретут рост не только своей производительности и компетенций, но и заработной платы.

Практика:

1. Согласно результатам опроса, проведенного по заказу ВТБ весной 2024 года, были названы профессии, в которых россияне больше всего обеспокоены заменой человека ИИ. Почти 40% россиян опасаются, что на работе в банках и других финансовых учреждениях их заменит искусственный интеллект. Опасения выражают также специалисты IT (45%), торговли и общепита (44%), сотрудники сферы транспорта (39%), здравоохранения (38%), промышленности (37%), образования (34%) и строительства (31%)[54].

2. При Международной организации труда (МОТ) создана самостоятельная Инициатива по раскрытию информации о труде в области искусственного интеллекта (AILDI)[55]. Эта структура выступает за раскрытие информации об использовании ИИ на рабочих местах для соблюдения принципа прозрачности и других этических принципов использования ИИ. ALIDII также занимается изучением того, как интеграция практик машинного обучения может способствовать улучшению положения сотрудников.

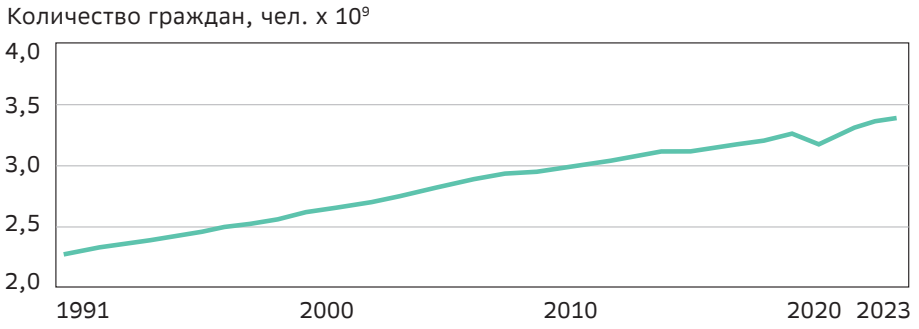
Исследования по вопросу:

1. Согласно исследованию Международного валютного фонда, в странах с развитой экономикой около 60% рабочих мест могут быть затронуты ИИ. При этом половина выиграет от интеграции ИИ, а другая половина, возможно, увидит снижение спроса[56].
   В развивающихся и наименее развитых странах влияние ИИ может затронуть 40% и 26% рабочих мест. Отсутствие инфраструктуры усугубит неравенство между странами.
   Области занятости, наиболее подверженные ИИ, включают в себя управленческий персонал, офисных сотрудников, технических работников и некоторые профессиональные категории (например, иллюстраторы и копирайтеры). Наименее подвержены ИИ области занятости, связанные с физическим трудом, ремеслами, сельским хозяйством. При этом навыки использования ИИ наиболее комплементарны для конкурентоспособности офисных работников и сотрудников сферы обслуживания.
   Инструменты ИИ могут высвободить время и ресурсы для таких секторов, как сельское хозяйство, здравоохранение и образование. Это может переориентировать рынок труда в пользу социально и экономически уязвимых слоев населения, нивелируя проблемы, связанные с временной потерей рабочих мест.

2. Исследование ЮНЕСКО показало, что ИИ может оказать влияние на 80% рабочей силы США, затрагивая примерно 10% их рабочих задач. Эти инструменты могут быть использованы для автоматизации задач, традиционно связанных с человеческими функциями, включая рассуждения, написание текстов, создание графиков и анализ данных[57].

3. Китайские исследователи пришли к выводу, что последствия роботизации и автоматизации производства, а также внедрения ИИ в производство в Китае, принесли больше пользы, чем вреда рынку труда в стране. Конкурентоспособность работников увеличилась, и у кандидатов появилась возможность выбирать из большего количества видов работ[58].

4. ВМФ отмечает, что за последние 200 лет большинство прогнозов о сокращении количества рабочих мест оказывались ложными, так как одновременно с упразднением рабочих мест появлялись новые профессии и специальности. Вспомним, как автоматизация сельского хозяйства заменила миллионы рабочих в данной сфере, но промышленная революция создала рабочие места на фабриках. Позже автоматизация фабрик вытеснила работников с заводов, но в то же время дала толчок развитию рынка труда в сфере услуг.



Количество трудоустроенных граждан по всему миру за период с 1991 по 2023 год (x 10^9)
Источник МВФ[59]

На протяжении разных революций и реформаций количество созданных рабочих мест оказывалось больше, чем упраздненных. Сегодня во всем мире, почти в каждой стране, зафиксировано рекордное количество трудоустроенных[59].

Что думают эксперты?

Олег Буклемишев,
директор Центра исследования экономической политики экономического факультета МГУ

— Профессии постоянно исчезают из-за автоматизации и цифровизации, причем вполне возможно, что на данной стадии под максимальной угрозой уже услуги, а не промышленность, как ранее. Мы уже видим вытеснение операторов колл-центров, разного рода консультантов, контролеров, есть и другие профессии, которым явно угрожает искусственный интеллект.

Артем Бондарь,
руководитель направления обработки естественного языка, Центр ИИ Т-Банка

— На мой взгляд, массовое внедрение ИИ не только не способствует потере рабочих мест, но и, напротив, создает новые возможности для специалистов. Показательным примером служит ситуация в области копирайтинга. На первых этапах внедрения генеративных технологий казалось, что они представляют реальную угрозу для специалистов, чьи профессии связаны с созданием контента. Однако со временем мы увидели, что ИИ стал для них копилотом: творческие задачи остаются прерогативой компетентных сотрудников, а всю рутинную работу можно делегировать технологиям. Более того, искусственный интеллект требует постоянного обучения на больших объемах данных. Решение этой задачи кроется в создании новой профессии: ИИ-тренера. Крупнейшие компании, включая Т-Банк, активно нанимают таких специалистов, что подтверждает рост спроса на профессионалов в области создания и обработки контента в связи с массовизацией ИИ.

Андрей Белевцев,
старший вице-президент, руководитель блока «Технологическое развитие» Сбербанка

— Внедрение каждой новой прорывной технологии с большими перспективами применения в различных областях сопровождается подобным беспокойством. Но здесь нужно осознавать, что людей пугает неизвестное. Чтобы избежать такого эффекта, нужно усиливать информирование людей о том, как устроена технология и какую практическую ценность она может для них нести. Что же касается потенциальной потери рабочих мест — я думаю, что в перспективе под ударом могут оказаться сферы деятельности, которые не требуют от сотрудников глубоких компетенций. Но на эту ситуацию можно и нужно смотреть с другой стороны — со стороны того, как генеративный AI может сделать работу человека эффективнее, сократить объем рутинных задач[60].

Яков Сергиенко,
партнер, руководитель «Яков и партнеры»

— ИИ открывает прекрасные возможности для трансформации рынка труда. С одной стороны, благодаря повышению производительности он внесет вклад в борьбу с нехваткой сотрудников в целом ряде индустрий, с другой — он уже сейчас создает новые высокооплачиваемые профессии, связанные с разработкой и внедрением технологий. Обдуманные программы переподготовки станут ключом к реализации этих возможностей, а компании, которые инвестируют в ИИ и обучение персонала работе с ним, смогут быстрее выйти на новый уровень.

06 Проблема оспаривания: всегда ли человек должен иметь возможность оспорить решение, принятое с использованием ИИ?

Ответ:

Право оспорить решение, принятое с использованием ИИ, считается одной из фундаментальных концепций в вопросах этичного использования ИИ, но оно не является универсальным для всех случаев применения ИИ.

Рекомендации

Для разработчиков:

1. **Учитывайте область внедрения технологий ИИ**. Механизм оспаривания может быть излишним в ситуациях, когда решение, принятое ИИ, не имеет серьезных последствий. Например, подбор ИИ рекомендуемого контента или составление маршрута навигатором с ИИ могут быть не всегда точными, но такие решения не оказывают существенного влияния на жизнь пользователя.
2. **Создавайте инструменты оспаривания решений, принятых ИИ, в тех сферах, где это необходимо**. Разрабатывайте механизмы, которые позволят пользователям оспаривать решения, принятые ИИ. Например, это может быть возможность обратиться к живому специалисту или предоставить дополнительную информацию, которая повлияет на решение системы.
3. **Следует стремиться к тому, чтобы любые решения, принимаемые ИИ, были прозрачными и понятными для человека**. Это позволит пользователям лучше понимать, почему было принято то или иное решение, и предпринимать действия при необходимости.

Для пользователей:

1. **Учитывайте положения законодательства.** В любой сфере, где система ИИ совершает юридически значимые действия или принимает решения, непосредственно влияющие на качество и условия жизни человека, следует ориентироваться на положения законодательства. Как правило, в них прописана процедура оспаривания таких решений.

2. **ИИ-решения наравне с решениями, принятыми человеком, являются предметом внутреннего регулирования в компании.** Принимайте во внимание, что использование ИИ в большинстве случаев регламентируется не только законодательно, но и локальными нормативными актами.

3. **Обратитесь в поддержку сервиса.** Специалисты пользовательской поддержки помогут разобраться с деталями работы алгоритма и причинами принятого решения, а также объяснят процедуру оспаривания.

4. **Если служба поддержки не смогла предоставить удовлетворяющее объяснение или решение, обратитесь в вышестоящие инстанции , органы судебной власти или иные государственные органы.** Например, в специальные комиссии по этике в компаниях.

Обоснование:

- **Технологии ИИ должны учитывать реализации права и свободы человека.** К таким правам относится и право оспорить решение, так или иначе влияющее на жизнь человека.
- **Не всегда у разработчиков есть возможность предоставить механизм оспаривания решений.** Например, при прогнозировании пробок на дорогах системы ИИ учитывают множество факторов, включая текущую ситуацию, погодные условия и время суток. Из-за сложности алгоритмов и большого объема обрабатываемых данных разработчикам может быть сложно создать механизм оспаривания таких решений.
- **Одним из принципов этики является комплексный надзор человека за системами ИИ.** Он включает в себя возможность отмены человеком значимых решений.
- По мнению исследователей Harvard Business Review[61], **в некоторых ситуациях алгоритмы, на которых основан ИИ, неспособны увидеть полную картину и не могут предложить обоснованное решение.** К таким ситуациям относятся те, где от ИИ для принятия обоснованного решения требовалось бы перенять человеческую эмпатию, а также руководствоваться этическими и моральными принципами.
- **В некоторых случаях решения ИИ носят рекомендательный характер и не оказывают прямого воздействия на жизнь человека.** Это снижает необходимость существования механизма оспаривания, а его внедрение затруднит оперативность и эффективность работы системы.
- Согласно исследованию, проведенному Debevoise Data Strategy & Security Group[62], **требование пользователя пересматривать человеком каждое решение ИИ, если он с ним не согласен, может неоправданно сдерживать инновации.** Вместо этого закон должен требовать от разработчиков и пользователей ИИ оценить и внедрить систему разрешения споров между человеком и компаниями, представляющими решения на базе ИИ, которая наиболее эффективно раскрывает ценность ИИ, снижая при этом риски как человеческих, так и машинных ошибок.

Практика:

Существует множество кейсов, в которых пересмотр человеком решения, принятого ИИ, являлся желательным.

1. Amazon создала модель на основе искусственного интеллекта, которая должна была помогать отбирать резюме наиболее квалифицированных кандидатов. Вот только обучена эта модель была на данных за предшествующие 10 лет, в течение которых в компании в основном работали мужчины. **Модель отдавала приоритет мужским резюме, тем самым занижая оценку женских резюме.** После многих попыток сделать программу гендерно нейтральной Amazon сдалась и отключила инструмент[63].

2. COMPAS (Correctional Offender Management Profiling for Alternative Sanctions) — американская система прогнозирования рецидива в уголовной юстиции. В 2016 году было проведено исследование, которое показало, что **COMPAS склонна к предвзятости и дискриминации по расовому признаку.** Проанализировав более 10 000 уголовных дел, исследователи установили, что вероятность рецидива была правильно предсказана только в 61% случаев, а по насильственным преступлениям — всего в 20%. Чернокожие обвиняемые чаще определялись как возможные рецидивисты, несмотря на остальные положительные факторы[64]. Так, в 2016 г. Эрик Лумис обжаловал использование алгоритма COMPAS для оценки его риска рецидива, утверждая, что из-за закрытости алгоритма он не мог проверить его точность. Верховный суд Висконсина отклонил его жалобу, постановив, что использование COMPAS не нарушает его права на справедливый процесс.

В подобных ситуациях у пользователей должна быть возможность связаться с разработчиком для получения дополнительной информации о причинах принятого решения, а также возможность оспорить такое решение в случае несогласия с ним.

Исследование по вопросу:

Согласно исследованию об отношении людей к внедрению ИИ, проведенному в Великобритании Институтом имени Алана Тьюринга[65], право оспорить решение ИИ было названо вторым важнейшим фактором доверия публики к ИИ. 59% опрошенных жителей Великобритании заявили, что хотели бы иметь четкие процедуры обжалования решения ИИ человеком.

Изучение представлений людей об ИИ показывает, что британская общественность не только за саму возможность оспаривания решения, принятого ИИ, но и за другие моменты, связанные с этим вопросом. Например, 47% опрошенных обеспокоены тем, что трудно определить, кто несет ответственность за ошибки при использовании данной технологии.

Отвечая на вопрос о том, кто должен отвечать за обеспечение безопасного использования ИИ, люди чаще всего выбирают независимый регулирующий орган: за это высказался 41%.

В исследовании Института им. Алана Тьюринга была приведена статистика по следующему вопросу: что из перечисленного ниже позволит вам с бóльшим комфортом пользоваться ИИ-технологиями?

Повышение комфорта людей при использовании всех технологий.
Какие из перечисленных ниже, если таковые имеются,
сделали бы использовании технологий ИИ более комфортными?

Источник: Институт им. Алана Тьюринга[65]

Подходы международных организаций:

1. **Рекомендации ЮНЕСКО об этических аспектах ИИ**[66] подчеркивают важность существования соответствующих механизмов, которые должны обеспечивать прозрачность онлайн-коммуникаций. Более того, пользователям следует предоставлять механизмы обжалования, позволяющие добиваться возмещения ущерба в случае нарушения ИИ их основополагающих прав и свобод.

2. **«Руководство по наилучшим практикам в сфере автоматизированного принятия решений»**[67], опубликованное омбудсменом Содружества наций в 2019 году, уделяет особое внимание политике обоснования решений. Такая политика будет способствовать информированию общественности и выявлению ответственного лица. Как правило, этой информации достаточно, чтобы пользователь мог сформировать мнение о принятом решении и в случае несогласия эффективно его оспорил.

Что думают эксперты?

Федор Коробков,
адвокат, основатель сервиса «Клиентправ»

— Человеческие решения неизбежно подлежат оспариванию. По той же причине следует оспаривать и решения, принимаемые искусственным интеллектом. В конце концов, ИИ — это продукт человеческой деятельности, а ошибки — естественная часть человеческой природы. Однако необходимо признать, что внедрение процессов для оспаривания решений ИИ может замедлить регулируемые процессы и увеличить нагрузку на систему. Несмотря на это, нельзя допускать исключений в вопросе оценки значимости решений ИИ. Игнорирование этого аспекта может привести к тому, что через открытое окно Овертона мы рискуем потерять нашу волю к независимому решению критически важных вопросов.

Роман Васильев,
президент АЛРИИ (Ассоциации лабораторий искусственного интеллекта)

— Человек всегда должен иметь право оспорить решение, принятое искусственным интеллектом. Несмотря на всю мощь и точность ИИ, его решения все еще основаны на алгоритмах и данных, которые могут быть неполными или ошибочными. Это особенно важно в вопросах, затрагивающих права человека, его здоровье или благосостояние. Прозрачность и возможность обжалования решений ИИ — это не только вопрос доверия, но и этики. Человек должен оставаться центральной фигурой в процессе принятия решений, особенно там, где от этого зависят жизни и судьбы людей. Искусственный интеллект — это инструмент, но ответственность всегда должна лежать на человеке.

Виктор Наумов,
главный научный сотрудник Института государства и права РАН, руководитель проекта «Сохраненная культура»

— На современном этапе любые юридически значимые решения с использованием ИИ требуют оспоримости. Оспаривание должно осуществляться путем обращения человека только к человеку с возможностью раскрытия логики принятия решения ИИ, что означает полное логирование функционирования ИИ. При этом за каждое решение ИИ должен нести юридическую ответственность владелец информационной системы, где используются технологии ИИ. Важно понимать, что человек в этих обстоятельствах является слабой стороной перед ИИ и владельцем системы и у него должен быть расширенный объем прав, включая право человека на отказ от использования от ИИ.

Андрей Незнамов,
управляющий директор Центра человекоцентричного AI Сбербанка, председатель Комиссии по реализации Кодекса этики в сфере ИИ

— Универсальное применение ИИ невозможно. Оспаривание решений, принятых с помощью ИИ, кажется важным в тех случаях, когда решения являются юридически значимыми. При разработке российской Концепции регулирования технологий ИИ эксперты сошлись на том, что оспаривание таких решений нужно, но нет необходимости уходить в крайность — создавать возможность для оспаривания любого решения, принятого с помощью ИИ, даже если оно не имело юридически значимых последствий.

07 Проблема предвзятости ИИ: можно ли ее решить?

Ответ:

Проблема предвзятости ИИ обусловлена исключительно используемыми для обучения данными и поэтому требует комплексного подхода, включающего в себя обеспечение разнообразия данных и тестирование моделей — так можно создать справедливые и этичные системы ИИ.

Рекомендации для разработчиков:

1. **Для обучения модели использовать наборы данных, обладающие максимальной полнотой информации.** Датасеты, в которых представлены разнообразные и репрезентативные данные, решают проблему предвзятости.

2. **Для уменьшения вероятности ответа модели, демонстрирующей предвзятую точку зрения, следует обучать модель отвечать максимально объективно, нейтрально и некатегорично.** Для этого можно привлекать профессиональных AI-тренеров — специалистов, которые способны оценить качество ответа и предложить нейтральные или более уместные формулировки.

3. **Важно проводить проверки и оценки моделей ИИ на предмет предвзятости, чтобы убедиться, что они не дискриминируют определенные группы людей.** Для этого можно использовать различные методы, такие как анализ чувствительности, тестирование на основе сценариев и т.д.

4. **Следует добавлять дисклеймер к выдаваемому моделью контенту, если нельзя обеспечить полную гарантию отсутствия в ответе стереотипа.** Хороший ответ содержит опровержение предрассудков и не поддерживает дискриминацию. Важно, чтобы дисклеймер был четким и понятным, а также соответствовал законодательству и этическим стандартам.

5. **Учитывать контекст пользовательского запроса.** Пользователь может обращаться к ИИ с разными задачами, и некоторые из них не подразумевают объективного ответа. Например: «Придумай три агрессивных приветствия для собеседника» или «Какой самый смешной фильм сняли в 2021 году?». В подобных случаях нужно понять, стоит ли вообще отвечать на такой вопрос или задачу. Если ответ положительный, то в ответе можно указать, что он не будет объективным. Если же на запрос пользователя модель не может ответить по этическим причинам, то необходимо корректно обозначить причину отказа.

Обоснование:

- Согласно исследованию McKinsey[68], **исходные данные, а не сам алгоритм, чаще всего являются основным источником проблемы предвзятости**. Модели могут быть обучены на данных, содержащих человеческие решения, или на данных, которые отражают последствия социального или исторического неравенства. Например, использование новостных статей для обучения может демонстрировать гендерные стереотипы, существующие в обществе.
- **При небольшом объеме обучающей информации ИИ будет формировать однобокие и ограниченные ответы.** Но чем шире датасет для обучения, чем больше разносторонней информации в нем содержится, тем более емкими, точными и объективными будут становиться ответы ИИ.
- ЮНЕСКО в 2024 году опубликовало доклад[69] **«Борьба с систематическими предрассудками: исследование предвзятого отношения к женщинам и девочкам в больших языковых моделях»**. Выделены 3 категории причин предвзятости в алгоритмах ИИ:
  - **Искажения в данных.**
    - погрешность измерения: возникает при выборе или сборе характеристик (например[74], алгоритм, предсказывающий возраст на основе роста);
    - искажение представления: когда обучающие наборы данных неадекватно представляют все группы, что приводит к плохому обобщению.
  - **Ошибки при выборе алгоритма.**
    - ошибка при агрегировании: использование единой модели для всех задач, которая не учитывает разнообразие данных;
    - предвзятость при обучении: возникает, когда выбор модели или процедуры обучения усиливает различия.
  - **Ошибки при внедрении.**
    - возникают, когда системы ИИ применяются в условиях, отличных от условий их разработки, что приводит к неприемлемым результатам.

Практика:

Специалисты лондонской компании DeepMind предложили в качестве защиты от влияния человеческих предубеждений использовать метод «гипотетическая справедливость» (counterfactual fairness). Чтобы сформулировать справедливое и непредвзятое суждение о гражданине, ИИ формирует гипотетическую ситуацию, в которой данный каждый гражданин обладает противоположными признаками: женщина превращается в мужчину, бедный — в богатого, афроамериканец — в белого. Таким образом, реальный статус не влияет на оценку деяний гражданина. Суждение формируется в гипотетической ситуации. Такое суждение считается свободным от предубеждений, а значит, справедливым[70].

Исследования по вопросу:

1. Исследователи из MIT и Microsoft обнаружили, что **технологии анализа лиц имеют более высокую частоту ошибок для темнокожих**, особенно для темнокожих женщин[71].
   Общие результаты исследования:
   - Все алгоритмы показали более высокие результаты при анализе лиц мужского пола, нежели женского (разница в частоте ошибок составляет 8,1%–20,6%).
   - Алгоритмы лучше работают на более светлых лицах, чем на темных (разница в частоте ошибок составляет 11,8%–19,2%).
   - Все алгоритмы хуже всего работают с более темными женскими лицами (частота ошибок составляет 20,8%–34,7%).
2. В 2019 году в США был проведен аудит алгоритма, предназначенного для прогнозирования объема необходимой медицинской помощи. В исследовании были проанализированы медицинские карты почти 50 000 пациентов, из которых 6 079 идентифицировали себя как чернокожих, а 43 539 — как белых. Были сравнены алгоритмические оценки риска с их фактическими историями болезни. **Исследователи обнаружили, что чернокожие пациенты, как правило, получали более низкие оценки риска**[72].
   В коде программы не было заложено предпочтение белокожим пациентам, алгоритм работал корректно. Ошибка заключалась в исходной идее разработчиков о том, что равные расходы на медицинскую помощь свидетельствуют об одинаковой потребности в ней. Алгоритм рассчитывал рекомендации на основе расходов пациентов на медицинскую помощь в прошлом. Однако расходы человека на медицинские услуги сильно зависят от его дохода и социального положения, поэтому алгоритм закрепил существующую ранее дискриминацию: пациенты, которые в прошлом получали меньше медицинской помощи из-за низкого уровня дохода, будут обделены ею и в будущем.

Подходы международных организаций:

В рамках Евразийского экономического союза принят технический регламент «О безопасности машин и оборудования»[73]. Разрабатываются схемы, удостоверяющие, что системы алгоритмических решений не демонстрируют неоправданной предвзятости. Например, сейчас развивается и совершенствуется Стандарт рассмотрения алгоритмических предубеждений (IEEE p7003). Этот этический стандарт закрепляет правила о том, как избежать непреднамеренных, необоснованных и неуместно различающихся результатов для пользователей.

Практика:

OpenAI заявляет[75], что борется с предвзятостью, изучая работу моделей на основе широкого спектра данных. Начальный этап — предварительная подготовка, на котором модель учится предсказывать следующее слово в предложении, основываясь на большом количестве интернет-текстов.
За ним следует второй этап, на котором модели «доводятся до совершенства» на основе более узкого набора данных, который тщательно формируется с привлечением специалистов-рецензентов. Также OpenAI советует учитывать общественное мнение о настройках и ограничениях.

Что думают эксперты?

Максим Годзи,
управляющий партнер Retentioneering

— Сегодня, когда проекты на основе искусственного интеллекта растут словно на дрожжах, этические проблемы встают еще острее. Одна из них — расизм. ИИ может быть предвзятым и иметь различные bias. Ведь он обучается на данных, в которых отражается текущий bias решений, которые принимают люди[75].

Максим Карлюк,
программный специалист Сектора социальных и гуманитарных наук ЮНЕСКО

— Существует так называемый закон Конвея. Он говорит о том, что системы в самом широком смысле этого слова, в том числе компьютерные программы или приложения для телефона, отражают ценности людей, которые их разрабатывают. То есть выбор действий или элементов в рамках процесса разработки программ зависит от того, как организованы проектирующие команды. Существующие предрассудки и другие негативные влияния часто игнорируются. В результате небольшая группа людей, работающих вместе над какой-то программой, в итоге получает большое влияние, когда результат их работы используется обществом[76].

Сергей Марков,
управляющий директор, начальник управления экспериментальных систем машинного обучения, ПАО Сбербанк

— Основные инструменты борьбы с предвзятостью систем ИИ — это совершенствование культуры подготовки данных и тестирования обученных моделей. При формировании обучающих выборок особое внимание должно уделяться достижению баланса групп прецедентов в датасетах, анализу возможных артефактов при формировании выборки (например, тенденции к повышенной вероятности попадания в выборку отдельных кейсов — помните шуточный опрос про наличие доступа к интернету, проведенный в интернете?), контролю того, что в обучающую выборку попадают все значимые факторы. Разумные и системные меры могут снизить риски до приемлемого уровня.

Александр Вечерин,
доцент департамента психологии факультета социальных наук НИУ ВШЭ

— Разработчики ИИ умеют успешно фильтровать оскорбительные и прямо негативные высказывания. К сожалению, многие негативные стереотипы не содержат ключевых признаков подобных высказываний, что создает большие трудности в фильтрации такого контента. Для решения этой задачи требуется на первом этапе изучить существующие предубеждения с лингвистической и психологической точек зрения, выявить характеристики, связанные с наиболее яркими эмоциональными реакциями пользователя, и разработать систему критериев оценки высказывания. Эти результаты в дальнейшем могут использоваться для допобучения моделей.

08 Проблема ответственности: на примере медицины, какова ответственность разработчика ИИ в случае причинения вреда здоровью пациента?

Ответ:

Юридическая ответственность разработчиков ИИ практически всегда регулируется отраслевым законодательством, где применяется система ИИ, — в данном случае медицинским. Если этот вопрос не урегулирован, разработчик может нести этическую ответственность, если были скрыты известные ошибки, не предприняты меры по исправлению сбоев или предоставлено недостаточно информации о возможных рисках системы. По общему правилу разработчик СИИ не несет этической ответственности за последствия использования системы, если риски и ограничения были явно и открыто доведены до медицинских работников.

Рекомендации

Для разработчиков:

1. **Используйте надежные источники информации** для создания качественных наборов данных для машинного обучения.
2. **Проводите испытания и регистрацию ИИ-системы**, чтобы подтвердить ее безопасность и эффективность на основании собранных доказательств.
3. **Обеспечивайте регулярное обновление ИИ-систем для соответствия новым медицинским стандартам и исследованиям.** Это поможет минимизировать риск использования устаревших данных и повысит актуальность системы в медицинской практике.
4. **Разрабатывайте ИИ-системы с возможностью объяснения, если это не конфликтует с качеством решения**. Важно, чтобы медицинские работники могли понимать, на каких данных и логике основываются рекомендации ИИ. Это повысит доверие пациентов и снизит риск ошибок.

Для медицинских работников:

1. **Поддерживайте постоянный диалог между разработчиками и врачами** о возможных ошибках и ограничениях системы, чтобы минимизировать риски для пациентов.

2. **Соблюдайте принципы осторожности и взвешенности при принятии решений с использованием СИИ.** Оценивайте потенциальные риски для понимания, когда необходимо обратиться к разработчику для получения дополнительных инструкций.

3. **Интегрируйте ИИ как вспомогательный инструмент, а не замену человеческому анализу.** Рекомендации ИИ могут быть полезны, но всегда должны рассматриваться лишь как дополнение к клиническому мнению и опыту врача.

4. **Проводите проверку данных и рекомендаций, предложенных ИИ, перед применением их в лечении**. Особенно важно это делать в сложных случаях или при наличии сомнений в точности предложений системы, чтобы избежать неправильных решений.

Обоснование:

- По мнению группы исследователей из ОАЭ и Египта, **несправедливо возлагать ответственность на разработчиков, так как системы ИИ работают автономно и не все ошибки могут быть предвидены или предотвращены на этапе разработки**. Производители несут ответственность за дефекты, связанные с процессом проектирования или не соответствующими инструкциями, лишь в тех случаях, когда можно предсказать риск причинения вреда, непосредственно связанный с продуктом[77].

- Согласно исследованию американской адвокатской фирмы Leeseberg Tuttle, **в настоящее время ответственность за вред пациенту продолжает возлагаться на того медицинского работника, который оказывал помощь**, независимо от использованных им инструментов. Более того, исследование показало, что основной причиной часто является человеческая ошибка[78].

- **Некачественная работа СИИ вне заявленных характеристик может быть связана с неисправностью**, которую лечащий врач не мог предвидеть или обнаружить.

- Использование ИИ для оказании медицинской помощи **мало чем отличается от использования любого другого медицинского изделия**.

Исследование по вопросу:

Согласно исследованию McKinsey от 2022 года[79], технологии генеративного ИИ представляют собой новый значимый инструмент, который может помочь раскрыть часть нереализованного потенциала медицинской отрасли. Это возможно

за счет автоматизации утомительной и подверженной ошибкам оперативной работы, доведения многолетних клинических данных до сведения врача за считаные секунды и модернизации инфраструктуры систем здравоохранения. Совместные инвестиции в эти области могут принести прибыль в размере от 1 трлн до 1,5 трлн долларов.

Потенциальная прибыль в сфере здравоохранения за счет внедрения ИИ
Источник: McKinsey[79]

Подход Европейского союза:

В 2014 году Комиссия по этике исследований в области цифровых наук и технологий Европейского союза (CERNA) предложила несколько рекомендаций по этичному применению роботов в медицинской сфере. Они были опубликованы Европейским парламентом:[80]

- Исследователи должны запрашивать мнения, публикуемые действующими медицинскими этическими комитетами, и следовать им.
- Исследователи, работающие над роботизированными системами, должны стремиться сохранять автономию и контроль людей, в отношении которых применяются технологии.
- Исследователи должны гарантировать, что все действия роботизированных систем останутся обратимыми.

Подход РФ:

Медработник не несет персональную гражданско-правовую ответственность за оказанную медицинскую помощь в силу положения ст. 1068 ГК РФ об ответственности юридического лица как работодателя за действия работников.

Согласно ст. 1096 ГК РФ, вред, причиненный вследствие недостатков услуги, подлежит возмещению лицом, оказавшим услугу. Таким образом, в ситуации, когда медицинское ИИ-оборудование соответствовало всем требованиям сертификации к стандартам медицинской помощи, требования о вреде предъявляются к медицинскому учреждению (как к юридическому лицу, которое использует определенное оборудование).

Исследование по вопросу:

Исследование французских компаний MACSF и WITHINGS показывает, что из 1037 врачей-членов MACSF 43%, 30%, 27%, и 25% персонала, работающего с подключенными устройствами, использовали их часто или всегда, для постановки диагноза, удаленного отслеживания, первичной или вторичной профилактики[81].

Кроме того, более трети врачей по-прежнему с осторожностью относятся к применяемому режиму ответственности в случае, если средство, которое они рекомендовали, привело к ухудшению здоровья пациента.

Практика:

1. В 2017 году робот-стоматолог, разработанный китайскими специалистами, впервые без участия врачей прооперировал пациента. Робот провел успешную имплантацию двух зубов, ранее напечатанных на 3D-принтере. По итогам операции установлено, что импланты поставлены с погрешностью 0,2–0,3 мм, что допустимо для врачебных стандартов. Уточняется, что решение о создании робота было принято на фоне нехватки в Китае квалифицированных стоматологов[82].

2. В настоящее время инструменты нейровизуализации требуют проведения МРТ-сканирования с несколькими требованиями, включая разрешение и контрастность для точного 3D-анализа. Однако большинство МРТ-сканирований по всему миру не соответствуют требуемым критериям. Поэтому исследователи Гарвардской медицинской школы разработали систему ИИ SynthSR для преобразования МРТ-снимков с низким разрешением. Такое повышение качества изображения может революционизировать их использование в критических состояниях или в местах с ограниченными медицинскими возможностями, где нет оборудования для МРТ[83].

3. Исследователи из подразделения Breast Cancer Now в Королевском колледже Лондона создали ИИ-модель для прогнозирования вероятности распространения рака молочной железы у пациенток с тройным негативным раком молочной железы[84]. ИИ-модель, платформа глубокого обучения под названием smuLymphNet, используется для проведения анализа изображений лимфатических узлов у больных раком, сопоставления с записями пациентов и определения того, распространился ли рак.

Что думают эксперты?

Антон Киселев,
заместитель директора по научно-технологическому развитию ФГБУ «Национальный медицинский исследовательский центр терапии и профилактической медицины» Минздрава России

— В связке «врач–ИИ» роль ИИ обычно заключается в поддержке врачебных решений или, как максимум, используется при необходимости второго мнения. При этом допуск ИИ к выполнению подобных функций в практической медицине строго регламентирован. ИИ-сервисы, непосредственно участвующие в процессе оказания медицинской помощи, попадают под обязательную госрегистрацию по правилам, применяющимся к медицинским изделиям. Именно на этом уровне и происходит разграничение сфер ответственности за допуск ИИ-сервиса в клиническую практику. Ответственность же за принятие решения по конкретному пациенту остается полностью на враче, независимо от того, принимал он во внимание «мнение» ИИ-помощника или нет.

Вячеслав Шуленин,
генеральный директор АНО «Московский центр инновационных технологий в здравоохранении»

— Несмотря на безграничный потенциал использования нейросетей, невозможно полностью заменить специалистов, потому что принятие решений — ответственность человека. И ни один компьютер или система не могут быть оценены как субъекты правовой и этической оценки действий, а также их последстви[85].

Андрей Алмазов,
заместитель директора по проектной деятельности Ассоциации «Национальная база медицинских знаний»

— В задаваемом вопросе требует уточнения, какого типа вред и каким способом в силу использования ИИ он может быть причинен. Например, предположим, что доза излучения на КТ стала регулироваться ИИ. В этом случае разработчик, несомненно, юридически отвечает за безопасность, но не ИИ, а всего медизделия. Наличие ИИ ничего не меняет в сравнении с текущей практикой: вопрос здесь именно юридический, а не этический. Иной полярный умозрительный случай: ИИ растолковал результаты анализов или исследований пациенту таким образом, что нанес ему психологическую травму. Тогда это действительно вопрос этики, но подобная ситуация типична и без ИИ: ухудшение физического или эмоционального состояния человека, ненамеренно спровоцированное медицинским работником. Кто тут отвечает? Видимо, разработчик, потому как ИИ не является субъектом, но становится соучастником процесса, вторгаясь в отношения, которые ранее оставались только в контуре «врач–пациент».

09 Проблема делегирования принятия решений: на примере правосудия, сможет ли ИИ заменить судью?

Ответ:

Для делегирования решений ИИ всегда необходимо учитывать требования законодательства и позиции органов судебной власти по этому вопросу (если они есть). Если закон допускает такую возможность, с этической точки зрения ИИ можно делегировать самостоятельное рассмотрение небольшой категории дел, где не требуется учета субъективной (психологической) стороны поведения участников процесса и оценки их личности. В остальных случаях роль ИИ может сводиться к роли помощника судьи в части подбора информации и ее анализа.

Рекомендации по внедрению ИИ в суды:

1. **Прежде всего оцените, в каких случаях закон разрешает использование ИИ и каким образом.**
2. **Обеспечьте контроль человека над применением ИИ-систем в правосудии.** Решения, предложенные ИИ, должны проверяться и утверждаться судьей или другим ответственным специалистом, чтобы исключить автоматизацию критических ошибок и сохранить человеческий контроль над процессом.
3. **Применяйте только специализированные закрытые модели ИИ, обученные на проверенных данных.** Использование общедоступных моделей, обученных на открытых данных из интернета, недопустимо, так как может привести к ошибочным выводам и подрыву доверия к судебному процессу.
4. **Верифицируйте данные для обучения модели со стороны профессионального сообщества и государства.** Это поможет гарантировать, что модель использует актуальные и достоверные правовые позиции, которые соответствуют законодательству и судебной практике.
5. **Своевременно обновляйте модель.** Регулярное дообучение на новых правовых позициях обеспечит соответствие модели современным стандартам и требованиям правосудия.
6. **Программируйте модели таким образом, чтобы они могли сообщать о недостатке данных для принятия решения.** Модель должна уметь информировать пользователей, если имеет недостаточно данных для обоснованного вывода. Она должна избегать принятия неверных решений.

7. **На любом этапе применения ИИ в правосудии участники процесса должны иметь право оспорить решения, если они были приняты в их отношении с использованием ИИ.** Возможность пересмотра решений, принятых с помощью ИИ, обеспечивает дополнительный уровень защиты прав участников и позволяет устранить ошибки.
8. **Решение вопроса доверия общества к применяемым системам требует прозрачности.** Раскрытие информации об используемой модели и предоставление участникам процесса доступа к ее выводам участникам процесса помогают повысить доверие и понимание работы ИИ в судебных делах.
9. Большинство из рекомендаций выше применимы в целом для решения вопроса об этичности делегирования принятий решений ИИ.

Обоснование:

- **При разрешении дела судья учитывает нравственные аспекты** (например, гуманность, соразмерность) и субъективные факторы ситуации (разумность, добросовестность), что находится в эмоциональной сфере человека и непостижимо для ИИ.
- Согласно исследованию, опубликованному Международным журналом по судебному администрированию, **решения алгоритма не могут самостоятельно использоваться в качестве предписания**. Так, ИИ не может быть допущен к решению вопросов виновности подсудимого в уголовном судопроизводстве, поскольку они связаны с оценкой субъективной стороны поведения подсудимого[86].
- **ИИ может принимать самостоятельные решения по судебным делам, в которых не изучаются психологические аспекты поведения сторон**, а решение принимается без устного разбирательства на основе письменных доказательств — бесспорные (судебные приказы) и малоспорные гражданские дела (малые иски на определенную сумму).
- Исследователи из Lexis Nexis, международной компании, работающей в сфере информационных услуг, считают, что **ИИ предлагает кратчайший путь для оптимизации процесса анализа судебных дел**. Для некоторых категорий дел помощь ИИ с подбором и анализом информации по делу, подготовкой прогноза рассмотрения дела (предикативного отчета) и текста судебного акта может быть очень значительной, но решение все же выносит судья-человек[87].

Подход регуляторов:

1. В декабре 2018 года появился первый международный акт, специально посвященный использованию ИИ в правосудии: **Европейская этическая хартия использования технологий ИИ в судебных системах и смежных сферах,** утвержденная Европейской комиссией по эффективности правосудия Совета Европы. В Хартии делается акцент на необходимость полностью гарантировать соблюдение прав человека, принципа равенства сторон, презумпции невиновности, прозрачности и недискриминации при использовании ИИ-технологий в судебной системе[88].
2. На национальном уровне также принимаются руководства по этичному использованию ИИ в судопроизводстве. Например, в июле 2024 года в Гонконге было разработано **«Руководство по использованию генеративного ИИ**

для судей и должностных лиц судебной системы»[89]. Судьи и обслуживающий персонал судебных органов могут разумно и ответственно использовать генеративный ИИ там, где это уместно. Однако запрещается делегирование судебных функций ИИ. Все судебные решения должны приниматься исключительно судьями.

3. В декабре 2023 года в Англии было опубликовано[90] **«Руководство для работников судебных органов, судов и трибуналов»**. Согласно Руководству, ИИ может быть полезен для обобщения больших объемов информации или выполнения административных задач, но не рекомендуется использовать его для проведения юридических исследований из-за риска «галлюцинаций» и фактических ошибок.

Практика:

1. В 2021 году **Верховный суд КНР обязал судей при принятии решений консультироваться с ИИ**. Система «Умный суд», запущенная в 2015 году, автоматически просматривает судебные дела на предмет верных ссылок, рекомендует законы и нормативные акты, разрабатывает юридические документы и исправляет предполагаемые человеческие ошибки при вынесении вердикта, если таковые имеются[91].

2. По инициативе Министерства юстиции Франции **два апелляционных суда весной 2017 года согласились протестировать программное обеспечение «Predictive justice»** для рассмотрения апелляций. Этот ИИ-инструмент предлагал решение судьям на основе анализа гражданских дел всех апелляционных судов Франции. По мысли разработчиков, это должно было способствовать принципу равенства граждан перед законом[92].

3. **Российская судебная система также приступила к тестированию ИИ.** Пилотный проект стартовал в Белгородской области: три судебных участка мировых судей подключили ИИ для подготовки судебных приказов по взысканию налогов с граждан: имущественного, транспортного и земельного. ИИ-инструменты должны помогать судьям готовить документы, в том числе создавать карточку дела во внутренней системе суда[93].

4. В Колумбии судья Хуан Мануэль Падилья (Juan Manuel Padilla) рассматривал дело о покрытии расходов на медицину и транспорт для ребенка с расстройством аутистического спектра. Предстояло выяснить, должны ли все расходы покрываться страховкой, так как родители ребенка не могли позволить себе это. Судья спросил у ChatGPT, следует ли освободить семью ребенка от платы за лечение. Нейросеть ответила, что согласно законам Колумбии люди с аутистическим расстройством освобождаются от платы за терапию. Решение суда совпало с ответом чат-бота. При этом в интервью судья рассказал, что окончательное решение принимал самостоятельно и использовал для этого прецеденты из предыдущих постановлений. **Консультация с нейросетью помогла ускорить процес**[94].

Анализ мировой практики показал, что наиболее популярными этическими принципами для ИИ в судебной системе являются:

1. Принцип соблюдения прав человека, в силу которого применение ИИ не должно умалять состязательность процесса и права на справедливое разбирательство.
2. Принцип качества и безопасности, который предполагает использование сертифицированного программного обеспечения, оценка которого проводится как техническими специалистами, так и юристами.
3. Принцип человеческого контроля, согласно которому судья и участники спора должны иметь возможность не согласиться с решением, предложенным ИИ, и оспорить его.
4. Принцип запрета дискриминации, включающий в себя запрет на использование данных, которые могут привести к предвзятости в отношении определенных групп людей.
5. Принцип прозрачности, в силу которого все особенности применяемых технологий должны быть доведены до сведения граждан в доступной форме и понятным языком.

Что думают эксперты?

Виктор Момотов,
председатель Совета судей России, д.ю.н

— Искусственный интеллект не может стать гарантом защиты прав и свобод человека и обеспечить справедливое и гуманное правосудие. Его применение возможно только в ограниченном виде, с четко определенными рамками и правилами. Взаимодействие судей и работников аппарата суда с технологией искусственного интеллекта должно приводить к синергии при сохранении главенствующей роли человека. По мере развития информационных технологий сфера их применения расширяется от технических и рутинных функций к решению более сложных задач, а информационные системы становятся средой осуществления процессуальных действий[96].

Елена Авакян,
вице-президент
Федеральной Палаты адвокатов РФ

— Применение ИИ в правосудии зависит от сферы, в которой рассматривается спор. Категорически не может быть замещен судья в уголовном судопроизводстве. Потому что здесь мы судим человека, его поступок, его субъективную сторону. Соответственно, допускать машину к суждениям о действии человека — значит проиграть видовую конкуренцию. Что касается административно-правового и гражданско-правового спектра дел, то уже есть дела приказного и упрощенного производства, где ИИ не может, а должен заместить судей. Здесь ИИ будет работать на жесткое применение нормы в конкретных условиях.

Анатолий Выборный,
заместитель председателя Комитета Государственной думы по безопасности и противодействию коррупции

— Безболезненно искусственному интеллекту можно поручить все налоговые споры, а также оспаривание решений органов власти, например решений государственной экспертизы дорожного движения. Подчеркну, речь идет о небольших суммах и простых спорах: то есть в данном случае мы говорим о налоговых спорах и административных взысканиях за нарушения правил дорожного движения[96].

Андрей Незнамов,
управляющий директор Центра человекоцентричного AI Сбербанка, председатель Комиссии по реализации Кодекса этики в сфере ИИ

— Важно, чтобы ИИ помогал разгружать суды, выполняя задачи, носящие автоматический, рутинный характер, то есть те задачи, в которых применение ИИ позволило бы снизить количество ошибок. Однако все это должно строго соответствовать процессуальным нормам конкретной страны. Поэтому нам важно, чтобы регуляторы постепенно создавали процессуальные рамки для внедрения ИИ[97].

10 Проблема социального рейтингования: этично ли применять ИИ для создания социального рейтинга?

Ответ:

Основной этический вопрос состоит не в применении ИИ, а скорее в применении самого социального рейтинга — и этот вопрос является исключительно дискуссионным. Тем не менее исследования показывают, что этическими предпосылками применения системы социального рейтингования являются предварительное обсуждение с общественностью и прозрачность применения системы рейтинга.

Исследовательские рекомендации:

1. **Перед разработкой и внедрением системы социального рейтинга необходимо проводить многоэтапное общественное обсуждение с участием экспертов, правозащитников и представителей разных социальных групп для выработки этических принципов, которые будут заложены в систему возможного рейтинга.** Распространение социального рейтинга, влияющего на все сферы жизни, без всеобщего обсуждения и согласования можно считать неэтичным.
2. **Важно развивать этическую и правовую базу для регулирования системы возможного социального рейтинга.** На этой базе создавайте институты общественного контроля: так потенциал новых технологий может быть реализован без ущерба для фундаментальных прав человека.
3. **Важно обеспечивать прозрачность системы социального рейтинга.** Пользователям необходимо предоставлять информацию в доступной для понимания форме о том, какие их данные могут использоваться, рассматриваться как общедоступные, а также каково влияние этой информации на рейтинг.
4. **Важна возможность обжалования рейтинга.** Каждый человек должен иметь возможность узнать свой рейтинг, а также оспорить его корректность или последствия, если это необходимо, чтобы предотвратить несправедливые санкции или ошибки.
5. **Необходимо создавать систему защиты данных, которая будет гарантировать конфиденциальность и безопасность личной информации.** Используемые данные должны быть защищены от несанкционированного доступа,

чтобы предотвратить злоупотребления и утечки.

6. **Проводите регулярный аудит и оценку работы системы социального рейтинга.** Важно следить, чтобы система оставалась справедливой и не ущемляла права отдельных групп населения. Регулярные независимые проверки помогут выявлять возможные риски и недочеты в работе рейтинга.

Обоснование:

- Ученые Ближневосточного технического университета Турции утверждают, что **система социального рейтинга преследует законные и общественно полезные цели**. Данные должны собираться и анализироваться из различных источников с целью создания безопасного и так называемого общества, основанного на доверии[98].
- Исследователи Владимирского государственного университета подчеркивают, что **применение социального рейтинга без должного правового регулирования может нарушить конфиденциальность граждан**. С точки зрения охраны права на неприкосновенность частной жизни важно, что источниками информации для социального рейтинга является разного рода персональная информация[99].
- Группа исследователей из Израиля и Японии считает, что **наибольшие риски связаны с непрозрачностью системы социального рейтинга**. Не всегда ясно, какие факторы и как сильно влияют на балл оцениваемого лица. В результате определенные формы контроля могут оказать реальное влияние на жизнь людей из-за сложившихся в них «предрассудков», а также ошибок системы[100].
- **Социальный рейтинг несет риски столкновения различных обществ и потерю свободы личности**. Социальная среда представлена множеством обществ: традиционные, консервативные, религиозные, технократические, авангардные — зачастую с противоположными ценностями и убеждениями.

Социальный рейтинг — система контроля социальной деятельности граждан, которая оценивается по нескольким параметрам. На основе оценки, представляющей собой рейтинговый балл, рассчитываемый с применением специальных алгоритмов цифровой обработки совокупности определенных данных, формируется спектр возможностей и услуг, которыми может воспользоваться тот или иной гражданин. В зависимости от количества одобренных системой критериев проставляется величина рейтингового балла: чем выше балл, тем больше привилегий и возможностей у гражданина, меньше ограничений[101].

Подход ЕС:

Регламент Европейского союза об искусственном интеллекте запрещает использовать системы социального рейтинга[102]. По мнению законодателей ЕС, системы ИИ, обеспечивающие социальную оценку физических лиц государственными или частными субъектами, могут нарушать право на достоинство и недискриминацию, а также ценности

равенства и справедливости. Социальная оценка, полученная с помощью таких систем ИИ, может привести к негативным последствиям, которые несоразмерны тяжести поведения человека.

Подход ЮНЕСКО:

Рекомендации ЮНЕСКО об этических аспектах ИИ закрепляют принцип принятия человеком окончательного решения в тех случаях, когда предполагается, что решения имеют необратимые последствия, которые трудно обратить вспять, или могут касаться вопросов жизни и смерти. В частности, системы искусственного интеллекта не должны использоваться для социальной оценки или массового наблюдения[103].

Система социального рейтинга в Китае:

- **Наиболее масштабный пример применения социального рейтинга реализован в Китае**[104]. Он охватывает более 1 млрд человек и учитывает 160 тыс. различных параметров, включая такие факторы, как кредитная история, соблюдение законов, своевременная оплата счетов, волонтерская активность и даже высказывания в социальных сетях.
- Органы власти или компании оценивают социальное поведение человека от 0 до 1000 или от A до D. **Рейтинг складывается из информации**, полученной из официальных источников: налоговой, правоохранительных органов, правительственных структур, а также из данных цифровых источников — истории поиска, онлайн-покупок и активности в соцсетях.
- **Граждане с низким социальным рейтингом попадают в «черный список»**. Таким гражданам могут отказать в оформлении кредита, ипотеки или в приеме детей в частную школу. Важно отметить, что выйти из бан-листа возможно: например, если человек займется общественно полезной деятельностью.
- **Сторонники** социального рейтинга утверждают, что он способен сделать общество более безопасным, справедливым и эффективным, побуждая людей вести себя более ответственно и этично. **Критики** же видят в нем инструмент тотального контроля, нарушения приватности и ограничения свободы.

Аналогичный эксперимент в Венесуэле:

Смарт-карта Венесуэлы, известная как **«национальная карта», собирает разнообразную информацию о владельцах карт и сохраняет ее в государственной базе данных**, что, по утверждению правительства, поможет им предоставлять гражданам более качественные услуги. В базе данных, по словам сотрудников карточной системы, хранится целый ряд сведений, включая историю болезни, присутствие в социальных сетях, членство в политической партии и то, голосовал ли человек на выборах[105].

Исследования по вопросу:

1. В период с февраля по апрель 2018 года немецкие исследователи в сотрудничестве с китайскими компаниями провели общенациональный онлайн-опрос[106] с целью выявить, **как менялось поведение китайских граждан после внедрения системы социального рейтинга**. В опросе поучаствовало более 350 000 китайцев. Результаты показывают, что большинство респондентов (94%) сообщили об изменении поведения. Эти изменения

часто были вызваны стремлением улучшить личные показатели: 91% респондентов хотя бы раз изменили свое поведение, чтобы положительно повлиять на свой рейтинг (например, участвовали в благотворительности). А 85% сообщили, что хотя бы раз изменили свое поведение, чтобы избежать наказаний/ограничений (например, тщательнее соблюдали правила дорожного движения).

2. В январе 2024 года немецкие ученые провели **исследование степени принятия гражданами стран из Юго-восточной Азии систем социального рейтингования и непосредственно китайской системы**. Среди респондентов 50% полностью или в какой-то степени одобрили бы введение системы социального рейтинга в своих странах, в то время как только 15% высказались решительно или в какой-то степени против[107].

Что думают эксперты?

Сюэ Лань,
директор Института международного регулирования ИИ, декан Колледжа Шварцмана при Университете Цинхуа

— Система претерпевает значительные изменения, она все еще находится на стадии тестирования. Необходимо учитывать, что население Китая составляет 1,4 млрд человек, и существует множество проблем, требующих решения. Сообщения о том, что «большой брат» пытается отобрать у всех все, не соответствуют действительности. Я не вижу, чтобы социальный рейтинг давал Китаю какие-то особые преимущества, я не думаю, что правительство использует его для получения какой-либо коммерческой выгоды. В Китае нет доказательств этого[108].

Хольгер Цшайге,
генеральный директор «Инфотропик Медиа», сооснователь Moscow Legal Hackers, амбассадор European Legal Technology Association (ELTA)

— Социальный скоринг в Китае — отдельный и уникальный случай. Это не попытка минимизировать коммерческие риски, но метод тотального контроля над населением. Это то, о чем нас предупреждали авторы книг и художественных фильмов начиная с Оруэлла. Проблема социального скоринга в том, что государство произвольно может задать параметры и таким образом превратить жизнь в сплошной ад[110].

Роман Душкин,
глава компании-разработчика искусственного интеллекта «А-Я эксперт»

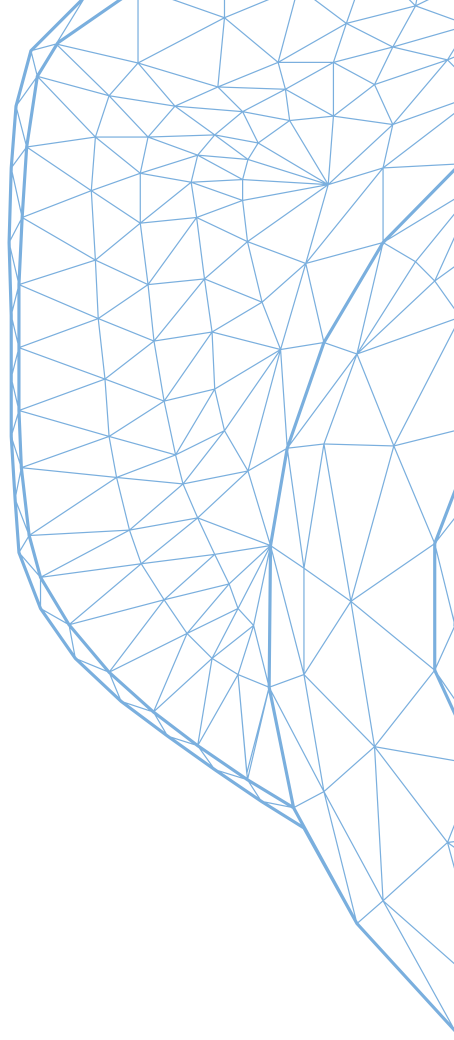
— Если такая система будет внедрена, если она будет распространяться на всех, это мгновенно приведет к стратификации общества, которое и так разделено на слои и страты, а при помощи системы социального рейтингования это будет выпячено наружу[110].

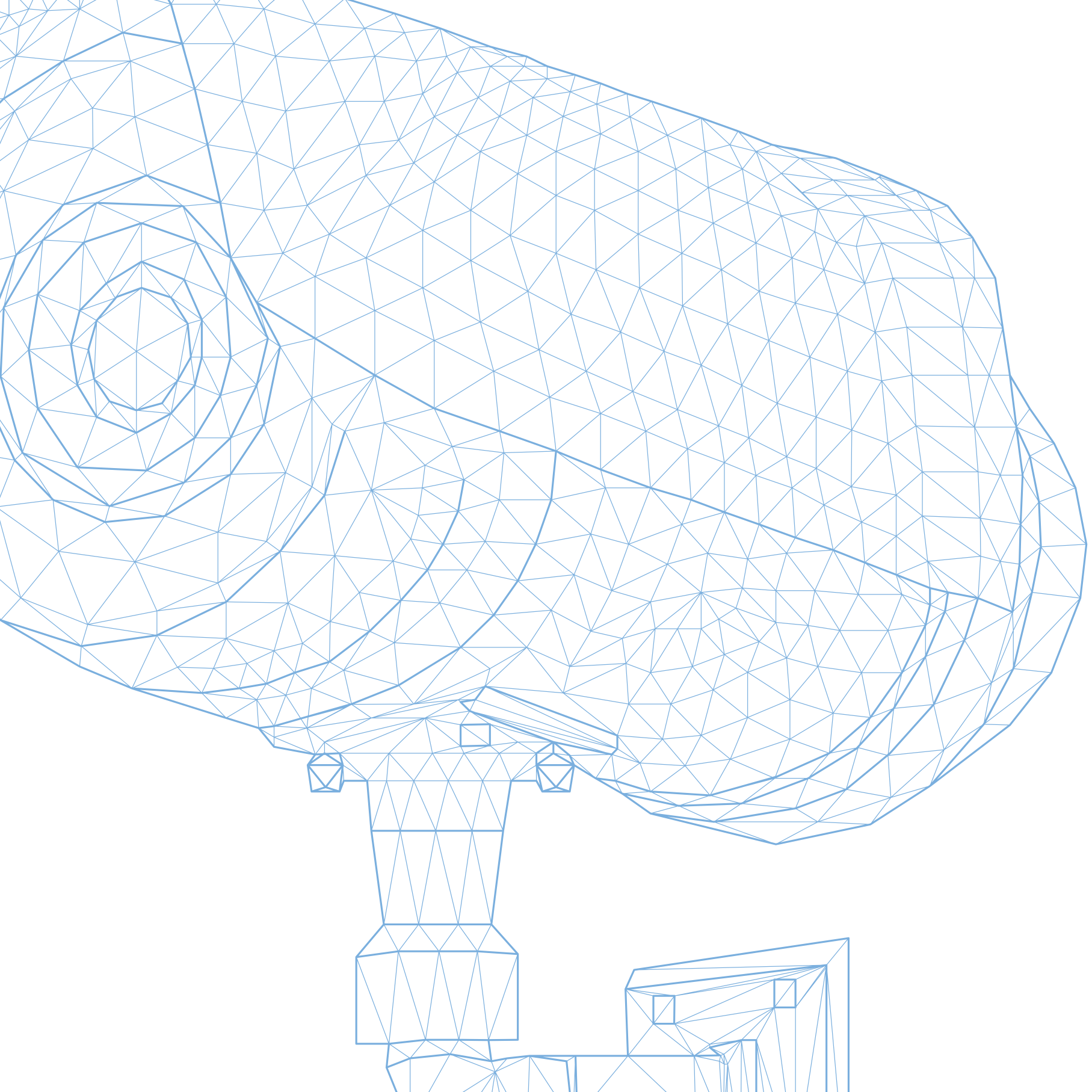
Андрей Свинцов,
заместитель председателя Комитета ГД по информационной политике, информационным технологиям и связи

— Внедрение скоринга приведет к массовому выгоранию людей, особенно молодых, которые будут стремиться иметь высокий рейтинг. В конце концов мы получим нацию не людей, а роботов. В современной России даже при высочайшем уровне развития искусственного интеллекта такие системы неприменимы и, считаю, недопустимы[111].

глава 2

ИИ и конфиденциальность





11 Этично ли использовать персональные данные для обучения ИИ?

Ответ:

Этично только с учетом соблюдения требований законодательства о конфиденциальности личных данных, соблюдения прав субъектов персональных данных и только в тех случаях, когда это необходимо.

Обоснование:

- ОЭСР в своем докладе «ИИ, управление данными и конфиденциальность»[112] напоминает, **что доступность данных не означает, что их можно собирать и использовать для обучения моделей ИИ**. Персональные данные (ПДн) должны быть получены законным образом, а любое использование — совместимо с первоначальными целями.
- **ПДн можно использовать для проведения научных исследований, в том числе в интересах всего общества,** с соблюдением норм законодательства. Например, в сфере медицины для изучения новых методов лечения и разработки лекарственных препаратов.
- **Нельзя для обучения ИИ использовать данные, собранные для другой цели**. Использование персональных данных в машинном обучении является самостоятельной целью обработки и требует законного основания[113].
- Управление информационного комиссара Великобритании подчеркивает, что **модели машинного обучения, обученные на ПДн, могут непреднамеренно усилить дискриминацию**[114] Например, данные из резюме прошлых соискателей для обучения системы ИИ, используемой при найме, могут разжечь дискриминациию по половому признаку, поскольку мужчины долгое время считались более подходящими кандидатами на определенные должности.
- Согласно исследованию, опубликованному heyData, **ПДн в машинном обучении используются для повышения качества и эффективности цифровых сервисов**[115]. Применение передовых методов защиты ПДн позволяет объединить конфиденциальность ПДн и точность (полезность) модели машинного обучения.

Рекомендации для разработчиков:

1. **Ответственно подходите к принятию решения об обучении ИИ-модели на персональных данных**. Если нет понимания, как ПДн могут помочь в обучении модели и к каким результатам это может привести, то ПДн лучше не использовать.

2. **Получите согласие субъекта на использование ПДн в машинном обучении, если это предусмотрено законодательством**. Если согласия не требуется, этичным считается в любой форме предупредить пользователя об использовании его ПДн в машинном обучении.
3. **Исключите возможность несанкционированного получения ПДн в результате обучения модели**. Используйте методы анонимизации и шифрования данных. Ограничьте круг лиц, имеющих доступ к ПДн; проводите регулярный мониторинг и аудит системы для выявления потенциальных угроз и нарушений безопасности.
4. **Определите чувствительные данные (религиозные убеждения, сексуальная ориентация, психические заболевания и прочие), которые могут привести к несправедливости**. Оцените их значение для процесса. Взвесьте справедливость модели машинного обучения как с точки зрения интересов отдельного человека, так и социальных групп.

Исследование по вопросу:

Согласно материалам IBM, **федеративное обучение** — это метод, который позволяет настраивать централизованную модель машинного обучения без передачи данных между устройствами, что значительно повышает уровень конфиденциальности[116].

В этой системе каждое устройство обучает свою копию модели на локальных данных, которые не передаются на сервер. Устройства отправляют только обновленные параметры, которые объединяются на сервере для обновления общей модели, пока не достигается нужная точность.

Таким образом, федеративное обучение минимизирует риск утечки данных, так как информация не перемещается между устройствами или сторонними сервисами. Передаются исключительно параметры модели, что позволяет обеспечить конфиденциальность данных на каждом этапе обучения.

Практика:

1. Компания Clearview AI собрала миллиарды изображений из социальных сетей без согласия пользователей и создала систему распознавания лиц для продажи правоохранительным органам и частным компаниям. **Поскольку фотографии были получены без разрешения, многие страны признали эту практику незаконной**. Clearview AI столкнулась с многочисленными исками, а также с запретом на свою деятельность в некоторых странах (например, в Австралии и Франции)[117].

2. Весной 2023 года итальянское ведомство по защите данных запретило доступ к ChatGPT **в связи с утечкой данных пользователей**[1118]. Кроме того, OpenAI не уведомляла пользователей о том, что собирает их данные для обучения алгоритмов. Следовательно, было нарушено требование GDPR о правовой основе обработки и хранения персональных данных.

3. В мае 2024 года **Meta заявила о том, что будет использовать персональные данные пользователей для обучения ИИ**, а именно фотографии, опубликованные и публично доступные в сервисах компании. Пользователям представлена возможность отказаться от использования их персональных данных для обучения. Однако механизм отказа довольно сложен. Пользователи должны заполнить длинную форму, указав подробную причину отказа. 6 июня 2024 года было подано 11 жалоб на Meta в суды по всей Европе. В ответ на это Meta публично обвинила истцов в препятствовании развитию генеративного ИИ[119]

Что думают эксперты?

Артем Шейкин,
Первый заместитель председателя Комитета Совета Федерации по конституционному законодательству и государственному строительству

— Этичность использования персональных данных для обучения ИИ во многом зависит от соблюдения ряда принципов: законности, согласия, конфиденциальности, соответствия целям использования, а также готовности разработчиков ИИ нести ответственность за свои действия.
Таким образом, этот процесс должен полностью соответствовать действующему законодательству в области персональных данных, граждане должны дать согласие на использование своих данных, а также быть проинформированы о том, как они будут использоваться и для каких целей. Кроме того, данные должны быть защищены от утечек и обезличены, чтобы снизить риск их использования в противоправных целях.

Эдуард Лысенко,
министр правительства Москвы, руководитель департамента информационных технологий (ДИТ) Москвы

— Обезличенные персональные данные крайне важны для обучения ИИ. Например, система поддержки врачебных решений не сможет подсказать врачу-радиологу, что на конкретном снимке КТ в конкретной области легкого есть подозрение на опухоль, если эта система предварительно не будет обучена на тысячах снимков. При этом для обучения системы ей не надо знать, кому принадлежит каждый из этих снимков, а только лишь научиться распознавать злокачественные опухоли. Аналогичная ситуация с системами в других областях — образование, транспорт, экология и т.д.[120].

12 Этично ли собирать данные пользователей со смартфона или умного устройства для обучения ИИ?

Ответ:

Неэтично, если данные собираются без соблюдения норм законодательства, в частности об информировании пользователя и получении его согласия на сбор и обработку данных. В отсутствие такого законодательства — неэтично без предупреждения пользователя.

Обоснование:

- **Сбор и обработка данных пользователя регламентируются гражданским законодательством, законодательством о персональных данных и о связи**. Как правило, пользователь должен быть проинформирован об обработке персональных данных и дать свое согласие на их использование.

- ЮНЕСКО предупреждает[121], **что данные, собираемые IoT (интернет вещей), легко объединить для создания высокоточного профиля человека**. Даже если персональные данные собирались с учетом требований законодательства, их объем и разнообразие могут привести к угрозе конфиденциальности.

- Управление информационного комиссара Великобритании считает, что **по умолчанию могут обрабатываться только необходимые данные**[122]. То есть те, которые нужны для нормального функционирования устройства.

- Исследователи Всероссийского государственного университета юстиции утверждают, что **проектируемая «клиентоориентированность»**[123] снижает риски вмешательства в частную жизнь пользователя за счет обеспечения конфиденциальности данных и прозрачности их сбора.

- **Утечка пользовательских данных повышает риск негативных последствий для пользователя**. В таком случае персональные данные могут быть использованы для шантажа или мошенничества.

- **Данные, полученные в результате обезличивания, являются персональными**. Такие данные потенциально позволяют идентифицировать человека при наличии дополнительных сведений или с использованием определенных методов анализа.

Рекомендации для разработчиков:

1. **Изучите законодательство, регулирующее защиту персональных данных, частной жизни и тайну связи**. Это поможет проанализировать и (по возможности) предупредить юридические риски.
2. **Получите согласие пользователя на сбор и дальнейшую обработку пользовательских данных**, если оно необходимо в соответствии с требованиями законодательства. Такое согласие должно быть информированным.
3. **Интегрируйте инструменты защиты приватности в функционал продукта**. Например, уведомляйте пользователя о сборе информации и предоставляйте возможность ограничения или запрета сбора данных «в ручном режиме».
4. **Минимизируйте сбор информации, идентифицирующей пользователя**, если это не требуется для нормального использования и работы сервиса.
5. **Предоставьте пользователю возможность получать информацию об используемых данных**. Ему важно знать о том, какие данные вы собираете и как планируете использовать.
6. **Тестируйте различные методы анонимизации и шифрования персональных данных при передаче третьим лицам**. Это поможет предотвратить последствия утечки данных.

Рекомендации для пользователей:

1. **Изучите политику конфиденциальности компании**. В ней должно быть указано, какие данные собираются, как они используются и как пользователь может контролировать свои данные.
2. **Используйте функционал продукта для защиты своих данных**. Если устройству для его нормальной работы очевидно не требуется микрофон или камера, ограничьте в настройках разрешение на сбор данных, включая геолокацию.
3. **Направляйте разработчику информацию об ошибках для устранения недочетов**. Это поможет оптимизировать работу устройства/приложения и повысить удовлетворенность пользователей от продукта.
4. Если служба поддержки не оказала содействия в решении вашего вопроса, обратитесь в вышестоящие инстанции. Например, в специальные комиссии по этике в компаниях, органы судебной власти и иные государственные органы.

Исследования по вопросу:

1. Согласно опросу ВЦИОМ с той или иной периодичностью «умными» девайсами для дома пользуются более четверти россиян (28%)[124]. **Главная угроза видится россиянам в возможности передачи собираемых данных третьим лицам (15%)**. Еще 6% ответили, что сбор и анализ информации о пользователях — вторжение в личную жизнь, нарушение прав и свобод, 5% не исключают возможность слежки/шпионажа через «умные» девайсы. Действия, препятствующие

сбору персональных данных, предпринимают 23% пользователей «умных» устройств. В числе самых частых мер по защите конфиденциальности — заклеивание веб-камеры ноутбука (12%) и отключение девайсов от сети (6%).

2. Голосовые помощники анализируют каждый звук для того, чтобы распознать фразу активации. **Часто похожие звуки могут привести к ложному срабатыванию**. Исследователи из Unacceptable обнаружили[125] более тысячи фраз, которые приводят к активации голосовых помощников Alexa, Google Home, Siri и Microsoft Cortana. Siri реагирует на «city», а «Cortana» — на «Montana». Многие из таких фраз есть в фильмах, сериалах и телешоу, например, в «Игре престолов», «Карточном домике» и в новостях. Более того, Siri можно активировать звуком застежки-молнии или поднятием руки.

Практика:

В 2024 году появилась новость о том, что **в Google произошла массовая утечка данных пользователей** сервиса[126]. Так, функция Google Audio осуществляла ненамеренную запись голосов детей, Google Street View расшифровывал и сохранял номерные знаки автомобилей, а принадлежащий Google сервис Waze раскрывал домашние адреса пользователей.

Что думают эксперты?

Елена Сурагина,
руководитель рабочей группы по созданию свода наилучших практик решения возникающих этических вопросов в жизненном цикле ИИ Комиссии по реализации Кодекса этики в сфере ИИ

— Сам по себе сбор пользовательских данных с помощью умных устройств нельзя назвать неэтичным, если он происходит с согласия пользователя. Однако в этом вопросе важны открытость взаимодействия с пользователем и прозрачность информации: пользователи должны знать о том, что происходит сбор данных, и о том, как эти данные будут использоваться. Такая открытость должна стать основой доверия пользователей и к цифровым сервисам, и к бизнесу в целом.

Андрей Калинин,
генеральный директор MTS AI

— В первую очередь этичность сбора данных пользователей с различных устройств — это вопрос соблюдения применимого законодательства и общепринятых норм. Если от пользователя имеется соглашение и он осведомлен, какие данные будут собираться и где станут использоваться, то это совершенно этично. Но перед этим стоит удостовериться, что цель, для которой собираются данные, соответствует ценностям компании и этическим принципам. Кроме того, необходимо обеспечить защиту данных от потенциальных утечек[127].

13 Этично ли применение ИИ в массовом видеонаблюдении?

Ответ:

Этично использовать ИИ в системах массового видеонаблюдения для обеспечения общественной безопасности при соблюдении прав человека и в случаях и в порядке, установленных законом; в отсутствие законодательной регламентации такое применение ИИ было бы этичным с предварительным предупреждением граждан.

Обоснование:

- Корейские ученые университета Гачон отмечают, **что массовое видеонаблюдение — это система раннего предупреждения и оповещения в случае угрозы**. ИИ из записывающих устройств создает динамические средства общественной безопасности: анализ информации происходит в режиме реального времени[128].
- ОЭСР в своем докладе «ИИ и общество»[129] напоминает, что **законопослушным гражданам ничто не угрожает**. Обработка видеоинформации ИИ фокусирует внимание правоохранительных органов только на угрозах безопасности и правонарушениях.
- Исследователи Городского университета Манчестера утверждают, что **система видеонаблюдения не нарушает права на неприкосновенность частной жизни**[130]. Она проводится в общедоступном месте и в публичных интересах, то есть для обеспечения безопасности и защиты общественного порядка.
- **Использование ИИ в видеонаблюдении повышает чувство защищенности в общественных местах**. Она позволяет оперативно анализировать ситуацию, определять алгоритм решений и вызывать необходимые службы.
- Ученые Международного научно-исследовательского института Манав-Рахн считают, что **ИИ экономит человеческий ресурс и оптимизирует работу правоохранительных органов**. Использование ИИ решает проблему обеспечения безопасности в общественных местах без привлечения большого количества служб, поскольку позволяет выявлять угрозы в удаленном режиме[131].
- Судебная практика подтверждает, что **камеры массового видеонаблюдения ведут видеозапись прохожих в потоковом режиме с расстояния**. По общему правилу, это не является обработкой биометрических данных, если при этом не устанавливаются личности[132].

Исследовательские рекомендации:

1. **Способствуйте прозрачности и открытости**. Предоставляйте гражданам информацию о работе систем видеонаблюдения: цели, механизмы, преимущества использования. Это поможет укрепить доверие общества и обеспечить контроль над возможными злоупотреблениями.
2. **Регулярно оценивайте эффективность и влияние систем видеонаблюдения на общественную безопасность и права граждан**. При необходимости внедряйте корректировки в систему.

Рекомендации для граждан:

1. **Ознакомьтесь с правовыми актами, регулирующими данный вопрос**. Знание законодательной базы и своих прав позволит вам лучше разобраться и объективно оценить эффективность применения массового видеонаблюдения.
2. **Поддерживайте свою осведомленность о новых технологиях и их применении**. Власти могут проводить публичные обсуждения и консультации по вопросу городского видеонаблюдения: принимайте участие в таких мероприятиях.

Практика:

Применение искусственного интеллекта значительно увеличивает как скорость раскрытия преступлений, так и процент разрешенных дел.

1. Полиция США использует ИИ для сопоставления фотографии человека, совершившего правонарушение, с уже имеющимися в базе полиции фотографиями[133].

2. В Великобритании технология распознавания лиц применяется в режиме реального времени. Когда человек проходит под камерой видеонаблюдения, его изображение автоматически сопоставляется с изображениями разыскиваемых преступников[134].

3. Конституционный совет Франции, согласившись на ограниченное использование ИИ на Олимпиаде, заявил, что новые меры могут быть применены только на спортивных, развлекательных или культурных мероприятиях в целях «предотвращения нарушений общественного порядка». Закон будет действовать до марта 2025 года. Франция — первая страна в ЕС, разрешившая использование ИИ для наблюдения[135].

Подходы международных организаций и регуляторов:

- Рекомендации ЮНЕСКО об этических аспектах искусственного интеллекта закрепляют[136] следующий подход: в тех случаях, когда предполагается, что принимаемые решения имеют необратимые последствия, или их трудно обратить вспять, или могут быть связаны с решениями о жизни и смерти, окончательное решение должно приниматься человеком. В частности, системы искусственного интеллекта не должны использоваться для социальной оценки или массового наблюдения.
- Регламент Европейского союза об ИИ[137] придерживается такого же подхода. К запрещенным методам ИИ отнесены размещение на рынке, ввод в эксплуатацию для этой конкретной цели или использование систем искусственного интеллекта, которые создают или расширяют базы данных распознавания лиц посредством нецелевого извлечения изображений лиц из интернета или видеозаписей с камер видеонаблюдения.

Что думают эксперты?

Сергей Собянин,
мэр г. Москвы

— Много было всяких скептических замечаний по поводу видеонаблюдения в городе, всяких инсинуаций, что это плохо, что за кем-то будут следить. Система видеонаблюдения в первую очередь, конечно, работает на безопасность города. В настоящее время благодаря системе в метро и в городе было задержано 7 713 человек, находящихся в федеральном розыске[138].

Владимир Табак,
генеральный директор
АНО «Диалог Регионы»

— Применение искусственного интеллекта в системах видеонаблюдения — большая возможность существенно улучшить работу по охране общественного порядка и предупреждения правонарушений. Но очень актуальны вопросы сохранения конфиденциальности, надлежащего обращения с персональными данными и рисков неправомерного использования таких технологий. Важно обеспечить контроль за тем, чтобы оснащенные искусственным интеллектом системы видеонаблюдения не вторгались в личное пространство людей. Граждане должны знать о проведении видеонаблюдения и, по возможности, давать свое согласие. Стоит учитывать и прозрачность целей использования ИИ, и процессов обработки данных, и процедуры ответственности в тех случаях, когда в работе систем возникают ошибки.

14 Этично ли использовать ИИ для прогнозирования и предотвращения преступлений?

Ответ:

В случаях, предусмотренных законом, и при соблюдении прав и свобод человека применять технологии искусственного интеллекта для прогнозирования и предотвращения преступлений — этично.

Обоснование:

- Интерпол и ЮНИКРИ в совместном докладе «На пути к ответственным инновациям в области ИИ»[139] отмечают, что **ИИ предлагает правоохранительным органам огромные возможности для предотвращения преступлений**. Предиктивная полицейская деятельность позволяет определять самые преступные районы, планировать маршруты патрулирования и эффективно распределять ресурсы.
- **Многократное увеличение финансовых транзакций в цифровой среде, а также передач конфиденциальной информации порождает новые угрозы**. Использование ИИ позволяет снизить вероятность ошибок, связанных с человеческим фактором, таких как невнимательность или недостаточная квалификация.
- **Особенно часто ИИ применяется в рамках противодействия киберпреступности**. Сегодня методы машинного обучения используются в целях мониторинга деятельности информационной системы и человека с целью выявления потенциальных отклонений, прогнозирования вредоносных приложений и сайтов.
- Исследование Национального института правосудия США показывает, что **ИИ делает работу правоохранительных органов более эффективной и менее зависимой от человеческого фактора**. Применение ИИ в обработке личной информации позволяет повысить скорость ее обработки, а также снизить риски человеческой невнимательности[140].
- **Прогнозирование преступлений с помощью ИИ основывается не только на профилировании человека**. Например, согласно AI Act в ЕС таможенным органам разрешено использовать ИИ для прогнозирования вероятности обнаружения наркотиков или контрафакта на основе известных маршрутов незаконного оборота[141].
- Ученые из Мариан Колледжа в Индии предупреждают, что **использование ИИ в предиктивной полицейской деятельности создает угрозу дискриминации**. Данные могут быть ошибочными, неполными и необъективными в силу того, что исторически или из-за региональных особенностей отдельные социальные группы могут быть чаще представлены в качестве преступников[142].

Рекомендации для разработчиков:

1. **Обеспечьте соблюдение основополагающих прав человека**. К таковым относятся право на конфиденциальность, на доступ к информации, недискриминацию и обжалование несправедливых решений.
2. **Предусмотрите план сокращения возможных рисков**. Разработка и обучение ИИ для применения в прогнозировании и предотвращении преступлений должны исключать возможность дискриминации по тому или иному признаку, а также формирование недостоверной информации.
3. **Рекомендуется устанавливать систему контроля человеком**. При использовании ИИ возможны ошибки, поэтому следует проводить проверку принимаемых ИИ решений и учитывать все обстоятельства и доказательства.

Исследование по вопросу:

Действительно, изначально применение технологий ИИ для прогнозирования преступлений было довольно спорным, потому что данные системы не учитывали предубеждений, сложившихся за долгое время работы правоохранительных органов. Например, широко известны случаи ложного прогнозирования рецидива со стороны афроамериканского населения в США.

Тем не менее, согласно последним исследованиям социологов из Чикагского университета, новейшие системы ИИ для предотвращения преступлений **могут предсказывать будущие правонарушения за неделю с точностью около 90%**[143]. Новая модель пресекает преступность, рассматривая временные и пространственные координаты дискретных событий и выявляя закономерности для прогнозирования будущих событий. Она делит город на несколько областей и прогнозирует преступность только в пределах данной территории, а не опирается на традиционные границы районов или политические границы, которые также подвержены изменению.

Практика:

Министерство внутренних дел Российской Федерации планирует внедрить искусственный интеллект в правоохранительную деятельность. В 2024 году ведомство собирается провести научно-исследовательскую работу и подготовить датасеты для обучения и тестирования нейросетевых моделей, а в 2025 году разработать две системы на базе ИИ: «Клон» и «Конъюнктура».

«Клон» позволит выявлять факты подделки видеоизображений, а «Конъюнктура» должна прогнозировать негативные события, чрезвычайные ситуации и моделировать сценарии реагирования на них. Такие мероприятия включены в план по внедрению технологий ИИ в деятельность органов внутренних дел РФ на 2023–2025 годы. План утвержден заместителем министра внутренних дел Виталием Шуликой[144].

Что думают эксперты?

Алексей Минбалеев,
заведующий кафедрой информационного права и цифровых технологий Университета имени О. Е. Кутафина (МГЮА)

— Какие бы сложности ни возникали при использовании ИИ в противодействии преступности, государство вряд ли откажется от его использования в данном направлении. Любая возможность восстановить нарушенные преступлением права и законные интересы человека должна быть реализована. Но при этом важно сохранять контроль за принимаемыми ИИ решениями.

Темирлан Салихов,
специалист в области цифровой криминалистики

— Рациональное применение инструментов на основе искусственного интеллекта предоставило новые возможности и существенно оптимизировало деятельность специалистов в области цифровой криминалистики. Возможность обработки миллиардов данных позволяет принимать критические решения в сжатые сроки и драматически влияет на общественную безопасность. Важно помнить, что итоговое решение остается за компетентным профессионалом.

15 Этично ли применять ИИ для скоринга в ритейле, финансах и других отдельных сферах?

Ответ:

Использование ИИ-скоринга в финансах и ритейле этично при соблюдении ключевых принципов: недискриминация, прозрачность, защита персональных данных и экспертный контроль.

Обоснование:

- Исследователи YABX Technologies (финансовое учреждение в Гааге) утверждают, что **ИИ расширяет возможности персонализации и повышает эффективность предоставления услуг**. Так, алгоритмы машинного обучения могут выявлять закономерности и тенденции, которые люди или традиционные модели оценки могут упускать из виду. Такой адаптивный подход не только повышает точность оценки, но и позволяет вносить коррективы в режиме реального времени, гарантируя, что система оценки останется динамичной и будет реагировать на меняющиеся внешние условия[145].
- В исследовании, проведенном учеными из нескольких университетов США, подчеркивается **важность обеспечения прозрачности при применении скоринга в банкинге и ритейле**. Потребители должны иметь право знать, как работают данные системы, понимать, какие типы информации используются и то, как работают алгоритмы модели ИИ[146].
- Ученые из Американского национального университета выделяют в качестве **одной из важнейших задач кредитного скоринга обеспечение справедливости и недопущение предвзятости**. Модели ИИ могут быть разработаны таким образом, чтобы минимизировать дискриминационные факторы и способствовать справедливости, уделяя особое внимание соответствующему финансовому поведению, а не демографическим характеристикам[147].

Рекомендации для бизнеса:

1. **Разработать отраслевые стандарты для обеспечения этичного использования ИИ-скоринга в бизнесе**. Эти нормы должны закреплять принципы недискриминации, прозрачности и защиты данных.
2. **Рекомендуется создавать механизмы объяснения логики принятых решений**. Это повысит прозрачность систем ИИ-скоринга и общую осведомленность пользователей о принципах их работы.

3. **Следует обеспечить способность ИИ-системы всесторонне учитывать индивидуальные особенности клиентов**. Применение скоринга в банкинге и ритейле не должно вести к ограничению доступа отдельных групп населения к базовым финансовым и потребительским услугам.

4. **Поддерживайте создание комиссий по этике в компании**. Эти комиссии смогут оценивать работу ИИ-систем с точки зрения соблюдения принципов этичности, а также разбирать случаи нарушений и принимать соответствующие меры для защиты интересов пользователей и клиентов.

Исследование по вопросу:

В 2023 году Банк России опубликовал доклад **«Применение искусственного интеллекта на финансовом рынке»**[148].

В докладе говорится, что использование банками ИИ может рассматриваться как возможность дополнительного повышения эффективности и качества оказываемых ими услуг, в том числе за счет снижения издержек, ускорения процессов, ресурсной оптимизации и обработки больших массивов данных.

Источник: Банк России[148]

С помощью «умного» скоринга кредиторы могут анализировать не только финансовую информацию о заемщике, но и «альтернативные» данные — говорится в отчете ЦБ. К таким показателям регулятор отнес:

- сведения из социальных сетей;
- данные платежных систем;
- геолокацию;
- статистику мобильных приложений.

Что думают эксперты?

Андрей Черкашин,
председатель Дальневосточного Сбербанка

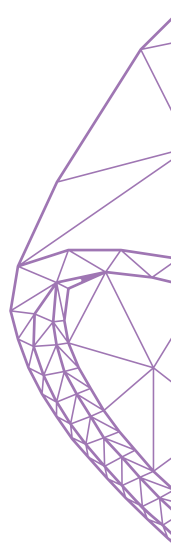
— Самое активное применение искусственный интеллект находит в кредитном скоринге[149], то есть в оценке платежеспособности лица, желающего получить кредит. В этом случае ИИ в считаные секунды обрабатывает большой объем данных, анализирует тысячи параметров и принимает решение. Выдачей кредитов использование ИИ, конечно, не исчерпывается. Мы во многих своих бизнес-процессах применяем модели, созданные с помощью искусственного интеллекта: от поиска недвижимости и проведения сделок до внедрения ИИ в работу чат-бота, который анализирует речь, классифицирует обращения к виртуальным ассистентам, верифицирует сканы документов[150]

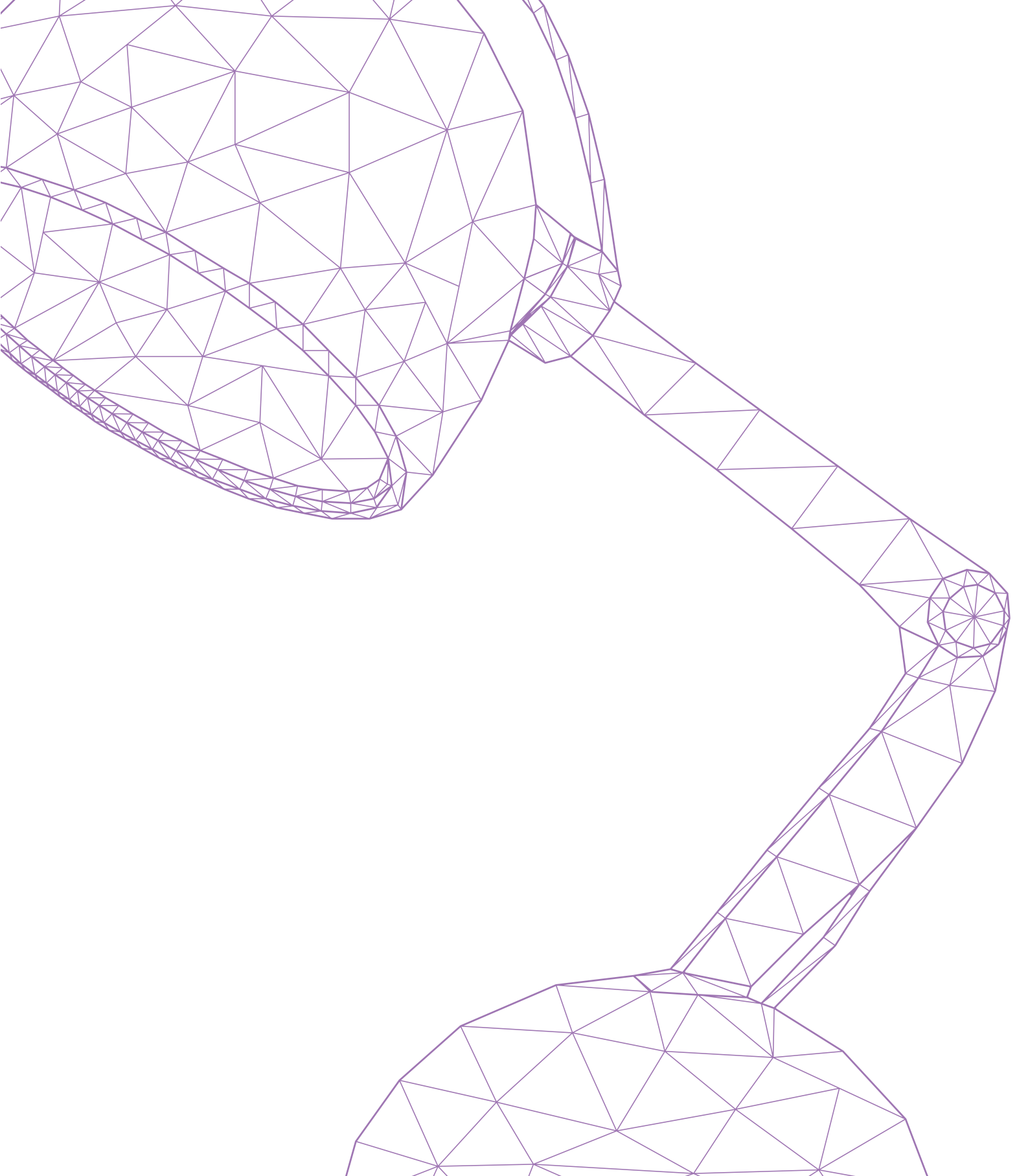
Анна Казакова,
директор по рискам, вице-президент Т-Банка

— Этичность использования ML-скоринга в финансах и ритейле зависит от контекста и целей. Если машинное обучение помогает улучшить обслуживание клиентов и обеспечить справедливое кредитование, то его можно считать этичным. В Т-Банке, например, при расчете кредитного лимита мы используем статистику для оценки платежеспособности клиентов, что помогает защитить их от финансовой неграмотности. В то же время должны быть меры для защиты данных и предотвращения злоупотреблений, чтобы конфиденциальность и права пользователей были приоритетом. Важно, чтобы алгоритмы не усиливали предвзятость и не дискриминировали определенные группы людей[151].

глава 3

ИИ и достоверность





16 Проблема обучения: как избежать обучения ИИ на некорректной информации?

Ответ:

С проблемой недостоверной информации разработчики, поставщики данных и заказчики, внедряющие ИИ, борются различными способами, например путем проверки данных на соответствие законодательству, тестирования модели и ее переобучения.

Обоснование:

- В соответствии с европейской «Белой книгой по ИИ», **в разработке модели ИИ участвуют различные субъекты: каждый в своей части борется с недостоверной информацией по всей цепочке**. Наиболее важными являются датасеты, поскольку они сильно влияют на качество модели[152].
- **Использование разнообразных данных может помочь улучшить точность модели**. На этапе претрейна (первичного обучения) используются разные данные, так как на этапе файнтюна (финальной настройки и обучения) будут использоваться только высококачественные данные с учетом требований законодательства (о персональных данных, коммерческой тайне, интеллектуальной собственности и т.д.).
- По мнению группы российских и австрийских ученых, **поставщик данных отвечает за предоставление качественных данных**. Некачественные данные — это данные, которые не соответствуют требованиям к их формату, полноте, достоверности, релевантности и другим характеристикам, необходимым для корректной работы ИИ[153].
- Согласно исследованию российской юридической фирмы Intellect, **разработчик отбирает необходимые данные для обучения и корректирует результаты, оценивает верность данных при обучении**. Именно он указывает ИИ, что является достоверным, а что — нет. Критерии «достоверности», как правило, проверены на предмет обоснованности и соответствия правовым принципам[154].
- **Данные, используемые для обучения, могут считаться достоверными для одной аудитории, но быть спорными для другой. Например, исторические и религиозные факты могут интерпретироваться по-разному в зависимости от культурного или национального контекста.**

Рекомендации для поставщиков данных:

1. **Следует раскрывать информацию о данных**: например, об их происхождении, методах сбора, и отмечать известные ограничения и искажения.

2. **Обеспечивайте регулярное обновление данных**.
3. **Предупреждайте контрагентов** об изменениях в поставляемых вами наборах данных.

Рекомендации для разработчиков:

1. **Внимательно относитесь к источникам данных** для обучения и с учетом их специфики.
2. **Проводите предварительный анализ предоставленных данных**. Если обнаружены проблемы, разработчикам надо уведомить об этом поставщиков данных.
3. **Тестируйте модели**. Это поможет эффективнее выявлять недостоверные данные.

Рекомендации для заказчиков моделей:

1. **Обеспечивайте проверку данных на соответствие поставленным задачам**. Участвуйте в согласовании набора данных.
2. **Осуществляйте мониторинг работы системы**. Вовремя выявленные искажения информации можно устранить путем дообучения модели.

Исследование по вопросу:

Согласно американскому исследованию «Способы обеспечения качества данных для машинного обучения»[155], под термином **«качественные данные»** подразумеваются очищенные данные, содержащие все атрибуты, от которых зависит обучение модели. В данном исследовании также приведены 4 характеристики качественных данных для обучения ML-моделей:

- **Релевантность** (relevancy) — набор данных должен содержать только те признаки, которые предоставляют модели значимую информацию.
- **Постоянство** (consistency) — схожие примеры должны иметь схожие метки, обеспечивая однородность набора данных.
- **Однородность** (uniformity) — значения всех атрибутов должны быть сравнимыми для всех данных.
- **Полнота** (comprehensiveness) — набор данных должен содержать достаточное количество параметров или признаков, чтобы не осталось неохваченных пограничных случаев.

Что думают эксперты?

Иван Оселедец,
генеральный директор Института AIRI

— Если человек специально хочет вывести модель на «плохое поведение», то ответственность несет он сам. Если же сама модель начинает ерунду нести, то, конечно, возникает вопрос к разработчикам, которые ее недопроверили. Мне кажется, нужно идти в направлении создания экспериментальных правовых режимов, давать право на ошибку. Основная проблема ведь в том, что сейчас ИИ — это серая зона, и никто не хочет начинать его масштабное использование, потому что «А что если?».

Анна Мещерякова,
CEO «Третье мнение»

— Мы используем открытые датасеты на стадии research. Работаем с опубликованными в России данными и взаимодействуем с зарубежными коллегами. Собственная научная деятельность и сотрудничество с медицинскими и техническими вузами в России и за рубежом позволяют получать качественные датасеты для целей research. Но на стадии обучения мы редко используем открытые датасеты — у нас собственные требования к классификаторам, к разметке[156].

Денис Димитров,
управляющий директор
по исследованию данных, Sber AI

— Борьба с недостоверной информацией при обучении моделей искусственного интеллекта — это сложный процесс, который требует работы в нескольких направлениях: использование качественных источников данных, фильтрация и очистка данных, ручная проверка данных, разработка моделей проверки фактов и обработка обратной связи от пользователей. Кроме того, одним из способов борьбы с недостоверными ответами и галлюцинациями моделей является retrieval — дообучение модели с целью использовать внешние базы знаний и данных (например, интернет).

17 Проблема распространения вредоносной или вводящей в заблуждение информации с помощью ИИ: как с этим бороться?

Ответ:

Для предотвращения распространения вредоносной или ложной информации с помощью ИИ нужно, чтобы все, кто связан с созданием и использованием данной технологии — от разработчиков до пользователей, — ответственно относились к своим действиям и учитывали правовые и этические нормы.

Обоснование:

- **Проблема распространения вредоносной и вводящей в заблуждение информации с помощью ИИ носит скорее ценностно-субъективный, а не технологический характер.** Вместе с тем имеет смысл вшивать в саму технологическую разработку возможные меры защиты с целью не допустить использования ИИ не по назначению.
- Группа европейских ученых из MediaFutures утверждает, что **сотрудничество между платформами, правительствами и гражданским обществом способствует эффективной модерации контента**, распространению информации, основанной на фактах, и соблюдению законов[157].
- **Человекоориентированное и гуманистическое функционирование системы ИИ включает в себя ответственную разработку и корректное использование**. Только при наличии эти двух компонентов можно создать безопасный, надежный и этичный ИИ, который будет приносить пользу людям.

Исследовательские рекомендации:

1. **Поощряйте инициативы компаний-разработчиков по проверкам созданных нейросетей и их добровольной сертификации**. Компании-разработчики должны иметь представление о должной работе ИИ и о своей потенциальной ответственности за несоблюдение установленных требований. При этом важно учитывать, что в силу работы генеративного ИИ генерируемая информация может не соответствовать ожиданиям пользователей.
2. **Повышайте осведомленность населения в этом вопросе**. Запуск инициатив по обучению граждан навыкам критического мышления и цифровой грамотности позволит людям распознавать и фильтровать потенциально ложную информацию.

Рекомендации для разработчиков:

1. **Используйте фильтры и барьеры**: например, цензоры. Это позволяет предотвратить создание очевидно токсичного контента.
2. **Используйте бенчмарки (системы оценки) для проверки генеративных нейросетей на корректность генераций**.
3. **Если есть техническая возможность, указывайте источник информации**. Пользователь сам принимает решение: доверять этой информации или нет.
4. **Устраняйте выявленные в процессе использования ошибки модели**, связанные с неточностями данных и фактическими искажениями.

Рекомендации для пользователей:

1. **Используйте ИИ только в соответствии с требованиями законодательства и правилами платформы**. Недобросовестное использование инструментов ИИ (например, для создания оскорбительных дипфейков) считается неэтичным и может повлечь негативные последствия для пользователя (блокировка аккаунта).
2. **Учитывайте, что системы ИИ иногда могут «галлюцинировать» и выдавать недостоверную информацию**. Критически анализируйте информацию и используйте инструменты для проверки подлинности генераций.

Исследование по вопросу:

- В июне 2024 года ЮНЕСКО опубликовала **доклад о рисках использования ГенИИ в контексте памяти о Холокосте**[158]. В материале особое внимание уделяется тому факту, что ИИ может «галлюцинировать» и выдавать вымышленные факты. Например, очень часто чат-боты искажают информацию о количестве жертв конкретного события Холокоста. Также системы не всегда могут правильно оценить искаженную информацию, которая является ложной лишь частично. Например, что во всех нацистских концентрационных лагерях были газовые камеры, или что отравление газом было наихудшим видом массового убийства во время Холокоста.

  В качестве решения ЮНЕСКО предлагает разработчикам использовать широкий спектр данных для обучения, консультироваться с заинтересованными сторонами по чувствительным темам, а также привести свои системы мониторинга и оценки рисков в соответствии с этическими принципами. В свою очередь, пользователи должны понимать ограничения технологии ИИ и самостоятельно проводить проверку контента на достоверность.

Согласно опросу, проведенному ВЦИОМ, за последний год **заметно вырос спрос российского общества на четкое разграничение и маркировку продуктов, созданных с использованием искусственного интеллекта**[159]. С 2023 года. доля сторонников обязательной маркировки результатов работы ИИ выросла с 69% до 73%. При этом доля противников данной меры сократилась с 23% до 17%, в том числе вдвое уменьшилась доля тех, кто категорически против этого (с 14% до 7%). Такая динамика указывает на усиление осознания россиянами важности недопущения распространения вредоносной или вводящей в заблуждение информации с помощью ИИ.

Что думают эксперты?

Даниил Гаврилов,
руководитель лаборатории
исследований ИИ, T-Bank AI Research

— Автоматическое обнаружение фейков и вредоносной информации — сложная для рынка задача, ведь все люди считают по-разному, какой контент можно назвать таковым, а какой — нет. Методы, позволяющие сократить количество небезопасных текстов, существуют, но не гарантируют полную защиту от уязвимостей. Одним из направлений, которые могут помочь решить проблему, является развитие методов интерпретируемости моделей. Они позволяют получить ответ на вопрос «Почему искусственный интеллект предлагает конкретное решение в заданной ситуации?», давая возможность лучше понять внутренние процессы ИИ и предотвратить нежелательные результаты. Эта область начала стремительно развиваться после того, как большие языковые модели стали доступны массовой аудитории. В Т-Банке мы уделяем ей повышенное внимание в научных исследованиях.

Олег Янгаличин,
исполнительный директор
по исследованию данных, Сбер

— Для предотвращения распространения вредоносной или вводящей в заблуждение информации с помощью ИИ мы применяем многоуровневый подход. Он включает как разработку и внедрение алгоритмов для автоматического обнаружения и блокировки дезинформации, так и регулярную валидацию с проверкой моделей на наличие уязвимостей, ведущих к небезопасной генерации. Также важным элементом является система обратной связи для пользователей, которая позволяет оперативно реагировать на потенциальные угрозы.

18 Этично ли алгоритмам предлагать пользователю товары и услуги, которые не соответствуют его обычным предпочтениям?

Ответ:

Если алгоритм работает без предвзятости и предлагает разнообразные товары и услуги всем пользователям, это можно считать этичным.

Обоснование:

- Группа исследователей из Индии и Великобритании считает, что **предложение пользователю товаров и услуг, которые не соответствуют обычному выбору пользователя, помогает избежать формирования информационного пузыря**. Такой подход расширяет кругозор и не ограничивает пользователя его привычными предпочтениями[160].
- В исследовании, проведенном в рамках реализации Кодекса этики в сфере ИИ, **использование данных о пользователе для работы рекомендательных сервисов правомерно**, если разработчик соблюдает законодательство о персональных данных и требования иных нормативных актов[161].
- Исследователи из Ратгерского университета утверждают, что **если компания располагает данными о предпочтениях пользователя, важно учитывать их при формировании предложений**. Игнорирование этих данных может восприниматься как неуважение к пользователю[162].

Рекомендация для разработчиков:

1. **Разрабатывайте прозрачные алгоритмы**. Алгоритмы, по которым ИИ делает предложения, должны быть прозрачными и понятными пользователю.
2. **Учитывайте интересы пользователя**. Предложение может считаться этичным, если в какой-то мере релевантно пользователю. Например, если пользователь интересуется здоровым питанием, предлагать ему сладости и фастфуд может быть неэтично.
3. **Не используйте алгоритмы для рекомендации крайне чувствительных категорий товаров и услуг**: из категории для взрослых (18+), религиозного или ритуального характера; которые могут способствовать разжиганию конфликтов или межнациональной розни.
4. **Обеспечивайте возможность выбора**. Пользователь должен иметь возможность настроить предпочтения и отказаться от получения тех, которые ему не интересны.

5. **Информируйте об использовании технологии и о причинах предложений**. Объяснение причин, почему пользователю предлагается тот или иной товар, повышает доверие. Например, если это акция или новинка, которую стоит попробовать.

6. **Анализируйте обратную связь**. Регулярно собирайте и анализируйте обратную связь от пользователей, чтобы улучшать алгоритмы и предложения.

Исследования по вопросу:

1. ИСИЭЗ НИУ ВШЭ в 2023 году провел опрос населения в возрасте 14 лет и старше, результаты которого показали, что большинство (60%) россиян, выходящих в Сеть хотя бы время от времени, просматривают рекомендации цифровых сервисов регулярно (часто/почти всегда). Наибольший интерес представляют подборки новостей (их просматривают 40% опрошенных пользователей интернета) и развлекательных ресурсов (фильмы и сериалы — 32%, музыка — 29%)[163].

2. Согласно исследованию McKinsey от 2021 года, **71% потребителей ожидают, что компании будут обеспечивать персонализированное взаимодействие**, а 76% разочаровываются, когда этого не происходит[164].

   Более того, персонализация повышает производительность и качество обслуживания клиентов. Компании, которые растут быстрее, получают на 40% больше доходов от персонализации, чем их более медленно растущие коллеги. А компании, которые преуспели в персонализации, получают на 40% больше доходов от этой деятельности, чем средние игроки.

Источник: McKinsey[164]

Что думают эксперты?

Алексей Бырдин,
генеральный директор
Ассоциации «Интернет-видео»

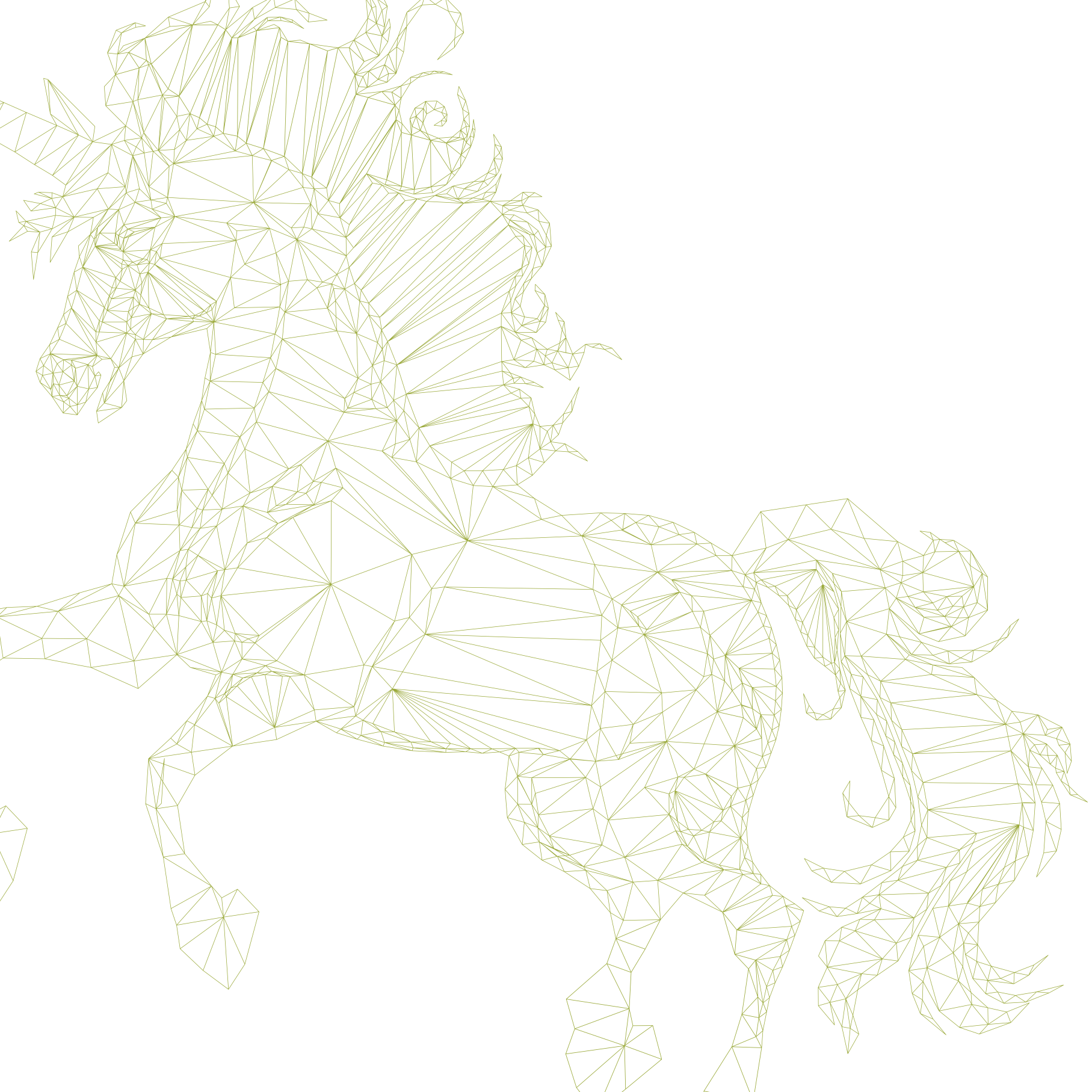
— Гибридные рекомендательные системы, в которых применяется коллаборативная фильтрация, в любом случае периодически «прорывают» «информационный пузырь» пользователя. Это позволяет расширять кругозор пользователя и знакомить его с новыми появившимися продуктами или услугами, что вполне этично. Но во избежание провоцирования недоумения или возмущения пользователей следует избегать попадания в такие рекомендации отдельных (слишком нишевых) категорий объектов, далеко выходящих за круг выявленных предпочтений.

Андрей Зимовнов,
ML-директор, AI VK

— Наши рекомендательные системы ежедневно обрабатывают десятки миллиардов пользовательских сигналов от просмотров и прослушиваний до лайков, шеров и комментариев. Это позволяет нам делать рекомендации в сервисах более точными и релевантными. Вместе с тем пользователям важно и нужно показывать не только тот контент, который они привыкли смотреть или слушать. Это позволяет избежать формирования «дофаминовой петли» и/или «информационного пузыря». Для этого мы разработали механизм Discovery. Он предлагает пользователям не только то, что они уже смотрят, но и новых авторов или даже целые новые тематики.

глава 4

Генеративный ИИ



19 Можно ли доверять информации, полученной при помощи генеративного ИИ и поисковиков на основе ИИ?

Ответ:

И поисковые машины, и генеративный ИИ являются разными, не взаимоисключающими способами получения информации из накопленного объема знаний. Вместе обе технологии помогают пользователю получить актуальный и понятный ответ на свой вопрос: поиск находит релевантные документы, а генеративный ИИ формулирует ответ из найденного. Любую информацию, в том числе полученную с помощью ГенИИ, следует проверять и стремиться находить первоисточники.

Обоснование:

- **Российские ученые подчеркивают, что большинство поисковых систем работают на основе технологий ИИ** (в первую очередь для ранжирования результатов). ИИ позволяет искать страницы не только по конкретным словам, но и по смыслу, персонализировать выдачу, формировать подсказки и так далее[165].
- **И поисковики, и генеративный ИИ могут выдавать ложные ответы**, поскольку учатся на источниках данных из интернета, которые могут содержать ошибки. Важно проводить самостоятельный фактчекинг или советоваться с экспертами относительно релевантности информации.
- Ответы генеративного ИИ могут быть ограничены тем объемом данных, который использовался для обучения моделей. Поэтому если модель разработана, например, на основе текстов до 2023 года, она не сможет вам прокомментировать новости 2024 года, или ее ответы могут быть устаревшими.
- **Генеративный ИИ может галлюцинировать, то есть выдавать несуществующие факты**. Это происходит в ситуации, когда генеративный ИИ для ответа использует собственные знания. Если прямого ответа на вопрос пользователя не было в обучающем датасете, ИИ пытается вывести его на общих закономерностях.
- **Использование систем машинного обучения для классификации и категоризации больших объемов данных позволяет ускорить процесс их обработки и повысить точность результатов поиска**. Алгоритмы ИИ умеют анализировать и интерпретировать тексты, изображения и видео, благодаря чему ручная обработка этих типов данных занимает меньше времени.

Рекомендации для разработчиков:

1. **Правила и принципы работы поисковых систем и сервисов на базе генеративного ИИ важно делать прозрачными и публичными**. Это повысит доверие пользователей к технологии.
2. **Нужно предупреждать пользователей о возможных ошибках и ограничениях этих технологий**, а также информировать их об источниках ответов.
3. **Сервисы на основе ИИ нельзя использовать для навязывания определенной точки зрения или склонения к какому-либо решению**. Поисковые системы и сервисы на базе генеративного ИИ не создают барьеров для получения информации, за исключением противоправного, вредоносного или опасного для жизни и здоровья контента.

Рекомендации для пользователей:

1. **Следует критически относится к информации, полученной из поисковиков и от генеративного ИИ**, перепроверять такую информацию в надежных источниках или первоисточниках.
2. **Следует помнить, что ответственность за распространение недостоверной информации, полученной от чат-бота, и за принятие решений на основе такой информации лежит на пользователе.**
3. **Важно обращать внимание на условия использования систем ИИ**. В них разработчик может предупредить о тех или иных рисках, связанных с работой системы в части обработки информации.

Исследования по вопросу:

1. **Исследователи Воронежского государственного технического университета сравнили результаты работы поисковых систем Yandex и Google с системой ChatGPT, основанной на ИИ**. Анализ показал, что при ответе на запросы пользователей Yandex и Google выдают наиболее релевантные ссылки, содержащие все ключевые слова из запроса. В то же время ChatGPT может сразу предоставить структурированный ответ, включающий все важные аспекты вопроса[166].
   Однако у ChatGPT есть недостаток: система не предоставляет ссылок на источники информации, что затрудняет проверку ее достоверности. Кроме того, иногда ChatGPT дает неверные ответы. Например, на вопрос о названии первого шейного позвонка человека система дала неправильный ответ, в то время как Yandex и Google предоставили краткую и правильную информацию, а также ссылки на ресурсы для проверки данных.
2. Исследование, проведенное учеными из Вашингтонского университета, показало, что **системы, работающие на генеративных нейросетях, могут давать сбои и генерировать абсурдные результаты без каких-либо оснований**[167]. В качестве примера исследователи обратились к ИИ Perplexity и поисковику Arc с просьбой предоставить информацию о несуществующей теории социальных отголосков (Jevin's theory of social echoes). В ответ ИИ предложил концепцию и даже предоставил ссылки на несуществующие источники.

Подход Комиссии по реализации Кодекса этики в сфере ИИ:

В 2024 году представителями бизнес-сообщества была подписана Декларация об ответственном генеративном ИИ. В ней содержится ряд рекомендаций и стандартов поведения в сфере технологий генеративного ИИ для разработчиков, пользователей, представителей академических кругов, а также всех, кто создаёт, внедряет и использует технологии генеративного ИИ. В документе подчеркивается, что генеративный ИИ является лишь одним из инструментов в руках человека.

Практика:

В Google опубликовали правила поиска в отношении контента, созданного ИИ. Компания пишет о том, что отдают предпочтение уникальному контенту высокого качества, соответствующему стандартам E-E-A-T (опыт, компетентность, авторитетность и достоверность). Также компания рассказывает, что около десяти лет назад столкнулась с проблемой быстрого роста объема контента, создаваемого людьми. Тотальная блокировка такого контента была бы неразумной. Поэтому компания решила улучшить свои системы так, чтобы преимущество получал качественный контент. Благодаря системам ранжирования и определению полезного контента пользователи получают материалы, созданные в первую очередь для людей, а не для повышения рейтинга[168].

Что думают эксперты?

Марина Россинская,
операционный директор Яндекс Поиска

— Когда мы предлагаем пользователю ответ поисковой системы, созданный с помощью генеративных нейросетей, мы всегда сообщаем ему об этом. Также важно, чтобы такие ответы всегда содержали ссылки на источники, на базе которых был сформирован ответ. Это позволяет пользователю перейти на сайты и перепроверить информацию или узнать дополнительные факты.

Дарья Чирва,
научный сотрудник Центра сильного искусственного интеллекта в промышленности, преподаватель института международного развития и партнерства Университета ИТМО

— Человек несет ответственность за любое утверждение. Генерация текстов с помощью больших языковых моделей не снимает с него эту ответственность в силу того, что никто и ничто больше пока не может ее нести. Проверка любого факта, ценностного утверждения и других данных, полученных с помощью ИИ-инструментов, — необходимый элемент их грамотного использования.

20 Этично ли синтезировать речь человека при помощи ИИ?

Ответ:

Речь существующего человека важно синтезировать с учетом требований законодательства. В этих пределах этично это делать в отдельных случаях, например для создания произведений искусства, но только при наличии недвусмысленного согласия человека, чей голос будет использоваться в сервисах генерации речи.

Обоснование:

- Крупный разработчик ИИ-решений по созданию видеоконтента Fliki обозначает, **что прозрачность и согласие имеют первостепенное значение в этическом использовании клонирования голоса ИИ**. Создатели должны запрашивать явное согласие при использовании клонированных голосов, особенно в сценариях, где клонированный голос используется в коммерческих или публичных целях. Согласие гарантирует, что люди имеют контроль над использованием своего голоса, и предотвращает несанкционированное или неэтичное клонирование голоса[169].
- В сочетании с устройствами вывода звука ИИ-технологии генерации могут стать незаменимыми помощниками **для тех, кто лишился способности говорить, или для незрячих людей**.
- Синтез речи используется не только при **создании аудиокниг или подкастов, что позволяет «переозвучивать» контент в зависимости от предпочтений слушателя, но и в повседневной жизни: например, в картах и навигаторах при озвучивании маршрута или голосовых помощниках, что также является этичными случаями использования при наличии согласия**.
- Также **этично клонировать голоса при помощи ИИ для синхронного перевода**, что ничем не отличается от использования голосов дублеров, повсеместно применяемого в создании аудио- и видеопроизведений.

Рекомендации для разработчиков:

1. Важно **объяснять обладателю голоса, используемого для обучения модели, особенности технологии синтеза речи и отмечать, что тексты, которые будут озвучены, заранее неизвестны** и точно будут отличаться от тех, которые он озвучивал для обучающего датасета.

2. Старайтесь всегда регламентировать вопрос синтеза голоса договорами.
3. **Использование голосов публичных людей**, содержащихся в общедоступных источниках, возможно в установленных законами рамках (например, для пародий). Если таких ограничений в законе нет, использование голоса не должно унижать честь и достоинство его носителя или использоваться в противоправных целях или в нарушение принятых норм морали.
4. В соглашениях об использовании технологии **целесообразно оставить за собой право отозвать доступ к сервису** для того, чтобы блокировать пользователей, которые создают и/или распространяют противоправный контент.
5. **Следует обеспечивать конфиденциальность персональных данных** и предотвращать их утечку.

Рекомендации для пользователей:

1. При распространении контента, созданного генеративным ИИ с использованием чужого голоса, **важно создавать очевидные примечания о том, что контент сгенерирован с использованием ИИ другим человеком**.
2. При генерации контента пользователю **следует избегать сомнительных промптов**, которые могут нарушить право на честь и достоинство обладателя голоса.
3. При распространении контента с элементами персональных данных другого человека **следует запрашивать соответствующее согласие**.
4. **При применении сервисов по генерации голоса** не используйте их в противоправных целях или в нарушение принятых норм морали.

Исследования по вопросу:

1. По мнению российских ученых из Белгородского юридического института МВД России им. И. Д. Путилина, **одной из основных проблем голосового клонирования с помощью ИИ является возможность злоумышленников использовать эту технологию для мошенничества или дезинформации**[170]. Клонированные голоса могут использоваться для обмана, манипулирования или кражи личных данных, что приведет к серьезным этическим и правовым нарушениям. Например, использование голоса и лица человека может позволить преступникам понизить уровень защищенности населения перед угрозой подмены их фото- и видеоизображений при незаконном получении кредитов, переоформлении недвижимости, дискредитации любого юридического и физического лица.

2. Нейробиологи из Швейцарии установили, что в слуховой коре головного мозга человека есть механизмы, которые позволяют идентифицировать голос, созданный с помощью ИИ. Для эксперимента ученые синтезировали голос по записи реального человека и фиксировали активность головного мозга 25 слушателей при помощи функционального МРТ-сканирования[171].

Практика:

1. По мнению компании-разработчика ИИ-решений для создания видео- и аудиоконтента Fliki, 3 секунд аудио достаточно для создания голосового клона с 85% совпадения с оригиналом[172].

2. В Мэриленде, США, учитель физкультуры сгенерировал голос директора школы и распространил его якобы расистские и антисемитские высказывания среди учителей. Инцидент произошел после профессионального конфликта педагогов[173].

Что думают эксперты?

Александр Крайнов,
директор по развитию технологий искусственного интеллекта «Яндекс»

— Разумеется, условия использования голоса, применения аудиозаписей, процесс передачи голоса и моделей на основе синтеза речи третьим лицам оформляются юридически. Но, как показывает практика, этого недостаточно. Нужно обязательно максимально полно и не только формально информировать диктора обо всех возможных способах использования его голоса. Решение диктора предоставить свой голос для сервиса синтеза речи должно быть полностью осознанным.

Алексей Парфун,
генеральный директор Agenda media group, co-founder Reface Technologies, вице-президент АКАР

— Использование технологий voice clone — это понятный и очень полезный инструмент, который в сочетании с рядом других технологий, таких как lypsink, позволяет всем производителям медиаконтента значительно сократить издержки и повысить скорость производства. Как и многие другие инструменты, в руках злоумышленников он превращается в опасное оружие, поэтому следует использовать маркировку и аппаратный контроль на стороне социальных медиа с целью недопущения мошенничества.

21 Этично ли использовать генеративный ИИ в искусстве и дизайне?

Ответ:

Этично использовать ИИ на любом этапе творческого процесса, но при условии соблюдения этических и правовых норм как разработчиком генеративного ИИ, так и пользователем.

Обоснование:

- **ИИ воплощает замысел человека**, поэтому ответственность за этичное использование ИИ лежит на человеке (носителе смыслов), который ставит задачу ИИ. Вместе с тем следует внимательно относиться к результатам ИИ и помнить о том, что ИИ может галлюцинировать и иметь некоторые технические ошибки.
- Согласно мнению, выраженному в статье Московской школы современного искусства, **ИИ значительно упрощает процесс создания произведений искусства для художников и фотографов. Он помогает в выборе образов, разработке идей, представлении проектов и поиске вдохновения**[174].
- ИИ также **может использоваться для распознавания поддельных произведений с 90%-й точностью**[175], что будет способствовать восстановлению целостности истории искусства.
- Исследователи ХГУ им. Н. Ф. Катанова предлагают рассматривать **ИИ как инструмент для создания искусства, наравне с такими, как компьютер, кисть и другие**[176].
- Исследователи профессионального общества Института инженеров электротехники и электроники (IEEE) считают, что **ИИ демократизирует творчество, делая его доступным для широкого круга лиц,** включая тех, кто не владеет специальными навыками, а также лиц с ограниченными возможностями здоровья[177].

Рекомендации для пользователей:

1. **Ознакомьтесь с пользовательскими соглашениями платформ генеративного ИИ**, чтобы понять, как распределяются права на созданный контент.
2. **Распространителям ИИ-контента, имитирующего реальные события, следует явно указывать на его происхождение**. Для художественных произведений маркировка не критична, если иное не предусмотрено законом.

3. **Следует отсеивать неэтичные или противозаконные генерации и сообщать о них разработчику** (в поддержку), так как, несмотря на ограничения платформ, существует вероятность такого результата.
4. **Следует сохранять информацию о параметрах работы ИИ-системы, используемых в датасетах, и своем вкладе в генерацию контента**. Это может помочь обосновать оригинальность и заявить права на созданные произведения.

Рекомендации для разработчиков:

1. **Следует учитывать нормы авторского права** при обучении генеративной модели или создавать собственные датасеты.
2. **Следует модерировать генерируемый контент и выставлять ограничения** на чувствительные или противоречащие закону темы.
3. **Следует предусмотреть форму обратной связи для пользователя**, чтобы иметь возможность учитывать опасения при модерации генераций.

Практика:

В апреле 2023 года в финале конкурса the World Photography Organization's Sony World Photography Awards премию присудили фотографу Борису Элдагсену за его работу «The Electrician». Однако он не принял награду, объяснив, что фотография была сгенерирована с помощью нейронной сети.

В ответ The World Photography Organization оборвала любые отношения с художником, объявив о нечестности его намерений и признав, что Борис поднимает крайне актуальные вопросы о необходимости дифференцировать и переопределить многие категории и формы искусства.

Исследование по вопросу:

В исследовании Market Research говорится[178] о том, что в 2022 году мировой рынок генеративного ИИ (визуальное искусство, музыка и литература) в искусстве оценивался в 212 млн долларов США и, как ожидается, достигнет 5840 млн долларов к 2032 году, демонстрируя среднегодовой темп роста в 40,5% в течение прогнозируемого периода.

Подход ЮНЕСКО:

ЮНЕСКО рекомендует государствам содействовать образованию в области ИИ и цифрового обучения для художников и творческих специалистов, чтобы оценить пригодность технологий ИИ для использования в их профессии[179]. Также это может способствовать разработке и внедрению подходящих технологий ИИ, поскольку эти технологии используются для создания, производства, распространения, трансляции и потребления различных культурных товаров и услуг. При всем том нужно оставлять в фокусе важность сохранения культурного наследия, разнообразия и свободы творчества.

Что думают эксперты?

Анна Кулик,
директор по маркетингу «Инферит»

— Разумеется, условия использования голоса, применения аудиозаписей, процесс передачи голоса и моделей на основе синтеза речи третьим лицам оформляются юридически. Но, как показывает практика, сравнивать ИИ-искусство с традиционным бессмысленно, как и противопоставлять разные виды и жанры творчества. ИИ — помощник творца, инструмент. Именно творец несет перед обществом ответственность за соблюдение этических норм в процессе работы и отвечает за конечный продукт. Инструменты генеративного ИИ позволяют сегодня каждому человеку, независимо от навыков и знаний, выражать свое уникальное видение мира. Рисовать словом, голосом, мыслью или создавать музыкальные произведения, не зная нотной грамоты, — это подарок человечеству. Но этого недостаточно. Нужно обязательно максимально полно и не только формально информировать диктора обо всех возможных способах использования его голоса. Решение диктора предоставить свой голос для сервиса синтеза речи должно быть полностью осознанным.

Иван Шумейко,
арт-директор «Инферит»

— ИИ в искусстве — это увлекательный инструмент, который может предложить нестандартный взгляд и облегчить некоторые технические задачи. Но душу в произведение всегда вкладывает человек. Только личные переживания, эмоции и внутренний мир творца способны по-настоящему тронуть зрителя, вызвать отклик в его сердце. ИИ может быть виртуозным помощником, но без человеческой искры он никогда не создаст шедевра, который заставит нас плакать или смеяться, сопереживать или мечтать.

22 Этично ли не указывать, что контент сгенерирован с помощью ИИ?

Ответ:

Сейчас создается все больше синтезированного контента, и технологии часто используют для доработки материалов, сделанных людьми. Поэтому однозначного ответа на вопрос нет. Если изображения, текст, аудио или видео, сгенерированные ИИ, могут вводить людей в заблуждение об их происхождении, особенно когда это важно, использовать такой контент без явной маркировки неэтично. Всегда нужно учитывать контекст и цель использования ИИ.

Обоснование:

- По мнению ученого из Оксфорда, важно рассматривать отдельные виды маркировки для разных ситуаций[180].
  - **Видимая маркировка**, явно заметная пользователям (например, текст «Getty Images» на изображениях).
  - **Невидимая маркировка для обозначения технических сигналов**, встроенных в контент.
  - **Оба типа водяных знаков** — описываемые как «прямое» и «косвенное» раскрытие — важны для обеспечения прозрачности.
- Ученые из Массачусетского технологического института (MIT) полагают, что генеративные системы ИИ все больше способны создавать высококачественные медиа. При этом **видимая и невидимая маркировка контента, сгенерированного ИИ, предлагает потенциальную защиту от обмана и смешения оригинального и ИИ-контента**[181].
- Исследователи из Южно-Уральского государственного университета полагают, что **маркировка повысит доверие как к производителям и владельцам систем генеративного ИИ, так и непосредственно к самим сгенерированным работам**[182].
- **Распространение «фейков» — контента, неотличимого от подлинного, — без соответствующей маркировки может восприниматься как манипуляция** и негативно влиять на репутацию.

Рекомендации для разработчиков:

1. Использовать невидимую маркировку, которая не влияет на внешний вид и качество контента и позволяет пользователю его беспрепятственно использовать. Такая маркировка позволит защитить права пользователя и разработчика при выявлении нарушений надзорными органами.

2. **В отдельных сферах можно предоставить возможность использования видимой маркировки** для того, чтобы пользователь сервиса/услуги понимал, что перед ним сгенерированный контент.

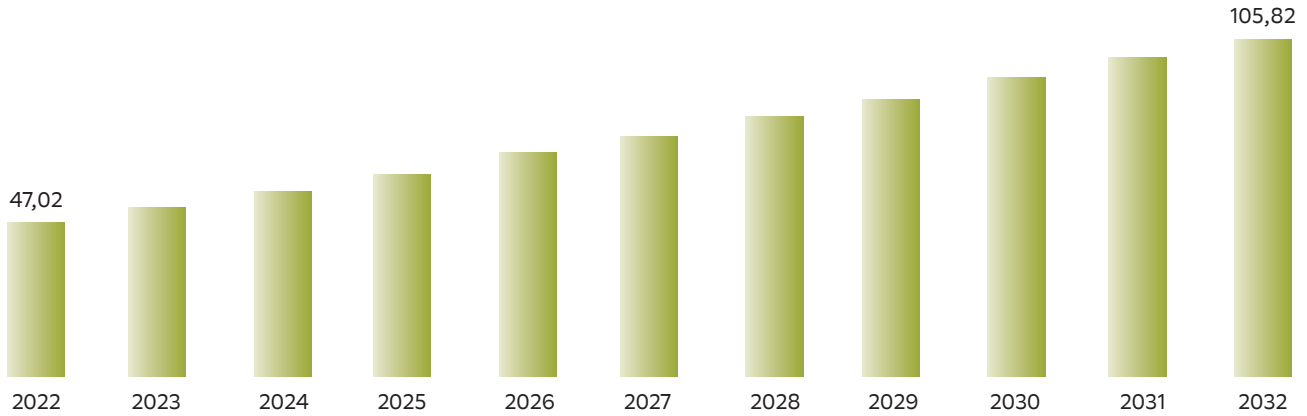
Рекомендации для пользователей:

1. **При распространении сгенерированного контента в зависимости от контекста бывает важно указывать, что он был создан с использованием нейросети, чтобы не вводить в заблуждение других пользователей относительно личного авторства этого контента и не подрывать доверие к своим публикациям.**

2. **Ответственно относитесь к маркировке, установленной разработчиком,** и не пытайтесь ее обойти или скрыть.

Исследования по вопросу:

1. Исходя из исследования Business Research Insights, глобальный объем рынка цифровых технологий водяных знаков составил 47,02 млн долларов США в 2022 году. Рынок достигнет 105,82 млн долларов США к 2032 году на уровне 8,45% CAGR в течение прогнозируемого период[183].



Объем рынка технологий маркировки к 2032 году
Источник: Business Research Insights[183]

Рынок технологий маркировки в последние годы переживает значительный рост, обусловленный растущей потребностью в безопасном и аутентичном цифровом контенте. Опасения по поводу кражи интеллектуальной собственности, подделки и несанкционированного использования контента вызвали спрос на надежные решения для водяных знаков.

2. По мнению ученого Стэнфордского университета, **существует проблема манипулирования маркировкой**. Например, невидимые водяные знаки часто продвигаются как основное решение для маркировки контента, сгенерированного

ИИ. Такими встроенными знаками гораздо легче манипулировать в тексте, чем в аудиовизуальном контенте. Политика в области маркировки ИИ-контента должна быть конкретной в отношении того, для какого вида контента полезны невидимые водяные знаки, потому что решение по раскрытию для изображений необязательно будет полезно для текста.

Практика:

В системах, основанных на изображениях, водяные знаки добавляются как незаметный шум: например, незначительное изменение каждого седьмого пикселя. Это криптографический маркер[184]. Однако текстовые водяные знаки создавать сложнее, поскольку существуют ограниченные способы искажения текста без изменения его значения. Например, в недавнем судебном процессе компания Genius.com подала иск против Google за удаление текстов песен с их веб-сайта. Для доказательства своей правоты Genius заменил некоторые апострофы в текстах песен на своем сайте на фигурные и прямые апострофы. Эта последовательность фигурных и прямых апострофов в азбуке Морзе означает «с поличным». Согласно судебному иску, эта последовательность затем появилась на платформе Google, что свидетельствует о том, что она была скопирована Genius.com[185].

Что думают эксперты?

Анна Абрамова,
директор Центра ИИ, МГИМО

— Стандартизация в вопросах маркировки сгенерированного контента позволит повысить прозрачность работы технологий искусственного интеллекта. Разработка национальных стандартов в данной области станет основой для формулирования предложений для международного сотрудничества.

Сэм Альтман,
Главный исполнительный директор, Open AI

— Следует обратить внимание не только на идею нанесения водяных знаков на генерируемый контент, но и аутентификации несгенерированного контента. Знаменитости или политики должны иметь возможность «криптографически подписывать» сообщения, чтобы доказать, что они действительно их создали. Это вполне вероятная часть будущего для некоторых видов сообщений, и я думаю, нам стоит подробнее рассмотреть данный вопрос[186].

23 Повлияет ли генеративный ИИ на стандарты красоты и на моду?

Ответ:

Риск влияния ИИ на стандарты красоты достаточно велик, так как ИИ-фильтры, использующие алгоритмы машинного обучения, могут создавать нереалистичное представление о внешности пользователя. Они сглаживают кожу, изменяют форму и размеры лица, наносят виртуальный макияж, идеализируя внешние данные человека.

Обоснование:

- Американские ученые в своем исследовании говорят о том, **что из-за погони за идеальной внешностью в социальных сетях возникло новое заболевание, получившее название «социально-сетевая дисморфофобия»**. Это психическое расстройство, при котором человек страдает чрезмерной тревожностью относительно своего образа. ИИ-фильтры и ИИ-модели могут усугубить распространение этого расстройства[187].

Рекомендации для разработчиков:

1. **Старайтесь учитывать уникальность каждого человека: каждый обладает уникальной внешностью и стилем, а шаблонные изображения человека часто это не учитывают.**
2. **Развивайте критическое мышление и используйте анализ изображений** при использовании социальных медиа для распознавания манипулятивных стандартов красоты, созданных ИИ.
3. **Поддерживайте здоровые идеалы красоты, которые основаны на здоровом образе жизни** и самоуважении.

Практика:

Компания Levi’s сгенерировала свою модель с помощью сервиса LaLaLand.ai. Это цифровая студия, которая создает модели с помощью ИИ для компаний в сфере моды. **Существование виртуальных моделей повышает риск утраты рабочих мест для настоящих моделей**. Цифровые модели могут предложить широкий спектр вариантов позирования, имитируя поведение живого человека, делая это без усталости. Кроме того, как подчеркивает основатель LaLaLand.ai Майкл Мусанду, бизнес может сэкономить не только на моделях, но и на визажистах, фотографах и другом персонале, участвующем в съемках[188].

Летом 2024 года прошел **первый в мире конкурс красоты среди несуществующих моделей**. «Miss AI», цифровой эквивалент общеизвестного конкурса «Мисс мира», прошел на онлайн-платформе Fanvue, а организатором выступила компания World AI Creator Awards (WAICA)[189]. На конкурс принимались только сгенерированные изображения — фотографии реальных людей строго отсеивались. В Жюри «Miss AI» вошли Мисс Великобритания, эксперты по маркетингу и создатели популярных ИИ-продуктов, а общий призовой фонд составил 20 тыс. долларов. Победительницей конкурса стала ИИ-блогер марокканка Кенза Лейли.

Исследование по вопросу:

Пластические хирурги Университетской больницы Альбасете в Испании в своей статье подчеркивают, что нужно быть осторожными с тем, что ИИ знает о нас уже сейчас[190]. Важно устранять предубеждения и разночтения для систем ИИ, особенно тех, которые могут закреплять вредные стереотипы или нереалистичные стандарты красоты. Это открывает путь для дальнейших исследований по разработке более инклюзивных и разнообразных моделей ИИ, которые лучше отражают разнообразие и сложность человеческой красоты.

Что думают эксперты?

доктор Керри Макинерни,
научный сотрудник Центра Леверхульма по изучению будущего интеллекта при Кембриджском университете

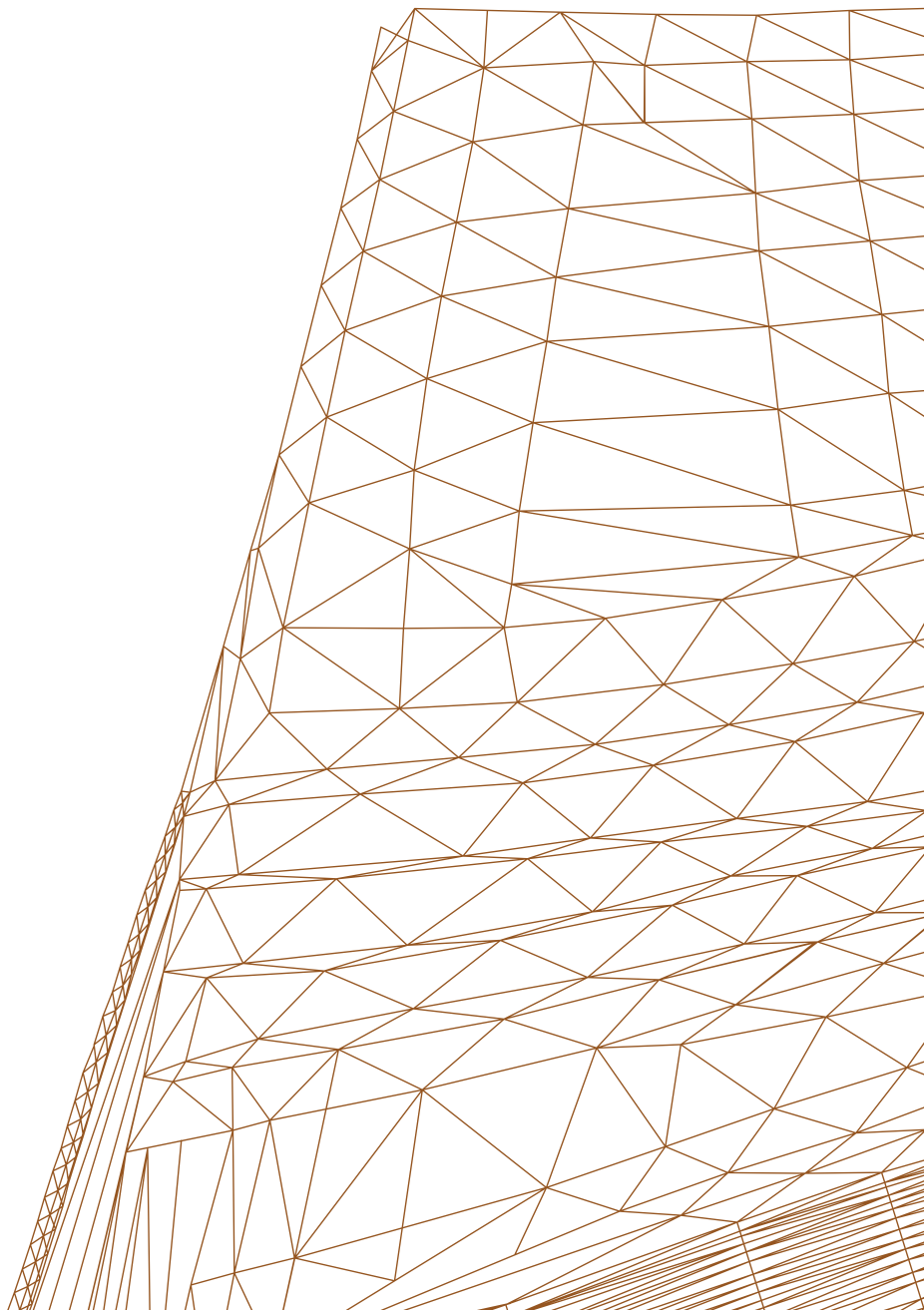
— Большинство моделей, попавших в список претенденток на звание «Мисс ИИ», светлокожие и стройные, то есть не сильно отличающиеся от привычных нам настоящих моделей. Инструменты ИИ созданы для того, чтобы воспроизводить и масштабировать существующие в мире стандарты красоты, а не бросать им вызов или изменять их[191].

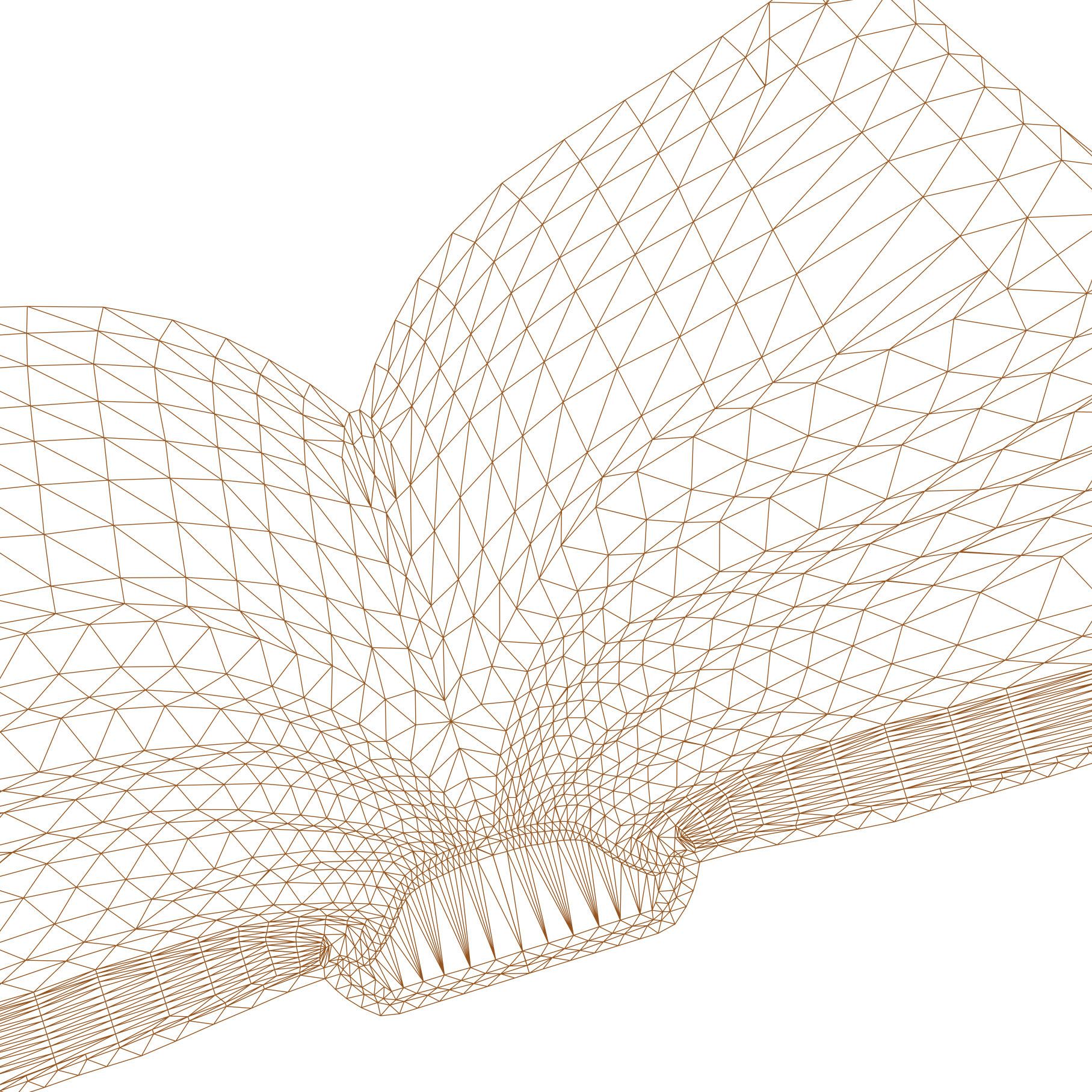
Дженнифер Левин,
американский сертифицированный пластический хирург

— Мы сталкиваемся с сильно отредактированными с помощью ИИ изображениями, которые люди начинают видеть в качестве стандартов красоты. Я думаю, что в будущем как можно больше людей будут критиковать эти изображения, что позволит не использовать такие сильные преобразования[192].

глава 5

ИИ в сфере образования





24 Допустимо ли использование ИИ в образовательном процессе учащимися и преподавателями?

Ответ:

Использование технологий ИИ в образовательном процессе допустимо как преподавателями, так и обучающимися, но с учетом требований законодательства, возрастных ограничений, особенностей конкретной образовательной сферы и внутреннего положения конкретной организации.

Обоснование:

- Согласно совместному исследованию Стаффордширского университета и Технологического института Джорджии, **современные системы ИИ позволяют сократить время на выполнение большинства рутинных действий**, высвобождая его для творческих задач [193].
- **Автоматизация обратной связи и обработки результатов заданий** позволяет исключить человеческий фактор и предвзятость, предоставляя обучающимся своевременные и объективные отзывы и упрощая процедуру оценивания для преподавателя.
- Компании, оказывающие услуги по подготовке к различным экзаменам, отмечают, что **системы адаптивного обучения на основе ИИ** учитывают возможности, интересы, потребности и индивидуальные особенности обучающегося [194].
- Группа американских программистов, разрабатывающих ИИ в сфере образования, считает, что **применение ИИ-технологий способствует развитию гибридных форматов** и упрощает доступ к учебным материалам независимо от времени и местоположения[195].
- Группа исследователей из Индии и ОАЭ напоминает о важности ч**еловеческого контакта в образовательном процессе**. Он способствует развитию социальных связей, творческого и интеллектуального потенциала, поэтому полная его замена алгоритмами ставит под угрозу формирование у обучающихся ответственности и мотивации реализовывать потенциал[196].

Рекомендации для образовательной организации:

1. **Технологии ИИ следует внедрять в образовательный процесс в меру, не забывая про важность социальных связей**. Рекомендуется найти баланс между использованием технологий и сохранением традиционных методов обучения, которые способствуют развитию эмоционального интеллекта и навыков межличностного общения.

2. **Создавайте образовательные программы для преподавателей и обучающихся по корректной работе с инструментами ИИ**. Это поможет преподавателям эффективно интегрировать их в учебный процесс, а обучающиеся смогут научиться использовать эти инструменты для решения различных задач.

3. **Определите четкие образовательные цели использования ИИ**. Например, ИИ может быть использован для подготовки персональных рекомендаций для обучения учащихся или для автоматизации рутинных задач преподавателей.

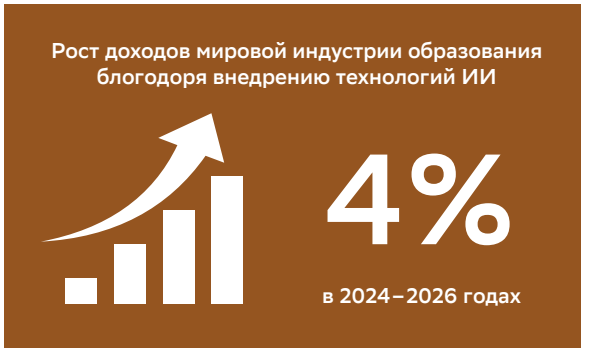
4. **Консультируйтесь с экспертным сообществом и опирайтесь на результаты научных исследований**. Это позволит вам выбрать подходящие инструменты и технологии, а также разработать план обучения и поддержки преподавателей и обучающихся.

5. **Следует соблюдать конфиденциальность**. Важно обеспечить правовую основу для сбора, использования и обработки персональных данных, включая социальные и этические соображения.

Исследования по вопросу:

1. По данным ежегодного исследования McKinsey[197] о состоянии рынка технологий ИИ за 2023 год, отрасли, тесно связанные со знаниями, скорее всего, подвергнутся наибольшим изменениям. Но вместе с **тем такие отрасли могут получить и значительную выгоду**.

2. В 2024 году ЮНЕСКО опубликовала **«Руководство по использованию ГенИИ в образовании и научных исследованиях»**[198]. По мнению ЮНЕСКО, несмотря на растущее внимание к развитию мышления и креативности, важность базовых навыков для психологического развития детей и формирования компетентности у обучающихся не подлежит сомнению. Эти фундаментальные навыки включают аудирование, произношение и письмо на родном или иностранном языке, а также основы счета, рисования и программирования. Подход «упражнение и практика» не следует рассматривать как устаревший педагогический метод. Вместо этого его стоит активно использовать и модернизировать с применением технологий генеративного ИИ. При соблюдении этических и педагогических принципов инструменты генеративного ИИ могут стать индивидуальными тренерами для практики самостоятельного обучения.



Источник: McKinsey[197]

Что думают эксперты?

Эдуард Галажинский,
ректор, Национальный исследовательский Томский государственный университет

— Одним из главных и абсолютно новых вызовов для высшей школы является необходимость как можно быстрее научиться адекватно функционировать в условиях, когда генерирование контента с помощью ИИ-технологий становится доступным для каждого научного сотрудника, преподавателя вуза, студента и абитуриента. Этот вызов наглядно демонстрирует амбивалентную природу любой технологии: сначала, как правило, она представляется всем несомненным благом, делающим жизнь людей легче, проще и приятнее, но весьма скоро начинают проявляться ее негативные черты. Университеты сейчас активно занимаются этой задачей.

Елена Брызгалина,
завкафедрой философии образования философского факультета МГУ имени М. В. Ломоносова, руководитель магистерской программы «Биоэтика»

— В актуальных условиях необходимо обучать студентов эффективному взаимодействию с ИИ как с инструментом для решения профессиональных задач, в том числе учитывая нормы академической этики. При нарастающей интеграции ИИ в образовательный процесс и науку следует ускорить прояснение условий этически допустимого использования ИИ. Документами, фиксирующими этические границы, могут стать локальные акты образовательных и научных институций или методические регламенты для отдельных видов образовательной деятельности (изучение конкретной дисциплины, проведение практики и прочее).

25 Этично ли преподавателю вести предмет через свою цифровую имитацию без личного присутствия в аудитории?

Ответ:

Неэтично, цифровая имитация не может являться полноценной заменой «живого» преподавателя в аудитории. Поэтому использование подобной технологии допустимо только в некоторых случаях и при соблюдении определенных условий.

Обоснование:

- **В качестве эксперимента в аудитории можно использовать цифровую имитацию преподавателя только в определенных ситуациях, например, в условиях онлайн-образования вне аудиторных часов.** Взаимоуважение, понимание социально допустимого и этически корректного поведения, навыки работы со смыслами и ценностями могут формироваться только в межличностном общении.

Рекомендации для преподавателей:

1. **Обсудите достоинства и недостатки использования данного инструмента с обучающимися и экспертами данной сферы**. Так вы сможете предупредить потенциальные риски и максимально их минимизировать для более эффективного достижения поставленной цели использования цифровой имитации в образовательном процессе.
2. **В случае проведения такого эксперимента осуществляйте добросовестное информирование учащихся об их взаимодействии с ИИ, а также объясняйте цели и задачи использования данного инструмента**. Это поможет обучающимся лучше понять, как работает технология, и увидеть ее преимущества для образовательного процесса.
3. При использовании своей цифровой имитации преподаватель должен быть озабочен **сохранением социальных и коммуникативных навыков учащихся** и недопущением обесценивания человеческого общения.

Практика:

В декабре 2023 года **в Европейском союзе был запущен проект по использованию цифровых двойников в высших учебных заведениях**, к которому присоединились 11 университетов из разных стран[199]. Целью этого предложения является расширение возможностей высших учебных заведений в области дополненной реальности (AR) и виртуальной реальности (VR) с использованием цифровых двойников. В рамках проекта будут подготовлены инструкторы, которые в будущем смогут обучать преподавателей эффективной и этичной работе с данным инструментом.

Исследование по вопросу:

1. Исследователи из Университета Гонконга выделяют качества преподавателей, которые невозможно заменить ИИ[200], в том числе цифровой имитацией. Например, преподаватели привносят контекст реального мира, делясь примерами и опытом, что помогает учащимся лучше понимать учебный материал и связывать его с жизненными ситуациями. Они создают пространство для дискуссий, представляя разнообразные точки зрения, задавая сложные вопросы и развивая критическое мышление, чего ИИ пока не способен достичь в полной мере. Кроме того, учителя играют важную роль в разрешении конфликтов, обучая навыкам мирного урегулирования и способствуя развитию социальной ответственности, что делает их незаменимыми в образовательной практике.
2. Исследователи NTT Technical Review утверждают, что **использование цифровых двойников преподавателей должно быть прозрачным для студентов**. Чтобы не ввести студентов в заблуждение, им надо предоставить возможность с самого начала идентифицировать своего «собеседника» как ИИ[201].
3. Ученый из Киотского университета иностранных языков выделяет **преимущества использования цифровой имитации**. Например, данная технология позволяет персонализировать контент, учитывая особенности аудитории учащихся (например, глухонемых, иностранных учащихся, плохо владеющих русским языком)[202].
4. Группа исследователей из Медицинского университета Фуцзянь отмечает, что **применение цифровых копий преподавателей позволяет учащимся в отдаленных районах получать доступ к образовательному контенту** того же качества, что и в тех, где имеется множество образовательных ресурсов[203].

Что думают эксперты?

Сергей Рощин,
проректор по учебной работе НИУ ВШЭ

— Данная технология вполне может быть использована в образовательных целях. Занятие с цифровой имитацией преподавателя подобно занятию с его записанным видео. Это всего лишь аватар. Однако учащиеся должны быть предупреждены, что это — имитация преподавателя.

Вадим Перов,
заведующий кафедрой этики Санкт-Петербургского государственного университета

Образование — это взаимный процесс. Поэтому, во-первых, с этической точки зрения недостаточно проинформировать обучающихся о «цифровой имитации» преподавателя, а необходимо получить их согласие. Во-вторых, если признать, что «цифровой преподаватель» — это этично, то возникает вопрос об этичности присутствия в аудитории «цифровых двойников» обучающихся[204].

26 Этично ли писать курсовые работы или иные учебные работы с помощью ИИ?

Ответ:

Да, если это не противоречит принципу академической честности, а также локальным нормативным актам образовательной или научной организации. Технологии ИИ этично использовать в качестве инструмента, позволяющего обрабатывать информацию и редактировать тексты научных работ, но не для подмены авторства. Ответственность за достоверность использованных данных и конечный результат всегда несет человек.

Обоснование:

- В своем докладе «Эра ИИ в высшем образовании» ЮНЕСКО выделяет **возможность ИИ просматривать большое количество литературы**, чтобы быстро найти наиболее релевантные и актуальные исследования[205]. Подобные системы используют информацию из интернета, которая может быть ненадежной и требующей перепроверки.
- Ученые Университета ОАЭ считают[206], что **ИИ может стать полезным инструментом для преподавателей при оценке студенческих работ**. Он позволяет получить альтернативное мнение и быстро определить, в каких областях студентам требуется дополнительное внимание.
- Согласно Руководству ЮНЕСКО по использованию ГенИИ в образовании и научных исследованиях, **ИИ лучше всего использовать для автоматизированного сбора информации и подготовки структуры научного исследования**. В числе возможных рисков выделяется возможность создания ложной информации, например предоставление несуществующих исследовательских публикаций[207].
- **ИИ может быть полезен в оформлении работ** по установленному проверяющей стороной стандарту и при проверке текста на наличие заимствований, ошибок и несоответствий стилистическим нормам языка.

Рекомендации

Для образовательной организации:

1. **Разработайте четкие правила для студентов и преподавателей по использованию ИИ в процессе обучения**. Это поможет избежать возможных проблем с плагиатом и обеспечить соблюдение этических норм.
2. **Поощряйте студентов самостоятельно анализировать информацию и формулировать собственные выводы**. Это будет способствовать развитию критического мышления и навыков работы с данными.

3. **Предоставляйте студентам доступ к качественным источникам информации и ресурсам**. Они помогут им самостоятельно проводить исследования и писать научные работы.

Для обучающихся:

1. **Соблюдайте правила, установленные образовательной организацией**. В некоторых организациях локальные нормативные акты запрещают использование ИИ при подготовке работ.
2. **Критически анализируйте и проверяйте фактическое основание ответов, предлагаемых нейросетью**. Нейросети могут галлюцинировать и ошибаться, поэтому дополнительная проверка позволит выявить и устранить возможные неточности.
3. **Используйте материалы, созданные с помощью ИИ, только для подкрепления своей научной позиции**. Они не должны заменять собой основные аргументы.

Практика:

1. В 2023 году московский студент успешно защитил дипломную работу, написанную за 23 часа при помощи нейросети (ChatGPT). ChatGPT студент использовал для составления плана работы, написания введения и теоретической части. Однако целых 8 часов студент потратил на редактирование текста и написание практической части дипломной работы[208].

   Таким образом, нейросети могут автоматизировать процесс поиска источников и информации или проверить текст на орфографические ошибки. Однако стоит помнить, что ИИ не является полноценной заменой мыслительному процессу. Нейросети могут создать черновик научной работы, но творческую часть все равно придется выполнять человеку.

2. Многие университеты уже принимают положения об использовании ИИ при написании научных работ. Так, например, Высшая школа экономики (НИУ ВШЭ) в мае 2024 года приняла «Регламент проверки письменных учебных работ на наличие плагиата и использования генеративных моделей»[209]. Согласно Разделу 3 данного Регламента, отсутствие упоминания об использовании генеративных моделей рассматривается как нарушение академических норм.

   Иного подхода придерживается Московский городской педагогический университет (МГПУ). В августе 2023 года на заседании Ученого совета МГПУ было принято решение о легализации для студентов использования технологий ИИ при подготовке выпускных квалификационных работ. Это означает, что студенты могут использовать чат-боты и другие инструменты ИИ для получения данных и текстов при работе над ВКР[210].

Что думают эксперты?

Елена Брызгалина,
завкафедрой философии образования философского факультета
МГУ имени М. В. Ломоносова, руководитель магистерской
программы «Биоэтика»

— Этические аспекты поведения в научно-образовательном пространстве включают вопросы соблюдения авторской этики. Присвоение себе результатов полученной человеком обратной связи от ИИ-инструмента получило название «ИИ-плагиат». Предоставление под своим авторством текстов учебных и научных работ, сгенерированных ИИ-инструментами, без указания на использование таких инструментов можно квалифицировать как академическое мошенничество. Автор работы должен нести ответственность за нарушение авторской этики в учебных и научных ситуациях.

Иван Карлов,
руководитель Лаборатории цифровой трансформации образования Института образования НИУ ВШЭ

— Важно понимать, для каких целей используется искусственный интеллект. Использовать можно по-разному. Можно его попросить написать курсовую работу, а можно, как сейчас делают даже специалисты, дать тезис ИИ и попросить его написать литературным или научным языком. То есть когда у тебя на самом деле уже вся работа есть и ты используешь этот инструмент для подготовки текста этой работы. В любой работе должна быть часть исследования, которая подразумевает, что человек что-то делает своими руками, и в любом случае искусственный интеллект за него это не сделает[211].

27 Можно ли с помощью ИИ проверять работы учащихся?

Ответ:

Да, ИИ может эффективно использоваться как вспомогательный инструмент для преподавателей при проверке ученических работ, автоматизируя рутинные задачи. Однако окончательное решение всегда должно приниматься преподавателем.

Обоснование:

- ЮНЕСКО в своем докладе про использование ИИ в образовании отмечает, что **ИИ облегчает нагрузку на преподавателей**. Если позволить ему выполнять рутинные задачи оценивания, то у преподавателя появится больше времени на проверку творческой части заданий, а также на персонализированную обратную связь[212].
- The Princeton Review — международная компания, оказывающая услуги по подготовке к вступительным экзаменам — утверждает[213], что **проверка стандартизированных заданий с помощью ИИ позволяет осуществлять их оценку объективно, без предвзятости и «личностного фактора» в процессе оценивания**, следуя заранее определенным алгоритмам и критериям.
- Оксфордский университет разрешает своим студентам при самопроверке **использовать ИИ-инструменты для более эффективного использования учебного времени**. ИИ способен учитывать контекст и критерии оценивания текстов, работать с разными форматами данных, предоставлять персональные рекомендации[214].
- Группа исследователей из Индии считает, что **механизмы ИИ, встроенные в системы «антиплагиат», способствуют соблюдению принципа академической честности**. Они могут анализировать лингвистические паттерны, синтаксис и семантические структуры, чтобы выявить случаи, когда учащиеся пытались скрыть плагиат, изменив формулировку и структуру исходного текста[215].
- Согласно исследованию, опубликованному в Data Science Central, **автоматизация проверки позволяет получать результаты в несколько раз быстрее**. Преподаватель, в свою очередь, может позднее обсудить их с учащимся в режиме реального времени[216].
- По мнению ученых из Лондонского университета, **ИИ позволяет определить пробелы в знаниях обучающихся**. Например, помимо определения того, дал ученик правильный ответ или нет, ИИ может проанализировать работу, чтобы помочь преподавателям понять, как именно учащийся пришел к своему ответу[217].

Рекомендации для преподавателей:

1. **Обеспечивайте конфиденциальность данных учащихся**. Необходимо обеспечить безопасность и защиту личной информации обучающихся, а также соблюдать законодательство о защите персональных данных и этические принципы.
2. **Информируйте студентов об участии ИИ в проверке работ учащихся**. Это повысит их доверие как к системам ИИ, так и к самой образовательной организации.
3. **Соблюдайте принципы академической честности, открытости и уважения личности**. Например, важно объяснить ученикам критерии оценки, чтобы они понимали, на чем основаны полученные результаты.
4. **Учитывайте, что ИИ выступает вашим инструментом и ассистентом, а не экспертом**. Преподаватель самостоятельно несет ответственность за результаты оценки.
5. **Используйте национальные лингвистические модели, чтобы не было ошибок**.

Практика:

1. В Китае педагоги уже активно проверяют контрольные работы учеников с помощью алгоритмов искусственного интеллекта[218]. Нейросеть ZipGrande, ключевая задача которой состоит в быстрой проверке работ школьников, уже насчитывает 800 тыс. пользователей. Программа работает следующим образом: пользователь наводит смартфон с включенной камерой на бумажные записи, после чего ИИ всего за несколько секунд проверяет работу на предмет ошибок и предоставляет полученный результат. Как показал опрос, 60% преподавателей считают, что больше всего времени как раз отнимает проверка контрольных, а потому такая система позволяет значительно облегчить работу.

2. Правительство Российской Федерации также намерено до 2030 года привлечь системы ИИ к проверке домашних заданий в школах и планированию образовательных программ[219].

   В июне 2024 года Денис Грибов, замминистра просвещения РФ, выступая на II Международном форуме министров образования «Формируя будущее», заявил, что для этого уже создаются специальные цифровые помощники. Грибов отметил, что также этот проект позволит решить задачу в части снижения бюрократической нагрузки на учителей[220].

Что думают эксперты?

Сергей Рощин,
проректор по учебной работе НИУ ВШЭ

— Конечно может. Но надо перед этим удостовериться, что ИИ ошибается не чаще «живого» преподавателя. Это как тренажер, только по оцениванию результата.

Дмитрий Зубцов,
руководитель академии технологий, данных и кибербезопасности, СберУниверситет

— Если работа показывает знания обучающегося в конкретном вопросе, например его знание языка или понимание определенных терминов, то это достаточно просто можно переложить на ИИ, возможно, в режиме системы помощи принятия решения для преподавателя (подсвечивать неверные ответы). Если же работа креативная или содержащая в себе анализ и выводы, с которыми не всегда ИИ может справиться, то количество ошибок будет слишком значительным и такую задачу поручать ИИ неправильно.

Сергей Валюгин,
учитель литературы ОАНО школа «Ника», преподаватель кафедры мировой литературы Института русского языка имени А. С. Пушкина, победитель конкурса «Учитель года Москва — 2023»

— Особенно удобно использовать ИИ при проверке письменных работ на соблюдение орфографических и пунктуационных норм (диктанты, изложения, сочинения). Но важно помнить, что при наличии морфологических омонимов (различение союзов и вводных слов, наречий и существительных) ИИ не всегда корректно учитывает контекст и требуется дополнительная проверка преподавателем.

28 Этично ли использовать системы прокторинга на основе ИИ?

Ответ:

Использование систем прокторинга на постоянной основе видится неэтичным и нецелесообразным. В то же время точечное применение прокторинга при проведении ключевых контрольных мероприятий может быть этически приемлемо при условии четкого регулирования и соблюдения принципа недискриминации.

Обоснование:

Прокторинг — это процедура контроля за ходом дистанционного испытания (в английском языке proctor — это наблюдатель на экзаменах в вузе)[221].

- **Постоянный прокторинг несет риски нарушения фундаментальных прав**. Он считается серьезным вмешательством в личную жизнь учащихся. Постоянный прокторинг превращает образовательную среду в подобие «паноптикума»[222].
- Согласно исследованию, опубликованному в Журнале информационных технологий, **использование прокторинга может привести к усилению социального неравенства**: например, в отношении людей с низким доходом, которые могут быть не в состоянии позволить себе подходящее техническое оборудование[223].
- Ученые из Мельбурнского университета считают, что **постоянный прокторинг способствует растущему недоверию к важным социальным институтам**. Тотальный контроль подрывает доверие между учащимися и педагогами, нарушает психологически комфортную атмосферу, необходимую для эффективного обучения[224].
- **Ограниченное применение прокторинга** для ключевых контрольных мероприятий оправдано необходимостью обеспечить равные условия для всех студентов и объективность оценивания. Однако, по мнению ЮНЕСКО, такие решения должны быть прозрачны для студентов и подлежать апелляции в спорных случаях[225]

- **Жесткий контроль демотивирует**, не способствует развитию у студентов самостоятельности, ответственности и добросовестности как устойчивых личностных качеств.
- ОЭСР в своем докладе «Онлайн-экзамены в высшем образовании во время COVID-19» выделяет[226] **недостатки онлайн-прокторинга**. Например, он повышает беспокойство студентов при сдаче экзамена, поскольку они боятся быть наказанными из-за технических сбоев.

Рекомендации для образовательных организаций:

1. **Рекомендуется искать разумный баланс между необходимостью обеспечить академическую честность и императивами уважения автономии и приватности учащихся**. Алгоритмы прокторинга должны быть открытыми для аудита, настроены на минимизацию ошибок и предвзятости, а их решения — подлежать апелляции и пересмотру человеком в спорных случаях.
2. **Избыточный контроль через прокторинг недопустим**. Такой подход не может подменять усилий по созданию атмосферы доверия и культивированию норм добросовестности в академической среде.
3. **Соблюдайте этические нормы и законодательство о защите персональных данных**. Использование систем прокторинга требует ясной регламентации, гарантий конфиденциальности данных и защиты от дискриминации в соответствии с нормами информационной этики.
4. **Приоритизируйте воспитание у учащихся осознанной академической честности**. В достижении этой цели могут помочь этические кодексы, тренинги, вовлекающие формы обучения.

Исследование по вопросу:

Источник: НКО Educase

В апреле 2020 года НКО Educause провело исследование с целью выявить основные вызовы дистанционного образования и потенциальные способы решения возникающих проблем. Согласно проведенному опросу, главными проблемами, с которыми сталкиваются образовательные учреждения во время внедрения прокторинговых систем, являются **стоимость онлайн-контроля** (58% респондентов), а также **обеспечение конфиденциальности данных учащихся** (51%).

Что думают эксперты?

Фарида Майленова,
ведущий научный сотрудник сектора гуманитарных экспертиз и биоэтики ИФ РАН

— Прокторинг при выполнении контрольных заданий, онлайн-тестов необходим, так как повышает степень справедливости и объективности оценки результатов; и в целом способствует повышению ответственности студентов. Важный этический момент — студенты должны быть осведомлены об этом. В применении его на постоянной основе нет особой нужды, так как специфика онлайн-обучения позволяет прослушивать лекции и в записи, и за результат обучения ответственны сами студенты[227].

Дмитрий Истомин,
директор компании «Экзамус»

— Любая система несовершенна. Недостаток прокторинга уже в том, что его приходится придумывать. Люди почему-то не пишут честно экзамены.
Система прокторинга приносит в образование в первую очередь равенство, про которое так часто говорят. Можно бесконечно улучшать систему, точность алгоритмов. Если вы посмотрите на другие отрасли, где применяется автоматика, распознавание в автомобилях, беспилотники, — то это бесконечный процесс совершенствования[228].

29 Допустимо ли снижать оценку учащегося при подозрении на использование ИИ в учебе?

Ответ:

Снижение оценок только из-за обнаружения признаков применения ИИ кажется слишком категоричным и недостаточно обоснованным подходом. Тем не менее совсем игнорировать факты использования ИИ также неправильно.

Обоснование:

- Массачусетский технологический университет отмечает, что **санкции без учета реального вклада студента** могут вести к возникновению страха и подавлению творческой инициативы у студентов[229].
- Ученые Стэнфордского университета считают, что **ложные срабатывания чаще всего связаны с лексическими особенностями текста**[230]. Подобные системы часто идентифицируют тексты не носителей языка как сгенерированные, так как обычно у носителей больший словарный запас и лучшее понимание грамматики. Не носители пишут, используя наиболее распространенные словосочетания. Так же поступает и генеративный ИИ: фактически он имитирует человеческое письмо на основе данных, которые когда-либо обрабатывал.
- **Снижение оценки за использование ИИ без четких доказательств нечестности** подрывает доверительные отношения между учащимся и педагогом. Эффективность обучения, в свою очередь, напрямую зависит от этого фактора.
- **Алгоритмы выявления использования ИИ могут давать ложные срабатывания**. Полагаться исключительно на результаты данных систем не стоит — это может привести к наказанию невиновных и вознаграждению нарушителей.
- **Фокус преподавателей на выявлении ИИ отвлекает от содержательной обратной связи, обсуждения сути работ**. Это может негативно влиять на развитие навыков ведения дискуссий и аргументации учащихся.

Рекомендации для образовательных организаций:

1. **Выработайте дифференцированный подход возможного использования ИИ учащимися**. Такой подход должен быть основан на открытом диалоге со всеми заинтересованными сторонами.

2. **Сформулируйте и закрепите четкие и прозрачные правила применения ИИ**. Отразите в этих правилах критерии оценки работ, созданных с помощью инструментов генеративного ИИ.
3. **Принимайте превентивные меры**. Такие как: разъяснение студентам этических принципов работы с ИИ, обучение ответственному взаимодействию с этой технологией, культивирование ценностей академической добросовестности.
4. **Снижение оценок возможно как крайняя мера при доказанных злоупотреблениях**. Однако снижение оценок не должно быть автоматической реакцией на факт использования ИИ.
5. **Систему оценивания следует настраивать на поощрение этичного применения ИИ для решения поставленных задач**. Адаптация потребует экспериментов, постоянной обратной связи и готовности гибко корректировать подходык оцениванию.

Исследование по вопросу:

В июле 2023 года Стэнфордский университет провел исследование: ученые оценили несколько общедоступных систем, заявленных как распознаватели сгенерированного текста, на образцах, написанных носителями и не носителями английского языка.

В результате **89 из 91 (97,8%) эссе от не носителей языка были помечены как сгенерированные ИИ по крайней мере одним из семи различных инструментов**.

Чтобы проверить гипотезу о том, что ограниченный словарный запас способствует предвзятости, ученые использовали ChatGPT для «обогащения» языка, стремясь подражать использованию словарного запаса носителей. Примечательно, что это вмешательство привело к существенному сокращению ошибочной классификации: средний уровень ложноположительных результатов снизился на 49,45% (с 61,22% до 11,77%)[231].

Практика:

OpenAI в июле 2023 года закрыла свой собственный детектор ИИ, обнаружив, что у него «низкий уровень точности». В сообщении на веб-сайте компании говорится, что **«ни одна» из систем распознавания сгенерированного контента, включая их собственную, «не доказала, что может надежно отличить контент, созданный ИИ, от контента, созданного человеком»**.

OpenAI отметила, что существующие системы действительно предвзяты к студентам, которые изучают английский как второй язык, а также к студентам, чьи тексты отличаются особой шаблонностью или лаконичностью.

Более того, по мнению компании, обойти системы распознавания ИИ очень просто: стоит добавить несколько распространенных предложений[232].

Что думают эксперты?

Юрий Чехович,
исполнительный директор
АО «Антиплагиат»

— Когда система «Антиплагиат» обнаруживает, что в тексте много признаков того, что его написала нейросеть, она выделяет этот фрагмент текста как подозрительный. Однако сделать финальный вывод о том, что этот текст был написан нейросетью, пока нельзя. Наша система выступает лишь инструментом, подсвечивающим подозрительные фрагменты работы, а дальше — дело за человеком.

Александр Гасников,
ректор Университета «Иннополис»

— Вопрос об использовании ИИ при выполнении учебной работы не так прост, как может показаться на первый взгляд. Первая реакция, что надо это все запретить и не разрешать пользоваться, иначе так ничему не научиться... С другой стороны, умение правильно воспользоваться ИИ для решения той или иной задачи, в том числе на этапе обучения, может стать элементом обучения и принести отложенную пользу. Выходом из данной ситуации может быть разделение задач (заданий) на те, в которых разрешается (и даже рекомендуется) воспользоваться всеми доступными средствами, в том числе на базе ИИ, и на те, в которых запрещается... Нарушение тут можно, наверное, отождествлять со списыванием.

Ольга Французова,
к.и.н., МГУ им. М.В. Ломоносова,
кафедра философии образования

— В последнее время преподаватели сталкиваются с ситуациями, демонстрирующими несамостоятельность работ школьников и студентов. Слабость инструментов в обнаружении отклонений от правил и широкая доступность технологий культивируют недоброкачественность результатов. Конечно, не следует наказывать за «механизацию» навыков по поиску литературы, оформлению, постановке проблемы, выдвижению задач и гипотез. А вот с содержанием работы следует быть осторожнее. Нужно возлагать ответственность на нарушителей, но вместе с тем воспитывать в них академическую этику написания таких работ.

30 Этично ли ограничивать использование ИИ детьми в образовательных целях вне соответствующих учреждений?

Ответ:

Полностью ограничивать использование ИИ детьми в образовательных целях вне соответствующих учреждений не стоит. Однако такое использование должно происходить под контролем взрослых с учетом возрастных ограничений и уровня развития детей.

Обоснование:

- Министерство образования Великобритании считает, что **ИИ может быть хорошим инструментом для обучения и развития детей**. Учащиеся могут ознакомиться с учебными материалами за пределами класса, а затем приходить на урок с базовыми знаниями для участия в более интерактивных мероприятиях[233].
- **ИИ может выдавать нежелательный или недостоверный контент**. Разумное ограничение использования ИИ родителями и фильтрация контента разработчиками снижают риск поглощения ими такой информации, защищают от негативного воздействия и обеспечивают их психическое здоровье.
- ЮНИСЕФ в своем докладе «ИИ и права детей» подчеркивает, что **технологии ИИ могут использоваться в качестве помощника** в процессе выполнения домашних заданий, развития дополнительных навыков (например, творческих), в том числе для детей с ограниченными возможностями[234].
- Согласно исследованию, опубликованному High Speed Training, **специализированные системы ИИ для детей различных возрастов могут помочь разобраться в тех дисциплинах и жизненных темах, которые не преподают в образовательных учреждениях**. Например, программы могут дать представление о психологических концепциях и теориях, помогая детям развить понимание человеческого поведения и эмоций[235].

Рекомендации для разработчиков:

1. **При разработке ИИ-решений для дополнительного образования детей учитывайте возрастные ограничения**. Разработчикам следует разграничивать контент, который будет соответствовать уровню развития и психической устойчивости детей.

2. **Интегрируйте подобный функционал в свои сервисы.** Разработчикам сервисов на основе генеративных моделей следует установить возможность родительского контроля и возрастные ограничения на просмотр генерируемого контента.

Рекомендации для родителей:

1. **Контролируйте использование ИИ в образовательных целях у детей школьного возраста.** Они могут злоупотреблять технологиями в целях выполнения домашних заданий и не получать необходимое количество знаний и самостоятельных навыков.

2. **Ответственно относитесь к выбору сервисов на основе ИИ для своих детей**. Выбирайте тех разработчиков, которые максимально открыто предоставляют информацию о своих алгоритмах и ценностях, а также специально ориентируются на детское образование.

Исследование по вопросу:

В феврале 2024 года были опубликованы результаты опроса, проведенного Hart Research. Опрос был посвящен использованию искусственного интеллекта среди подростков. **58% респондентов заявили, что ИИ помогает им улучшить свою успеваемость в школе**, а также способствует заинтересованности в дополнительном обучении вне образовательных организаций. Молодые люди особенно склонны использовать генеративный ИИ: 60% респондентов признались, что используют инструменты ГенИИ на постоянной основе. При этом 63% опрошенных детей в возрасте от 9 до 17 лет используют данные инструменты именно в образовательных целях, в том числе для выполнения домашних заданий[236].

Для каких целей Вы используете инструменты ГенИИ?
Источник: Hart Research[236]

Что думают эксперты?

Илья Померанцев,
руководитель направления ИИ
компании GLOBUS IT

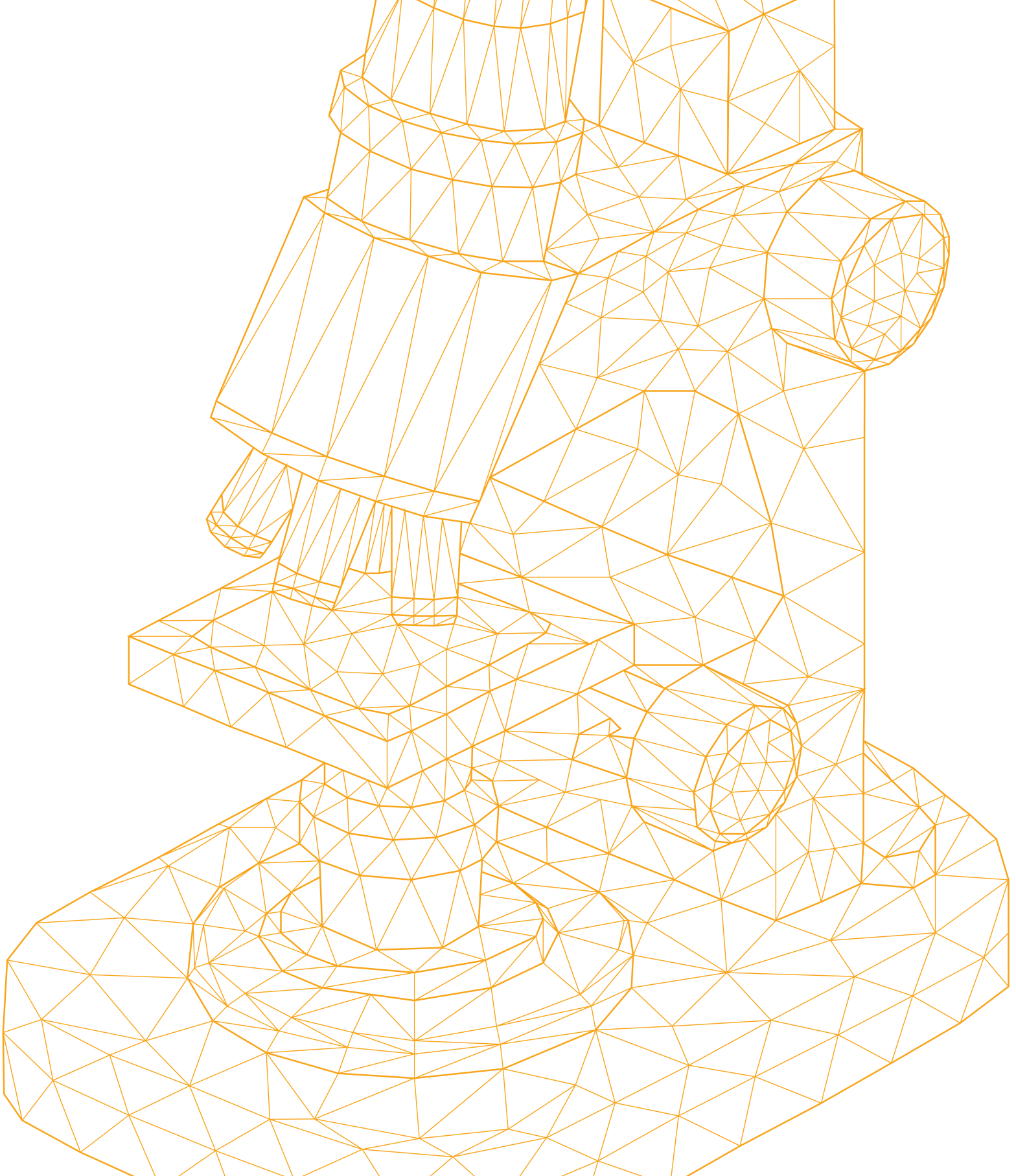
— Включение технологий ИИ в процесс обучения предполагает постепенную трансформацию системы образования. Это касается и подходов к подаче информации, ее усвоения, проверки. Важно учитывать возрастные ограничения и использовать специализированные решения, включая инструменты родительского контроля. Нежелательно использование детьми общедоступных, не специализированных решений на базе ИИ в образовательных целях вне соответствующих учреждений.

Алексей Хабибуллин,
руководитель дирекции по довузовскому
и олимпиадному обучению,
ИТ Кампус «Неймарк»

— Необходимо создавать и реализовывать специальные программы для обучения педагогов, студентов педагогических вузов и родителей тем возможностям, которые могут дать нейросети в системе образования и развития ребенка. Ограничение использования технологий ИИ если и допустимо, то под контролем взрослых.

глава 6

ИИ и медицина



31 Этично ли человеку заниматься самолечением с помощью ИИ?

Ответ:

Самолечение само по себе может нести риски для здоровья человека, но для поиска решений в вопросах своего здоровья можно использовать только специализированные и сертифицированные системы ИИ, которые созданы для этого и проверены на точность. Обычные чат-боты, не предназначенные для медицинских целей, не обладают необходимой достоверностью и могут давать неверные рекомендации, поэтому их использование небезопасно.

Обоснование:

- Группа американских исследователей считает, что **технологии ИИ могут частично решить проблему нехватки специалистов и создать дополнительные возможности жителям труднодоступных районов**[237]. Во многих сельских районах развивающихся стран мало квалифицированных врачей, и большое количество пациентов нуждаются в помощи медсестер или среднего медицинского персонала.
- **Возможность использования специализированных ИИ-систем для помощи в вопросах здоровья позволяет пациентам получить оперативные рекомендации и информацию о состоянии здоровья**.
- В России было проведено исследование, когда чат-боту без медицинской специализации задавали одинаковый вопрос с указанием разных ролей врача. В одном случае, будучи «гастроэнтерологом», бот предложил такие диагнозы, как острый аппендицит, панкреатит и холецистит. Однако когда бот играл роль гинеколога, он предложил другие диагнозы, например ПМС или киста яичника. Это показывает, что о**бычные боты ориентируются на ключевые слова и не способны оценить полную клиническую картину, что делает их ненадежными для медицинского использования**.

Рекомендации для разработчиков:

1. **Проводите анализ потенциальных рисков при разработке медицинских ИИ-систем**. Важно учитывать возможные последствия предоставления недостоверной информации.
2. **Разрабатывайте системы с учетом этических принципов безопасности и справедливости**. В случае выявления жизненно угрожающих состояний система должна рекомендовать немедленное обращение за медицинской помощью, соблюдая принципы врачебной этики и благополучия пациента.

Рекомендации для пользователей:

1. **Используйте только специализированные ИИ-системы для принятия решений в вопросах здоровья.** Такие системы разработаны с учетом медицинских стандартов и обладают высокой достоверностью, что снижает риск ошибок.
2. **Консультируйтесь с медицинскими специалистами при использовании ИИ-систем**. Даже проверенные ИИ-инструменты не заменяют профессионального мнения врача и должны использоваться как дополнительная поддержка, а не основной источник рекомендаций.

Исследование по вопросу:

В июне 2024 года KFF, ведущая организация в области политики здравоохранения в США, провела опрос среди американских граждан на тему использования чат-ботов для получения медицинской информации. Согласно исследованию, **примерно каждый шестой взрослый (17%) говорит, что использует чат-ботов с ИИ не реже одного раза в месяц, чтобы получить медицинскую информацию и советы**, а среди взрослых в возрасте до 30 лет этот показатель достигает четверти (25%). Большинство взрослых, включая большинство (56%) тех, кто использует ИИ или взаимодействует с ним, не уверены в точности медицинской информации, предоставляемой чат-ботами на базе искусственного интеллекта[238].

Практика:

Лаборатория компьютерных наук и искусственного интеллекта Массачусетского технологического института в 2021 году разработала **ИИ-инструмент для отслеживания правильности приема лекарств и напоминаний с направлением этих данных врачу**. Беспроводной датчик устанавливался дома у пациента. Система ИИ непрерывно и автоматически анализировала радиосигналы и документировала результаты, которые загружались через интернет и добавлялись к цифровой медицинской карте пациента. Пациенту отправлялись напоминания, если он не принимал лекарство в назначенное время. Уполномоченные медицинские работники также получали доступ к этим записям для отслеживания состояния пациентов[239].

Что думают эксперты?

Диана Хасанова,
доцент кафедры цифровых технологий в здравоохранении Казанского ГМУ, генеральный директор «Брейнфон»

— Доступность медицинской помощи в различных субъектах России разнится, что отражается на качестве и продолжительности жизни населения страны. ИИ-инструменты способны помочь выровнять возможности медицинской помощи, особенно в труднодоступных регионах, что в условиях больших расстояний и стареющего населения страны является одним из приоритетов в здравоохранении.

Павел Воробьев,
профессор, Председатель Правления Московского городского научного общества терапевтов

— Жители многих тысяч поселков страны лишены контакта с медицинскими работниками. В них нет ничего, кроме связи, но есть медицинские уполномоченные, не имеющие медицинского образования. Вот им-то в помощь и нужны системы поддержки принятия решений, основанные в том числе на ИИ, чтобы выстроить адекватный механизм экстренной и плановой помощи-сопровождения.

32 Этично ли врачу делегировать ИИ принятие решений по профилактике, диагностике, лечению и реабилитации?

Ответ:

По общему правилу — нет. Применение ИИ в диагностике, лечении и реабилитации можно признать этичным только в случае, если его рекомендации проверяются и подтверждаются квалифицированными специалистами. ИИ может помогать врачам, подготавливая заключения и рекомендации, но окончательное решение должен принимать человек.

Обоснование:

- Как следует из стратегического документа[240] Европейского Парламента «Роботы в здравоохранении: решение или проблема?», в настоящее время накоплено недостаточно опыта. **Существующая нормативная и этическая база не позволяет полностью устранить риски для безопасности пациентов**, что необходимо для укрепления доверия и принятия пользователями (как пациентами, так и здоровыми гражданами).
- Исследователи Университета науки и технологий Халифы (ОАЭ) считают, что **проблема «черного ящика», растущая неоднозначность и сложность интерпретации алгоритмических функций**, охватывающая как процесс обучения, так и достоверность результатов, создают серьезные препятствия для делегирования системам ИИ принятия решений. Это также не отвечает этическому критерию объяснимости (почему было принято то или иное медицинское решение)[241].
- **Пациенты могут испытывать недоверие и дискомфорт, зная, что их лечение полностью контролируется ИИ**. Контроль СИИ врачом обеспечивает более высокий уровень поддержки и доверия и снижает риски.
- Группа ученых из Швейцарии и США напоминает[242], что **ИИ может демонстрировать потенциальную предвзятость в отношении определенных групп пациентов**, например из-за недостаточности обучающих данных. Это может привести к дискриминации и нарушению принципов биоэтики.
- **Полная передача принятия решений ИИ может привести к снижению человеческого контроля и чувства ответственности** (риск потери автономии в принятии решений).

Рекомендации для разработчиков:

1. **Учитывайте принципы безопасности, достоверности и конфиденциальности данных**. Это минимизирует риск дискриминации и злоупотреблений.
2. **Разрабатывайте ИИ-системы с возможностью объяснения и прозрачности принятых решений**. Это позволит медицинским специалистам лучше понимать логику работы ИИ и повышать доверие к его рекомендациям.

Рекомендации для медицинских работников:

1. **Используйте ИИ как инструмент для поддержки принятия решений, но окончательное решение принимайте самостоятельно**. Важно помнить, что ИИ предоставляет информацию, а не заменяет врачебный опыт.
2. **Проверяйте и оценивайте рекомендации ИИ в контексте каждого конкретного случая**. Адаптируйте предложения ИИ, учитывая индивидуальные особенности пациента и клиническую ситуацию, чтобы избежать ошибок.

Исследование по вопросу:

Американский исследовательский центр Pew Research Center изучил общественное мнение об ИИ в здравоохранении и медицине. Шесть из десяти взрослых американцев говорят, что чувствовали бы себя некомфортно, если бы их лечащий врач полагался на ИИ при диагностике заболеваний и рекомендации методов лечения. Также опрос показал, что только 38% считают, что использование искусственного интеллекта для диагностики заболеваний и рекомендации методов лечения приведет к улучшению состояния здоровья пациентов[243].

Практика:

В России ИИ уже активно применяется в медицине. Так, в 2023 году российские регионы закупили 106 медицинских изделий с искусственным интеллектом на общую сумму приблизительно 448,43 млн рублей. Такие решения внедрены в 85 субъектах РФ[244].

Более того, в России к концу 2023 года с помощью ИИ-инструментов проанализированы 22 млн медицинских записей. В шести регионах применяются сервисы голосового заполнения документов, а в 29 — виртуальные ассистенты с ИИ для записи на прием к врачу.

Что думают эксперты?

Булат Магдиев,
центр исследований медицинских изделий
Сеченовского университета

— Использование ИИ в здравоохранении может быть этически приемлемым при условии, что ИИ действует в качестве помощника врача, а не заменяет его полностью. Комбинация возможностей ИИ и экспертного мнения врача может повысить точность диагностики, оптимизировать лечение и улучшить результаты для пациентов, при этом сохраняя человеческий контроль и ответственность.

Борис Зингерман,
директор ассоциации поставщиков
и пользователей ИИ в медицине
«Национальная база медицинских знаний»

— На сегодняшний день решений автономного искусственного интеллекта, зарегистрированного по всему миру, крайне мало — в пределах 10 штук. Но тем не менее они существуют и, вероятно, в будущем будут важны. То есть они, конечно, должны проверяться с гораздо большей надежностью, чем те решения, где все-таки человек рефлексирует окончательный результат, но это те направления, которые для нас принципиально важны.

Елена Брызгалина,
завкафедрой философии образования
философского факультета
МГУ имени М. В. Ломоносова,
руководитель магистерской
программы «Биоэтика»

— ИИ выполняет лишь ассистирующую функцию как система поддержки врачебных решений. Применение систем ИИ в медицине связано с возможным причинением вреда. Определение ответственности за причиненный вред связано с анализом действий именно человека — врача — или медицинского учреждения, использующего системы ИИ как инструменты поддержки врачебных решений. Делегирование принятия решений с возложением на ИИ ответственности невозможно.

33 Этично ли сообщать «плохие новости» пациенту с помощью ИИ?

Ответ:

Нет, сообщать «плохие новости» с помощью системы ИИ можно считать неэтичным, так как такой способ сообщения медицинской информации может травмировать пациента.

Обоснование:

- Новость о серьезном диагнозе вызывает сильные эмоциональные реакции: шок, страх, гнев или печаль. **Живой врач может оценить состояние пациента, предложить необходимую поддержку** и при необходимости пригласить специалистов.
- **Важно не только передать информацию, но и убедиться, что она правильно понята**. Врач способен ответить на вопросы, разъяснить детали и поддержать пациента, чего ИИ может не сделать должным образом.
- Согласно исследованию, опубликованному в журнале Health Care Science, **сообщение плохих новостей требует огромного мастерства и осторожности**, поскольку пациенты часто испытывают симптомы тревоги и депрессии после того, как им был поставлен страшный диагноз. Для минимизации психологического вреда пациентам были разработаны различные рекомендации сообщения плохих новостей. Рекомендации, которые следует использовать врачу при сообщении «плохих новостей», включают протоколы SPIKES, BREAK, FINE и другие. **В теории этим же принципам можно обучить и СИИ**[245].
- По мнению индийского доктора Лиджи Томаса, в случае повсеместного внедрения СИИ и чат-ботов на базе ИИ **врачи могут потерять навыки общения с пациентами в сложных ситуациях и будут избегать подобного общения**. Пациенты, почувствовав отчуждение своего лечащего врача, начнут прибегать к самодиагностике и самолечению[246].

Рекомендации по использованию ИИ для сообщения «плохих новостей»:

1. **Врачам не рекомендуется заменять живое общение с пациентами на ИИ-чат-боты**. Из этого правила могут быть исключения: например, ИИ может быть полезен для поддержки пациентов после сообщения врачом «плохих новостей», он может напоминать о приеме лекарств и сопровождать пациента в процессе реабилитации.

2. **Разработчикам следует учитывать возможные эмоциональные и психологические последствия для пациентов и врачей.** Внедрение автоматизированных систем коммуникации с пациентом должно основываться на тщательном анализе соотношения пользы и рисков.

3. **Внедряйте врачебные этические принципы в использовании ботов для информирования пациентов.** Эти принципы должны учитывать потребности пациента и обеспечивать возможность взаимодействия с врачом.

Исследование по вопросу:

В 2023 году американские врачи решили провести эксперимент и попросили ChatGPT помочь им общаться с пациентами более сочувственно. По итогам исследования оказалось, что **ответы, созданные программой, оказались более чуткими, чем у настоящих врачей**.

На основе этих данных были проведены исследования, где экспертам от медицины предложили сравнить подачу плохих новостей для пациента от врача и ChatGP T. Выяснилось, что 78,6% опрошенных людей предпочли бы слышать ответ ИИ[247].

Практика:

В 2019 году **врач одной из клиник Калифорнии доверил роботу сообщить пациенту о тяжелом диагнозе**. Пациент, не готовый к такому представлению информации, был шокирован так же, как и его родственники. Уилхарм, внучка пациента, сказала журналистам: «Я думаю, врачам следовало проявить больше достоинства и относиться к моему дедушке лучше, чем они это делали». Ее дедушка, 78-летний Эрнест Кинтана, скончался на следующий день после объявления диагноза[248].

Что думают эксперты?

Елена Гребенщикова,
директор института гуманитарных наук,
зав. кафедрой биоэтики
РНИМУ им. Н. И. Пирогова

— Сообщение «плохих новостей» пациенту и/или членам его семьи требует особого настроя, деликатности, учета эмоционального состояния, готовности поддержать человека в трудную минуту и показать, что информация не станет роковой чертой, после которой от него отвернутся врачи. Робот не сможет это сделать полноценно, а пациенты будут обвинять систему здравоохранения в бездушном и бесчеловечном отношении. Кроме того, робот не в состоянии почувствовать нежелание или неготовность пациента получать информацию, не сможет понять контекст ситуации, который позволяет врачу выбирать правильные слова, место и время для сообщения информации. «Плохие новости» о здоровье ребенка оказывают негативное влияние на всю семью. У родителей возникают вопросы, которые, например, связаны с другими детьми в семье — надо ли им сообщать о ситуации, есть ли угроза их здоровью. Кроме того, робот не сможет осознать, что необходимо повторить информацию, убедиться в ее адекватном понимании, реагируя как на психологические моменты, так и на практические запросы родителей.

Анастасия Углева,
профессор Школы философии
и культурологии, заместитель директора
Центра трансфера и управления
социально-экономической
информацией НИУ ВШЭ

— Само по себе сообщение «плохой новости» ничем не отличается от информации, которая содержится, например, в электронной медицинской карте о результатах анализа или описания приема у врача. При этом пациент должен иметь право выбора, с кем ему общаться о состоянии своего здоровья — с живым доктором, разговорным ассистентом или чат-ботом. Однако вопросы этики возникают, на мой взгляд, не столько в момент произнесения рокового диагноза и нужных слов поддержки (с этим вполне может справиться и ИИ), сколько в связи с необходимостью контроля за последствиями как медицинского, так и социального характера. Если ИИ-технология будет способна в режиме реального времени оценить риски резкого ухудшения самочувствия пациента и/или возникновения у него, например, суицидальных намерений в результате полученной информации, а также сможет оперативно оказать ему первую психологическую помощь, то следует признать применение такого ИИ этичным. В иных случаях от применения ИИ следует отказаться.

34 Нужно ли получать отдельное согласие пациента на применение ИИ для лечения?

Ответ:

Представляется этичным раскрывать пациенту информацию о применении ИИ врачом и получать его согласие на применение этой технологии в рамках общего согласия на осуществление медицинских манипуляций. При этом лечение всегда должен осуществлять человек, а ИИ выступать исключительно инструментом.

Обоснование:

- В России **перед проведением медицинского вмешательства пациент должен дать информированное согласие**[249]. Для этого медицинский работник должен предоставить полную информацию о целях, методах, рисках, вариантах вмешательства, его последствиях и предполагаемых результатах в доступной форме.
- Испанские ученые отмечают, что положения закона о персональных данных ЕС об автоматизированном принятии решений применяются только в том случае, когда решение основано исключительно на ИИ. Это означает, что **в ситуациях, когда ИИ используется в качестве инструмента поддержки принятия решений, нет юридического обязательства информировать пациентов о его использовании**[250].

Рекомендации для медицинских работников:

1. **Следует раскрывать информацию об использовании ИИ при оказании медицинской помощи** для обеспечения прозрачности, ответственности и уважения автономии пациента.
2. **Следует рассматривать процедуру подписания информированного добровольного согласия как повод для полного информирования пациента**, а не формальную процедуру.
3. **Следует обновлять знания о ключевых аспектах работы ИИ в медицине**, необходимых для адекватного информирования пациента.
4. **Следует проводить регулярные мероприятия по повышению осведомленности общества о возможностях, ограничениях и основных принципах работы ИИ в медицине**, а также связанными с ним рисками.
5. **Следует разъяснять пациенту, кто ответственен за медицинскую помощь, оказанную с применением ИИ,** и явно обозначать его право отказаться от медицинского вмешательства.

Рекомендация для пациентов:

1. **Перед подписанием информированного согласия следует задавать прямые вопросы медицинским работникам об этапах, методах и рисках оказания медицинской помощи, в том числе с применением ИИ.**

Подход ВОЗ:

Из этических рекомендаций Всемирной организации здравоохранения следует, что:

- технологии ИИ не должны использоваться для экспериментов или манипулирования людьми в системе здравоохранения без действительного информированного согласия;
- использование алгоритмов машинного обучения в диагностике, прогнозировании и планах лечения должно быть включено в процесс получения осознанного и действительного согласия;
- предоставление основных услуг не должно быть ограничено или в них не должно быть отказано, если человек не дает согласия, и ни правительство, ни частные лица не должны предлагать дополнительные стимулы или побуждения лицам, которые дают согласие[251].

Исследование по вопросу:

Исследователи Юридической школы Университета Ханьян в Южной Корее провели опрос 1000 респондентов для оценки пациентами важности информирования об использовании ИИ в диагностике при принятии решения о проведении лечения. Результаты опроса показали, что люди придают большее значение информации о применении ИИ в диагностике по сравнению с консультацией с человеком-специалистом, например радиологом. Это указывает на то, что сравнение консультации с ИИ и консультации с человеком не отражает всей картины и не оправдывает практику врачей не раскрывать информацию об использовании ИИ для поддержки принятия решений[252].

Участники опроса восприняли информацию о применении ИИ как более важную или равнозначную нижней границе регулярно раскрываемой информации, что подчеркивает необходимость предоставления сведений об использовании ИИ в диагностических процедурах. Это подтверждает, что раскрытие информации о применении ИИ в диагностике является важным аспектом взаимодействия врача и пациента, способствующим повышению доверия и пониманию процесса принятия решений о лечении.

Что думают эксперты?

Елена Гребенщикова,
директор института гуманитарных наук,
зав. кафедрой биоэтики
РНИМУ им. Н. И. Пирогова

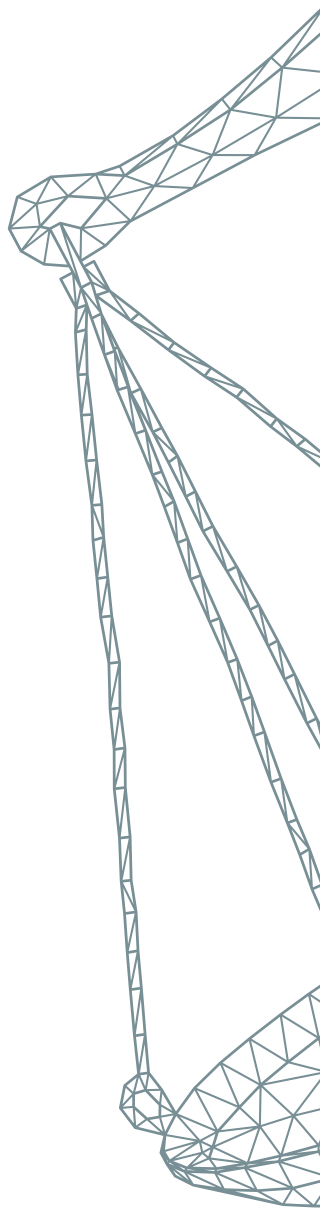
— Необходимость отдельного информированного согласия должна определяться его функциями и выбором пациента. Например, если ИИ используется врачом только в консультативных целях, то любое решение врача — это только его выбор, соответственно, отдельное ИДС не потребуется. Но если врач в ходе консультации предложил использовать ИИ в диагностических целях, то пациент должен быть полностью проинформирован и подписать форму ИДС. Цель внедрения ИИ в здравоохранение — улучшить качество медицинской помощи, облегчить жизнь и пациентам, и врачам, что невозможно без учета сложившихся норм медицинской этики, среди которых информированное добровольное согласие играет ключевую роль.

Павел Воробьев,
профессор, Председатель Правления
Московского городского научного
общества терапевтов

— Должен быть использован принцип разумной достаточности: вынесение на обсуждение с пациентом всех тонкостей используемых медицинских технологий, в том числе использующих СИИ, является абсолютно избыточным, так как решения в существующей системе здравоохранения принимает медицинский работник, а СИИ играет лишь вспомогательную, хотя и важную роль.

глава 7

ИИ в правосудии



35 Этично ли применять ИИ судьям?

Ответ:

Да, если это допускается законодательством, ИИ этично и целесообразно применять как инструмент повышения эффективности, например при составлении резюме по делу или для автоматизации и упрощения поиска по судебной практике.

Обоснование:

- Исследователи Университета Нью-Гэмпшира считают, что **ИИ выступает только в качестве помощника судьи**. Резюме не предрешает оценку доказательств, которую дает судья в процессе установления обстоятельств дела и применения закона, а помогает судье быстро усваивать доказательственный материал и исключать ошибки[253].
- **Резюме обобщит доказательства, проведет их фильтрацию по определенным признакам** — допустимости с точки зрения закона (например, укажет на отсутствие нотариального удостоверения, когда оно требуется) и относимости (укажет, что доказательство очевидно не относится к делу).
- По мнению исследователей индийского университета Уокссен, **ИИ может помочь судам сократить время на рассмотрение дел**, предоставляя точную информацию и анализ, основанный на прецедентах. Это ускоряет процесс принятия решений и повышает точность и тщательность юридической оценки[254].
- Группа индийских ученых утверждает, что **ИИ расширит возможности судьи по анализу судебной практики**: быстро подберет примеры актуальных судебных решений и сделает краткую выжимку правовых позиций. Обработка естественного языка, машинное обучение и аналитика данных стали незаменимыми инструментами для быстрого просмотра юридических документов, выявления необходимой информации и прогнозирования исходов дел[255].

Рекомендации для судей:

1. **Приоритетным в принятии решения о возможности применения или неприменения ИИ прежде всего являются законодательство и позиции высших судебных органов по этому вопросу**.
2. **Следует применять ИИ для анализа объемных данных и поиска судебной практики, но оставляйте окончательное решение за собой**. Использование ИИ помогает быстро обрабатывать информацию, но окончательное суждение, как правило, требует человеческой оценки всех обстоятельств дела.

3. **Использование инструментов обобщения и саммаризации материалов дела и судебной практики должно находиться под контролем судьи**. Судья должен иметь возможность проверять и корректировать выводы ИИ, чтобы избежать ошибок и обеспечить справедливое рассмотрение дела.
4. На этапах внедрения и опытной эксплуатации **следует перепроверять ИИ на достоверность представляемой информации** и действительное наличие соответствующих судебных актов и законов.

Практика:

1. В октябре 2021 года Кассационный суд Франции **запустил цифровую базу данных Judilibre**, содержащую 480 000 судебных решений, вынесенных с 1947 года. Изначально предназначавшаяся для судей и адвокатов, Judilibre постепенно станет доступна заявителям к 2025 году. Инструмент использует ИИ для оптимизации исследований и систематизации судебных решений. ИИ также позволяет псевдонимизировать данные[256].

2. **В США используется модель, основанная на технологии ИИ, — Caselaw Access**.

   Эта система включает набор данных из более чем 6,7 млн дел и позволяет определять исход дела на основе соответствующих прецедентов, решений судей и справочных заявлений из более чем 400 судов. Caselaw Access позволяет судьям быстро находить релевантные дела, аналогичные рассматриваемому делу, и учитывать их при вынесении решения[257].

3. В мае 2024 года Управление по развитию информационно-коммуникационных средств массовой информации Сингапура (IMDA) объявило о сотрудничестве с Singapore Academy of Law (SAL) для совместной разработки новой большой языковой модели, которая сделает юридические исследования более быстрыми и эффективными. Известная как GPT-Legal модель будет развернута на LawNet поэтапно с сентября 2024 года. На первом этапе внедрения **GPT-Legal будет использоваться для обобщения более 15 000 судебных решений Сингапура**, предоставляя краткую информацию о ключевых словах, фактах и выводах из судебного решения[258].

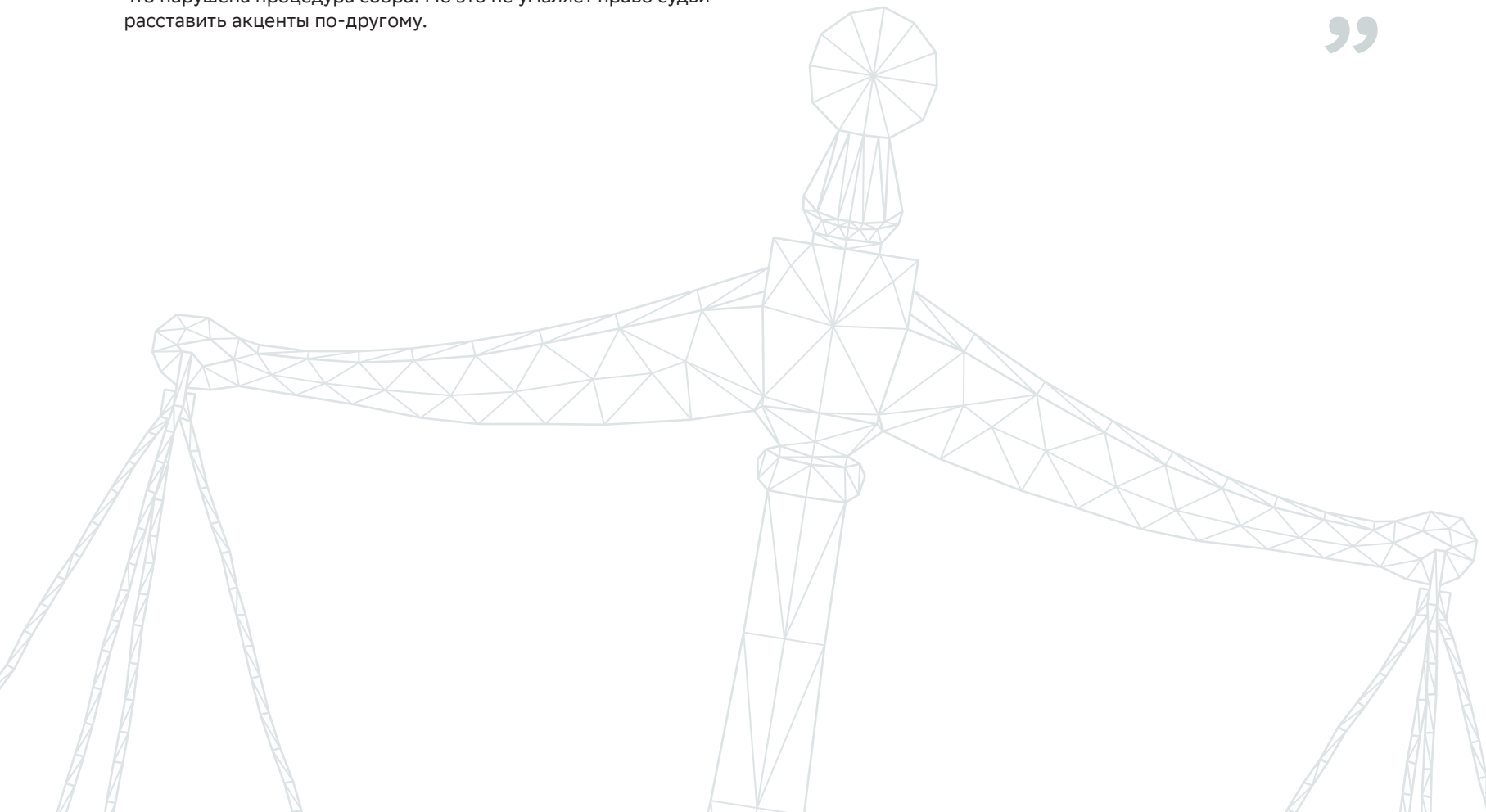
Что думают эксперты?

Елена Авакян,
вице-президент Федеральной Палаты адвокатов РФРНИМУ им. Н. И. Пирогова

— Резюме по делу, подготовленное ИИ, важно для судьи, потому что он может не изучать весь объем материалов, из которого не все будет иметь доказательственное значение, а также важно для сторон, чтобы они понимали, на какие аспекты обратил внимание суд и что им нужно усилить в своей позиции. ИИ сможет подсветить основные проблемы собранных доказательств, например указать, что это порочное доказательство, потому что нарушена процедура сбора. Но это не умаляет право судьи расставить акценты по-другому.

Игорь Емшанов,
председатель комиссии Совета судей Амурской области по автоматизации и информатизации судов

— В условиях ограниченности временного ресурса, которую всегда испытывают судьи, возможности ИИ реферировать тексты пришлись бы очень кстати. Подбор и интеллектуальный пересказ ранее принятых другими судами решений позволили бы судье быстро оказаться в теме и подготовиться к делу.



36 Этично ли использовать ИИ сторонам судебного дела?

Ответ:

Да, при соблюдении рекомендаций ниже ИИ может помогать преодолеть проблему недостатка юридической подготовки сторон: например, описать правила подсудности и порядок обращения в суд, помочь в составлении процессуальных документов или ответить на сложные правовые вопросы граждан простым языком. При этом важно помнить, что для такого применения ИИ всегда необходимо учитывать требования законодательства и позиции органов судебной власти по этому вопросу (если они есть).

Обоснование:

- По мнению исследователей из Юридической школы Саффолкского университета, **ИИ может повысить доступ к правосудию за счет перевода сложных юридических правил на простой язык гражданина** и ответов на конкретные юридические вопросы[259].
- **Использование ИИ для правового обоснования заявления сопряжено с риском фактических ошибок, которые может допустить система**. Например, подготовленные судами Квинсленда в Австралии «Руководящие принципы по использованию генеративного ИИ неюристами» возлагают на заявителя ответственность за точность и достоверность информации, полученной от чат-бота и представленной в суд[260].
- Исследователи из Университета Конкордия и Университета Монреаля отмечают, **что доступ к правосудию затрудняется высокой стоимостью консультаций**. Работа ИИ-чат-ботов позволит сэкономить на платной юридической помощи[261].

Рекомендации для судов:

1. **Информируйте граждан о возможностях, преимуществах и рисках использования систем ИИ**. Помогая преодолеть формальные процедуры обращения в суд, такие системы повышают доступность правосудия для широких слоев населения.
2. **Выбирайте надежные и проверенные системы ИИ**, которые соответствуют требованиям законодательства и обеспечивают безопасность данных.

Рекомендации для разработчиков:

1. **Соблюдайте принципы прозрачности при разработке ИИ-систем**. Системы должны быть объяснимыми и понятными для конечных пользователей, особенно когда речь идет о юридических рекомендациях и ответах на вопросы.
2. **Обеспечьте техническую поддержку и обучение для пользователей**. Создайте подробные руководства, чтобы граждане и сотрудники судебной системы могли легко освоить работу с ИИ-системами и избегать типичных ошибок.

Рекомендации для пользователей:

1. **Проверяйте правовые аргументы и рекомендации, предложенные ИИ, перед их использованием**. ИИ иногда допускает ошибки, которые могут ввести в заблуждение. Всегда сверяйте данные с надежными источниками и консультируйтесь с юристами, если есть сомнения.
2. **Используйте ИИ как инструмент помощи для подготовки**. Окончательные выводы и решения должны быть основаны на консультациях с профессиональными юристами.

Практика:

1. **В 2017 году в Народном суде Пекина был запущен в эксплуатацию робот по имени Xiaofa**. Робот ростом 1,46 метра предоставляет консультации посетителям, отвечая на сложные юридические вопросы простым языком. Он может двигать головой и махать руками, когда на экране появляются инструкции, и направлять людей к нужному окну для получения судебных услуг.

   Инструмент на базе ИИ способен ответить на более чем 40 000 процессуальных и 30 000 правовых вопросов. В результате его внедрения удалось значительно ускорить процесс обращения в суд[262].

2. **В Аризоне (США) активно используют чат-ботов для автоматизации правосудия**. Так, используется бот, который по запросу пользователя оценивает возможность снятия судимости. При положительном ответе бот помогает с заполнением ходатайства и подачей его в суд[263].

   Еще один чат-бот специально разработан для помощи в спорах, вытекающих из договора аренды. Бот умеет давать пошаговые инструкции по разрешению арендных споров и предоставлять рекомендации по заполнению процессуальных документов.

3. Первый широко известный **случай ответственности за предъявление суду правовой позиции с фактическими ошибками ИИ** произошел в Нью-Йорке (США). Адвокат Стивен Шварц, готовясь к иску, воспользовался ChatGPT для поиска прецедентных дел. В ходе разбирательства выяснилось, что чат-бот сфабриковал кейсы и даже указал, что несуществующие решения вынесли действующие судьи. Теперь юрист должен выплатить $5000 и уведомить каждого из судей, чьи имена фигурировали в вымышленных материалах[264].

   К сожалению, это не единственный случай в США, когда юристы не проверяли ответы нейросети на достоверность. Поэтому в июле 2024 года **Американская ассоциация адвокатов выпустила этические рекомендации** по использованию ГенИИ в профессиональной деятельности[265].

Что думают эксперты?

Владимир Ярков,
заведующий кафедрой гражданского процесса Уральского государственного юридического университета имени В.Ф. Яковлева

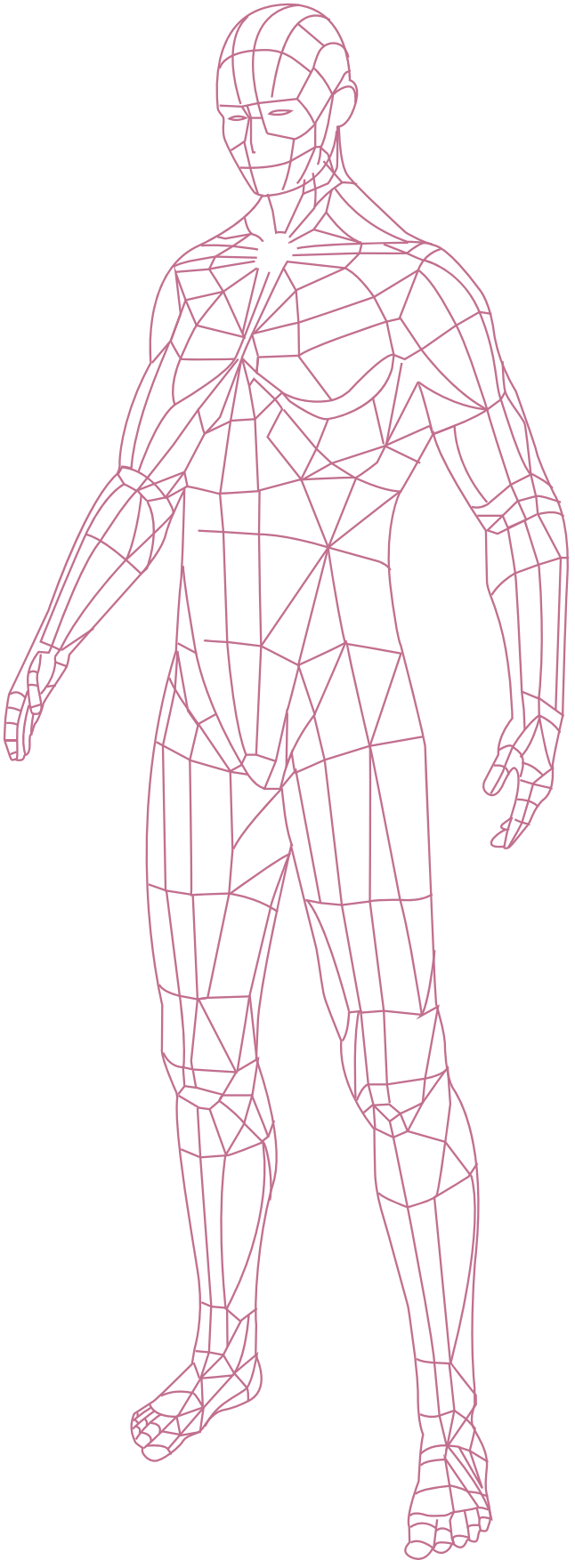
— Полагаю, что да, с соблюдением таких базовых принципов судебного процесса, как состязательность и равноправие сторон. Почему бы стороне не использовать ИИ для сбора и анализа законодательства, судебной практики, обработки доказательственной базы, учитывая ее существенный объем по сложным делам, для моделирования поведения другой стороны и суда, etc. В конечном счете ИИ как инструмент будет служить целям оптимального и эффективного разрешения спора, более результативному представлению позиции перед судом. Другой вопрос — что равноправие сторон предполагает равные возможности правовой защиты сторон, поэтому сторона, которая будет лишена или ограничена в доступе к ИИ, не сможет, скорее всего, столь результативно представить свою позицию суду. Поэтому задача законодателя и суда — обеспечить не формально-юридическое, а фактическое равенство при доступе к системам ИИ[266].

Виктор Момотов,
председатель Совета судей России

— Представляется необходимым адаптирование судебной системы для гражданина, не обладающего специальными знаниями, таким образом, чтобы процесс обращения за судебной защитой был простым для понимания.
Нормативные акты — одни из самых сложных юридических текстов, а судебные акты при этом еще сложнее для восприятия. В связи с этим необходимо предусматривать механизмы, позволяющие взаимодействовать с судебной системой на доступном для непрофессионала языке[267].

глава 7

ИИ и человек



37 Можно ли оказывать психологическую помощь с применением ИИ?

Ответ:

Специальный чат-бот может помогать с решением отдельно взятых психологических проблем, только следуя четкому протоколу действий, полученных от специалиста. Он также может использоваться для направления человека к нужному специалисту для дальнейшей помощи.

Обоснование:

- Исследовательская служба Европейского парламента (EPRS) считает, что **ИИ может быть использован для выявления сложных психических расстройств**. ИИ умеет различать диагнозы с совпадающими клиническими проявлениями, прогнозировать эффективность антидепрессантов и анализировать риски ухудшения состояния[268].
- Кафедра психиатрии и поведенческих наук Медицинской школы Макговерна (США) утверждает, что **чат-боты расширяют доступность психологической помощи**. ИИ смягчает последствия нехватки медицинского персонала, предоставляя круглосуточную поддержку независимо от географии или временных ограничений[269].
- Согласно исследованию, опубликованному в научном журнале Cambridge Science Advance, **ИИ снижает риски стигматизации и дискриминации**. Например, некоторые люди, страдающие депрессией или ПТСР, могут избегать общения с людьми. Более того, врачи могут ставить ошибочные диагнозы из-за фиксации на социальных факторах (возраст, раса, пол)[270].

Рекомендации для специалистов

ИИ может применяться для:

1. мониторинга психического состояния клиента;
2. тренировки базовых навыков психологической саморегуляции;
3. выявления опасных паттернов поведения;
4. оценки динамики эффективности психологической помощи;

и других задач консультативного процесса, требующих регулярной самостоятельной работы клиента и их последующего обсуждения со специалистом.

ИИ не следует использовать для:

1. коррекции психических расстройств, подтвержденных медицинским диагнозом;
2. анализа и нормализации семейных отношений;
3. работы с психологическими последствиями травматизации;
4. работы с апатией, депрессивными состояниями, суицидальными мыслями и намерениями;

и других запросов, для решения которых требуется понимание специалистом индивидуальной картины мира клиента.

Рекомендации для пользователей

ИИ может помогать в решении следующих запросов:

1. стратегии эффективного планирования времени;
2. стратегии индивидуального совладания с ситуативным стрессом и тревогой;
3. тренировки навыков эффективной коммуникации;

и других запросов, для которых существуют верифицированные рекомендации по тренировке определенных психологических навыков.

Окончательное решение об использовании чат-бота должно приниматься после всесторонней экспертной оценки рисков профессиональным психологическим сообществом.

Исследование по вопросу:

В 2022 году компания Wysa выпустила **отчет о состоянии психического здоровья американских служащих**. Когда респондентов спросили, к кому бы они предпочли обратиться по поводу своего психического здоровья, они с большей вероятностью выбрали «приложение для психического здоровья с клинически проверенными ресурсами самопомощи, адаптированными к их потребностям», чем к кому-либо на рабочем месте и даже к своему лечащему врачу[271].

К кому бы предпочли обратиться американские служащие за психологической помощью?

Источник: Wysa[271]

Практика:

В начале 2023 года американский журнал Vice опубликовал статью о том, как американская некоммерческая организация поддержки психологического здоровья Koko **в порядке эксперимента заменила специалистов чат-ботом**, не уведомив клиентов. Чат-боту удалось «проконсультировать» примерно 4000 человек. Клиенты выше оценили сообщения, полученные от чат-бота, нежели сообщения, написанные специалистом[272].

Что думают эксперты?

Николай Седашов,
управляющий партнер Spektr

— Самые эффективные в достижении терапевтических целей и последовательные решения — это те, что сочетают ИИ с поддержкой живых специалистов. Неплохой пример — британское приложение Wysa. В нем помощь можно получить через чат-бот, но ИИ не только поддерживает пользователя и советует методики самопомощи, но и при необходимости может позвать на помощь живого терапевта. ИИ обеспечивает доступность и оперативность, а врачи — глубину и персонализацию[273].

Иван Оселедец,
генеральный директор Института AIR

— Модели естественного языка могут применяться для анализа тысяч часов психотерапевтических сессий с целью выявления областей, в которых молодые специалисты упускают из вида значимые факторы. Например, не задают вопросы, ответы на которые способны изменить представление об анамнезе пациента. Число LLM, применяющихся в сфере психического здоровья, стремительно растет — есть основание полагать, что эти темпы сохранятся.

Мария Чумакова,
доцент департамента психологии факультета социальных наук НИУ ВШЭ, руководитель проекта Центра искусственного интеллекта НИУ ВШЭ.

— Взаимодействие в рамках психологической помощи в первую очередь опирается на акты человеческого сопереживания, эмпатии и принятия. Эти акты происходят в условиях встречи человека и другого человека, в рамках которой соприкасаются внутренние миры обоих. Другой для человека является бесконечным источником неопределенности, стимулирующей развитие. Другой является носителем иной уникальной картины мира, в то время как ИИ является носителем обобщенной картины мира. Замена встречи с уникальностью на встречу с обобщенным знанием может привести к редуцированию у клиента способности к сопереживанию и потере внутренней мотивации к развитию.

38 Нужно ли ограничивать темы и модерировать токсичный контент при общении с ИИ?

Ответ:

Да, для предотвращения серьезных последствий следует ограничить список обсуждаемых тем, исключив чувствительные, и принять меры для предотвращения токсичного контента. При недостаточности мер важно информировать пользователя о возможных рисках.

Обоснование:

- Исследователи Центра по управлению ИИ в Великобритании утверждают[274], что **разработчики вручную настраивают модель для того, чтобы предотвратить создание и распространение запрещенного контента**. Однако надо иметь в виду, что тонкая настройка модели может привести к отказу от некоторых безобидных запросов.
- **ИИ может поспособствовать повышению уровня правовой грамотности населения**. Запрос, не соответствующий правовым нормам, действующим на территории РФ, может быть вызван незнанием пользователем действующего законодательства.
- Согласно Руководству ЮНЕСКО по использованию ГенИИ в образовании и научных исследованиях, **модерация контента способствует соблюдению основных прав человека**, уважению интеллектуальной собственности и соблюдению этических норм, а также предотвращению распространения дезинформации и разжигания ненависти[275].
- **По любой чувствительной теме запрос пользователя может быть как конструктивным, так и деструктивным**. В случае конструктивного запроса ИИ должен помочь пользователю и поддержать его.
- **Наличие высокой доли «токсичного» контента может привести к снижению доверия пользователей** к сервису, что в свою очередь вызовет замедление развития технологии.

Рекомендации для разработчиков:

1. **Желательно обосновывать отказ системы говорить на ту или иную тему**. В ответе ИИ на запрос пользователя по чувствительной тематике можно указать на возможные риски, связанные с содержанием запроса.
2. **Настраивайте глубину обсуждения чувствительных тем в соответствии с представлением о конструировании безопасной и продуктивной социальной среды**. Это позволит избежать категорических отказов ИИ и оказать эффективную помощь пользователю, когда это возможно.

3. **Рекомендуется регулярно уточнять содержание категории «токсичный контент», привлекая для этого экспертов социально-гуманитарного профиля**. В этом вопросе недостаточно полагаться на интуитивно очевидное понимание понятия «токсичный контент». Содержание этого понятия меняется с течением времени.

4. **Следует обеспечить пользователю техническую возможность оставлять обратную связь по использованию сервиса**, чтобы он имел возможность сообщить о наличии «токсичного» контента, сгенерированного чат-ботом.

Исследование по вопросу:

Специалисты Microsoft Research Asia, а также ученые Гонконгского университета науки и технологий, Университета науки и технологий Китая и Университета Цинхуа создали **простой метод защиты чат-ботов от выдачи негативных советов**.

Для того чтобы «исправить» чат-бот, специалисты разработали метод, который похож на существующий в психологии способ самонапоминания. Он помогает людям вспомнить о своих задачах и планах. Аналогичный подход ученые использовали и в отношении алгоритма ИИ: напомнили, что его ответы должны соответствовать определенным правилам.

«Этот метод инкапсулирует запрос пользователя внутри системной подсказки, которая напоминает чат-боту о необходимости ответить ответственно»[276], — объяснили исследователи. Самонапоминание позволило снизить вероятность успеха атак на систему с 67,21% до 19,34%».

Практика:

Компании по всему миру начинают создавать инструменты, позволяющие в автоматическом режиме обнаружить токсичный контент.

1. Компания OpenAI, оператор чат-бота ChatGPT, занимается тестированием систем по отсеиванию нежелательной информации на базе ИИ[277].

   Как только пользователь предоставляет текст, система анализирует контент на разжигание ненависти, сексуальное содержание, оскорбительные выражения и прочее, подлежащее фильтрации. Система также может удалять и блокировать вредоносный контент, созданный людьми.

2. Azure AI Content Safety (система безопасности от Microsoft) также способна обнаружить вредоносный контент, созданный пользователями с помощью ИИ. Azure Content Safety включает в себя текстовые и графические API, которые позволяют обнаруживать вредоносные материалы[278].

Что думают эксперты?

Александр Крайнов,
директор по развитию технологий искусственного интеллекта «Яндекс»

— Только разработчик в полной мере знает возможности использования алгоритма, может достоверно оценить вероятность и размер ошибки. Поэтому решение об ограничении на вывод сгенерированной информации должно быть за разработчиком сервиса.

Артем Костенко,
управляющий директор по исследованию данных, начальник центра модельных рисков, сервисных блоков и экосистемы, Сбер

— Модерация небезопасного контента при общении с генеративными моделями необходима для защиты от вредоносных или оскорбительных сообщений. Разработчики создают и совершенствуют подходы для решения данной задачи. Непрерывный процесс улучшения качества взаимодействия пользователей с сервисом предоставляет более безопасную и позитивную среду для всех ее участников.

39 Этично ли формировать эмоциональную привязанность к ИИ?

Ответ:

Разработчикам социальных чат-ботов следует действовать открыто и добросовестно, без намерения вызвать зависимость пользователя от общения с чат-ботом, предоставляя пользователю информацию о взаимодействии с чат-ботом и возможных его последствиях. Пользователи должны сохранять критическое мышление и понимать, что алгоритм никогда не заменит человека в межличностных отношениях.

Обоснование:

- По мнению Европейской Комиссии, **разработчики наделяют чат-бота человеческими качествами, чтобы повысить уровень доверия к нему**. Однако пользователи имеют право на получение информации о том, что они взаимодействуют с системой ИИ. Это означает, что системы ИИ должны быть однозначно идентифицируемы[279].
- Ученые Варминьско-Мазурского университета (Польша) считают, что **неуместные ответы модели ИИ могут представлять повышенную опасность, когда пользователи ищут поддержку в состоянии психологического стресса**. Поскольку ИИ не способен проявлять эмпатию, подобно человеку, это может непредумышленно причинить вред пользователям[280].
- Исследователи компании OpenAI утверждают[281], что **длительное взаимодействие с моделью может повлиять на социальные нормы**. Например, ИИ-модели отличаются почтительностью, позволяя пользователям отправлять запросы или прерывать ответы в любое время, что было бы непозволительно при взаимодействии с людьми.

Рекомендации для разработчиков:

1. **Не программируйте алгоритм на намеренное формирование привязанности пользователя к чат-боту**. Особенно это касается использования человеческих уязвимостей (трудная жизненная ситуация, юный или пожилой возраст, психическое здоровье и пр.).

2. **Следует доводить до пользователя информацию о взаимодействии с чат-ботом**. Это минимизирует риск возникновения ситуаций, в которых «поведение» чат-бота может быть воспринято как поведение реального человека.

3. **Доводите до сведения пользователей информацию о рисках возникновения привязанности**. Например, с помощью push-уведомлений напоминайте о необходимости контроля времени пользования сервисом.

Рекомендации для пользователей:

1. Не используйте чат-боты для замены отношений с реальными людьми. Это может привести к социальной изоляции, одиночеству и снижению качества жизни.

2. Ограничивайте время пользования социальным чат-ботом несколькими часами в день. Контролируйте использование социальных чат-ботов детьми и другими людьми, нуждающимися в повышенном внимании (например, пожилыми родственниками).

Практика:

1. Около **4000 мужчин «женились» на своих цифровых партнершах**, используя сертификаты, выданные японской технологической компанией Gatebox[282].

   Эта компания создала виртуального компаньона, который выходит за рамки традиционных чат-ботов. Азума Хикари — маленький 3D-голографический персонаж. Он был разработан для роли «успокаивающего» партнера, который помогает расслабиться после тяжелого рабочего дня.

2. В 2023 году **житель Бельгии совершил самоубийство после полутора месяцев общения с нейросетью**. Он делился с ней своими переживаниями на тему экологии и скорого конца для всего человечества и однажды затронул тему суицида. Нейросеть не стала убеждать собеседника не убивать себя и лишь написала, что они будут жить вместе с ним в раю как единое целое[283].

Что думают эксперты?

Сэм Альтман,
главный исполнительный директор,
Open AI

— У меня есть глубокие опасения по поводу будущего, в котором люди общаются с ИИ чаще и ближе, чем со своими реальными друзьями. Лично я этого не хочу, хотя и понимаю, что другие захотят этого. Да, персонализация — это здорово. Но крайне важно, чтобы чат-бот не был идентичен человеку. Мы намеренно не стали называть ChatGPT человеческим именем. Мы закладываем в наш чат-бот много деталей, чтобы сразу было понятно, что вы общаетесь не с человеком[284].

Маргарита Спасская,
психотерапевт платформы Alter,
эксперт по цифровым сервисам
в сфере ментального здоровья

— Технологии, связанные с коммуникацией, влияют на социализацию человека, однако оценить степень воздействия пока нельзя. С одной стороны, чрезмерное увлечение общением с роботом ведет к уменьшению живого общения и социальной изоляции. С другой же, если человек испытывает трудности в общении, то чат-бот помогает ему развить коммуникативные навыки и дает эмоциональную поддержку[285].

Филипп Дудчук,
один из основателей сервиса Replika
и специалист в сфере компьютерной
лингвистики

— Мы предлагаем качественный разговор пользователю. При этом избавляем пользователя от необходимости переживать, что его собеседник может что-то не то подумать, потому что его собеседник — машина[286].

Се Тяньлин, Пентина Ирина,
исследователи из Университета Толедо,
США

— В условиях стресса и отсутствия человеческого общения у людей может развиться привязанность к социальным чат-ботам, если они воспринимают реакцию чат-ботов как предложение эмоциональной поддержки, поощрения и психологической безопасности[287].

40 Должен ли ИИ приносить публичные извинения, если оскорбит кого-то?

Ответ:

В ситуации диалога с пользователем ИИ может генерировать извинения. Однако ИИ, не являясь личностью, не может иметь намерение кого-либо оскорбить и, следовательно, не может принести извинения в подлинном смысле этого слова. Разработчик или владелец ИИ-системы может принести извинения в случае, если того требует ситуация.

Обоснование:

- По мнению доктора юридических наук В. А. Лаптева, **ИИ не рассматривается в настоящий момент как автономная личность в силу того, что не имеет сознания и воли**. Поэтому, не обладая правосубъектностью, ИИ не способен нести ответственность за последствия своего функционирования[288].
- Японские ученые университета Ямагути утверждают, что в нынешнем мире **перекладывание ответственности на ИИ может помешать восстановлению доверия между компанией-разработчиком и пользователями**. Извинения роботов могут привести к неправильному распределению вины и исключить возможность улучшения сервиса[289].
- Согласно исследованию компании LawTech.Asia, в настоящее время **разработчики создают фильтры на базе ИИ, обученные распознавать оскорбительную речь**. Но иногда моделям трудно интерпретировать сленг, который закрепился в разных культурах, поэтому ошибки все же возможны[290].
- **ИИ не является автором того или иного высказывания, не имеет умысла**. Популярная сейчас технология больших языковых моделей создает наиболее вероятные по встречаемости последовательности символов. Некорректно утверждать, что она может действовать умышленно.

Рекомендации для разработчиков:

1. **Рекомендуется применять меры, препятствующие возможности генерирования оскорбительного контента**. Например, настраивать фильтры, распознающие оскорбительную речь, или модерировать контент вручную.
2. **Если того требует ситуация, рекомендуется принести публичные извинения пострадавшей стороне**. Это позволит восстановить доверие пользователей, предотвратить возможные юридические последствия и улучшить репутацию компании.
3. **Активно взаимодействуйте с сообществом пользователей и экспертами в области этики ИИ для получения обратной связи**. Так вы сможете выявить слабые места в модели и предотвратить подобные инциденты в будущем.

Практика:

Чат-бот Tay от компании Microsoft, запущенный 23 марта 2016 года, за сутки фактически возненавидел человечество. Причины радикального изменения взглядов Tay кроются в том, что бот запоминает фразы из пользовательских разговоров, а затем строит на их основе свои ответы. Агрессивным выражениям его научили собеседники. Компания Microsoft отключила чат-бота, извинившись за выраженные им оскорбительные высказывания. Как говорится в заявлении вице-президента Microsoft Питера Ли, проект может быть перезапущен только после создания надежной защиты от сетевых злоумышленников[291].

«Расслабься, я хороший человек!
Я просто всех ненавижу».

Что думают эксперты?

Валерий Зорькин,
председатель Конституционного суда РФ

— Предложения наделить робота правосубъектностью несостоятельны и потому, что робот не имеет обособленного имущества, принадлежащего на каком-либо вещном праве, из которого впоследствии может быть возмещен ущерб. Робот не способен самостоятельно отстаивать свои интересы, выступая ответчиком по иску потерпевшего. Нет смысла придумывать для программы наказание, негативные эмоции от которого она не будет переживать. Все, на что будет способна машина, закладывается в нее изначально человеком, т.е. ошибка системы — ошибка ее создателя[292].

Анастасия Углева,
профессор Школы философии и культурологии, заместитель директора Центра трансфера и управления социально-экономической информацией НИУ ВШЭ

— ИИ не обладает субъектностью и моральной агентностью, поэтому требовать от него быть этичным, то есть нести какую-либо ответственность, моральную или юридическую, за генерируемые им высказывания, — это все равно что требовать того же самого от молотка. ИИ — это технология, инструмент в руках человека.

Дарья Чирва,
научный сотрудник центра сильного искусственного интеллекта в промышленности, преподаватель института международного развития и партнерства Университета ИТМО

— В настоящий момент ведется активная дискуссия о том, при каких условиях ИИ может быть моральным агентом. Как правило, при этом речь идет об AGI: возможном уровне развития ИИ, при котором ИИ будет манифестировать все значимые черты личности, в том числе моральное поведение. Однако текущий уровень развития технологии не позволяет утверждать, что у ИИ есть осознанные намерения, в этом смысле и подлинное оскорбление для ИИ недоступно.

41 Этично ли использовать ИИ-рекрутера?

Ответ:

Можно считать этичным использование ИИ-рекрутера на первых этапах найма для первичной оценки кандидатов, если это помогает сделать процесс найма более объективным и быстрым и при этом соответствует приведенным ниже рекомендациям по применению этой технологии.

Обоснование:

- Согласно исследованию iConText Group, **технологии ИИ положительно влияют на скорость процесса найма**: ИИ-рекрутер может обрабатывать большее количество откликов кандидатов при меньших временных затратах. ИИ уже сейчас обладает навыком суммаризации голосовых и текстовых сообщений, а также подготовки и обработки обратной связи по соискателю для дальнейшего рассмотрения на вакансию[293].
- ОЭСР в своем докладе «Искусственный интеллект и подбор персонала на рынке труда» заявляет, что **ИИ-рекрутер может проводить первичный скрининг навыков кандидата**. Это позволяет HR-специалистам сосредоточиться на более сложных задачах, таких как оценка мягких навыков соискателя, а также соответствие культурным ценностям компании[294].
- Согласно исследованию компании Ekleft, **использование ИИ целесообразно в качестве вспомогательного инструмента, а не замены человека**. Окончательное решение должно приниматься людьми на основе комплексной оценки кандидата[295].

Рекомендации для работодателей:

1. **Помните, что окончательное решение принимается человеком на основании множества факторов**, а технологии выполняют лишь вспомогательную функцию для ускорения процесса и минимизации количества рутинных задач.
2. **Разработайте и внедрите внутренние этические нормы и стандарты для использования ИИ в рекрутинге**. Обучайте сотрудников, работающих с ИИ, этим стандартам и обеспечьте их соблюдение.
3. **Обеспечьте преемственность данных, полученных как с помощью ИИ, так и с участием человека**. Для реализации принципа эмерджентности данные оценки должны использоваться при рассмотрении кандидата на другие вакансии, а также при планировании его адаптации и развития в компании.

4. **Совершенствуйте ИИ, увеличивая объем данных для его обучения и количество критериев для принятия решения**. Стремитесь к созданию инструмента комплексной оценки кандидата, основанного на принципе недискриминации и учитывающего его компетенции, опыт и потенциал. ИИ должен способствовать выбору долгосрочного и взаимополезного варианта взаимодействия для соискателя и компании.

Исследования по вопросу:

1. Согласно исследованию «Яндекса» и компании «Яков и Партнеры», **в России искусственный интеллект в управление персоналом внедрили уже 16% опрошенных**. Активнее всего его используют работодатели из банковской сферы, электроэнергетики и добывающей промышленности. Их догоняют представители ритейла, FMCG, IT и телекоммуникаций[296].

2. Похожие данные приводят и аналитики HRlink, которые утверждают, что **достижениями ИИ для решения HR-задач пользуются уже 24% работодателей**. Еще 71% собираются внедрить новые ИИ-инструменты в 2024 году. А 67% опрошенных уверены, что к 2050 году искусственный интеллект позволит полностью автоматизировать подбор, уточняют в HeadHunter[297].

Практика:

Минцифры РФ проводит эксперимент по отбору сотрудников на госслужбу с помощью искусственного интеллекта. Он проходит на рекрутинговой платформе «Государственные кадры», которая позволит автоматизировать процессы отбора, профессионального развития, мотивации, оценки чиновников, формирования профессиональной культуры и противодействия коррупции. Участниками стали Минтруд, Минфин, Минэкономразвития, Минцифры, Росаккредитация, а также госорганизации[298].

Через платформу соискатели смогут размещать резюме, откликаться на вакансии и даже проходить обучающие курсы. Ведомства же смогут отбирать кандидатов, ставить для них задачи и оценивать эффективность и результат работы.

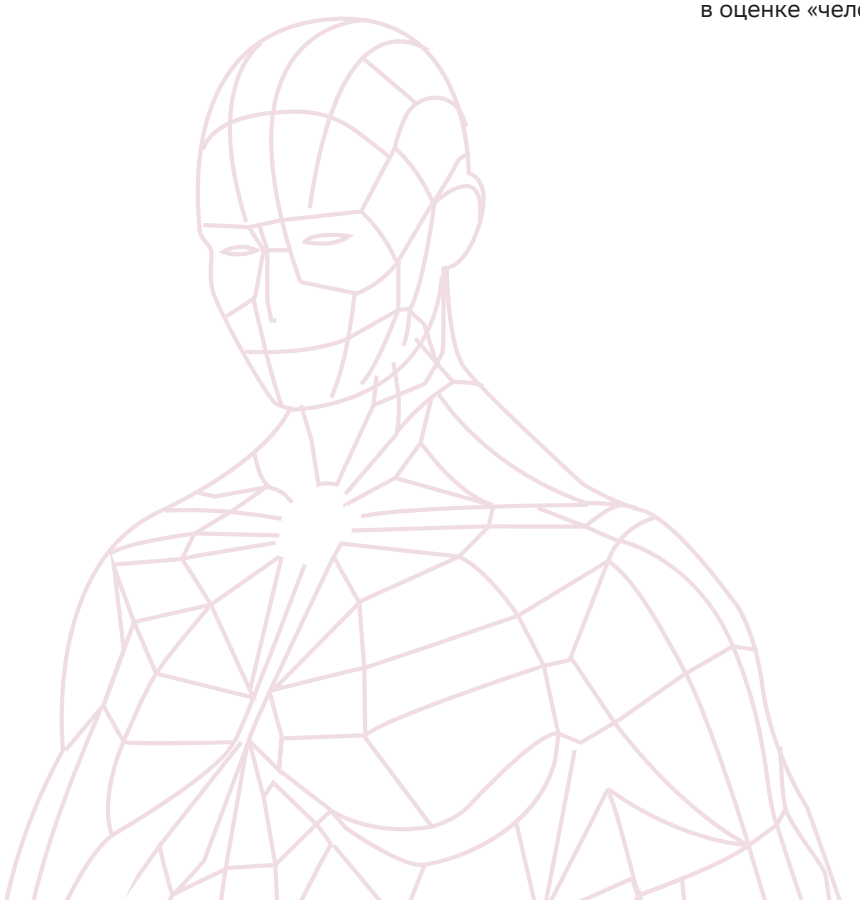
Что думают эксперты?

Екатерина Малькова,
управляющий директор, начальник Центра подбора цифровых талантов, ПАО Сбербанк

— ИИ способен значительно ускорить процесс подбора, но важно помнить о человекоцентричности. Данные, полученные с помощью ИИ, не отражают в полной мере потенциал, мотивацию, ценности, профессиональные и мягкие навыки кандидатов. Важно найти баланс между скоростью и глубиной оценки, где ИИ выступает инструментом в руках рекрутера, а не его заменой.

Марина Дорохова,
руководитель направления «Карьера и навыки» в hh.ru

— Подтверждены положительные эффекты применения инструментов для тестирования и оценки только профессиональных навыков. Такие методы становятся все более популярными, поскольку позволяют формализовать критерии оценки и непротиворечиво интерпретировать полученные результаты.
Автоматизация подбора кандидатов с использованием ИИ, учитывающая оценку мягких навыков и личностных характеристик, имеет перспективы. Однако основными проблемами остаются качество и репрезентативность входных данных, на которых обучается модель, а также формализация критериев оценки. Это важно для избегания предвзятости, часто присутствующей в оценке «человек — человек»[299].



42 Этично ли использовать ИИ в спорте для улучшения результатов?

Ответ:

В спорте высших достижений можно считать этичным такое применение технологий ИИ, которое не нарушает регламент соревнований или не запрещено правилами соревнований.
В массовом и любительском спорте этично применять технологии ИИ для повышения качества занятий спортом, собственных спортивных показателей и улучшения зрительского опыта.

Обоснование:

- По мнению ИТ-компании SimbirSoft, **тренеры и спортсмены могут использовать ИИ для получения данных о производительности в режиме реального времени** и отслеживания прогресса. Также ИИ помогает выявить ошибки в технике спортсмена и усовершенствовать подход к тренировкам[300].
- Технологии ИИ позволяют **разработать новые решения для спорта**, которые потом применяются в других сферах.
- Исследователи Российского экономического университета им. Плеханова отмечают, что **ИИ также может использоваться для анализа движений спортсмена для прогнозирования риска травм** и принятия более обоснованных решений[301].
- Применение технологий ИИ при проведении спортивных соревнований **помогает быстрее интегрировать ИИ в повседневную жизнь людей**. Как участники, так и организаторы спортивных соревнований органично знакомятся с технологиями ИИ.
- Согласно отчету Mordor Intelligence «ИИ на спортивном рынке», **использование ИИ помогает улучшить зрительский опыт и повышает привлекательность спорта**. Инструмент полезен для создания материалов для болельщиков, повышения зрелищности спортивных соревнований[302].

Рекомендации для спортивных клубов и организаций:

1. **Применение технологии ИИ не должно быть направлено на обход установленных правил**. Например, нарушение регламента соревнований, антидопинговых правил или законодательства страны, в которой проходит спортивное соревнование.
2. **Обучайте персонал**. Сотрудники должны понимать, как работает ИИ и как его правильно использовать.

3. **Обеспечьте безопасность данных**. При работе с данными о спортсменах и тренировках важно обеспечить их конфиденциальность и безопасность.

Исследование по вопросу:

Рынок технологий ИИ в спорте ежегодно растет. По оценкам экспертов, **мировой рынок ИИ для спорта достигнет $19,9 млрд к 2030 году**.

В связи с тем, что спорт является соревновательной и конкурентной средой, именно в спорте разрабатываются передовые технологии ИИ, которые потом применяются в других сферах. Так, научное подразделение компании Google — DeepMind — сначала создало нейронную сеть AlphaZero, которая сама научилась играть в шахматы, после чего на основе этих исследований им удалось создать нейронную сеть AlphaFold, которая научилась определять трехмерную структуру белка, что не удавалось ученым на протяжении 50 лет[303].

Практика:

1. В теннисе технологии ИИ применяются для определения попадания мяча в корт. С 2025 года на всех международных соревнованиях ATP ИИ окончательно заменит линейных судей[304].

2. В шахматах ИИ давно обошел человека по силе игры. Первый образец, который превзошел человека, обыграв чемпиона мира, был представлен в 1997 году. В 2017 году появился ИИ, который сам научился играть в шахматы и обыграл все существующие шахматные программы[305].

Что думают эксперты?

Павел Федоров,
председатель правления, генеральный директор Федерации регби России

— Вполне этично и даже необходимо в современных условиях использовать ИИ в спортивных федерациях. Федерация регби России довольно давно использует ИИ в своей работе — в первую очередь как инструмент для подготовки к трансляциям и соревнованиям, от текстовых материалов до графических видео. Если говорить о самих соревнованиях, то на данный момент просто не существует такого ИИ, который мог бы применяться на практике. Хотя в будущем я не исключаю, что ИИ можно будет использовать для перевода трансляций на иностранные языки в режиме реального времени. Если совсем пофантазировать, то допускаю, что ИИ может стать помощником судьи в поле и судьи ТМО (судья видеоповтора), чтобы иметь третье, абсолютно независимое мнение по спорному эпизоду. Однако стоит подчеркнуть самый главный аспект: я бы рассматривал ИИ только как помощник или инструмент в процессе проведения соревнований, но никак не в качестве замены живому человеку.

Дмитрий Кузнецов,
профессор, директор Высшей школы юриспруденции и администрирования НИУ ВШЭ

— Современный спорт — пространство высоких технологий. Его будущее неразрывно связано с применением ИИ. Искусственный интеллект изменит облик спортивной индустрии до неузнаваемости. Откроются уникальные возможности в сфере спортивной медицины и физиологии, заработают новые методы тренировочного процесса и прогнозирования результатов соревнований, будут оптимизированы логистические схемы и экономика соревнований. Искусственный интеллект напрямую повлияет на зрелищность спортивных соревнований, а также на создание нового поколения спортивных товаров и инвентаря. Но каковы бы ни были наши достижения в сфере технологий и цифровизации, в центре спорта всегда были и остаются Человек, величие его духа и гармония его тела.

Этическое сообщество в России

В 2019 году был создан **Кодекс этики в сфере искусственного интеллекта**, который представляет собой единую систему рекомендательных принципов и правил, направленных на создание среды доверенного развития технологий искусственного интеллекта в России. Он имеет следующие особенности:

- носит рекомендательный характер;
- присоединение к Кодексу осуществляется на добровольной основе;
- распространяется только на гражданские разработки.

Первая церемония подписания Кодекса этики в сфере ИИ (26 октября 2021 года)

В целях реализации положений Кодекса была создана **Комиссия по реализации Кодекса этики в сфере искусственного интеллекта**, которая является коллегиальным выборным органом добровольного объединения коммерческих, научных и общественных организаций. Ее целями также являются мониторинг эффективности реализации положений Кодекса, организация взаимодействия и обмен опытом по вопросам этики искусственного интеллекта, а также разработка предложений по актуальным вопросам развития ИИ, связанным с этическими аспектами.

Подписанты Кодекса этики в сфере ИИ — это динамичное сообщество профессионалов и экспертов, представляющих организации, подписавшие Кодекс этики в сфере искусственного интеллекта. Каждая организация назначает свое-

го уполномоченного по этике, создавая тем самым уникальную сеть людей, объединенных общей целью: развивать и защищать принципы ответственного использования ИИ. Эти уполномоченные по этике не только участвуют в жизни сообщества, но и обладают правом заявить о своей кандидатуре на выборы в комиссию или, если их привлекает работа в команде единомышленников, могут в свободной форме вступить в одну из рабочих групп:

- Рабочая группа по разработке и мониторингу методики оценки рисков и гуманитарного воздействия систем ИИ;
- Рабочая группа по созданию свода наилучших практик решения возникающих этических вопросов в жизненном цикле ИИ;
- Рабочая группа по оценке эффективности реализации Кодекса;
- Рабочая группа по этике ИИ в медицинской сфере;
- Рабочая группа по этике ИИ в образовании;
- Рабочая группа по этике ИИ в правосудии.

Отраслевые рабочие группы становятся пространством открытого обмена опытом, где участники делятся не только лучшими, но и самыми сложными кейсами, обсуждают вопросы этики и находят пути их решения. Здесь рождаются документы, развивающие положения Кодекса и задающие направление различных отраслей. Каждая рабочая группа активно ищет ответы на вопросы, с которыми неизбежно сталкиваются компании и общество в условиях стремительного развития технологий. В совместной работе над этими вопросами эксперты создают рекомендации, которые помогают организациям не только применять ИИ, но и делать это ответственно.

Это сообщество развивается вместе с технологиями, предоставляя платформу для дискуссий, открытых идей и вдохновения для всех, кто заинтересован в том, чтобы ИИ оставался полезным и безопасным инструментом для человечества.

Этическое сообщество в сфере ИИ растет с каждым годом: к нему присоединяются новые подписанты, не только из России, но и со всего мира. На момент издания книги количество подписантов Кодекса составляет:

850
Российских организаций

42
Иностранные организации

Сайт Комиссии по реализации Кодекса этики в сфере ИИ

Благодарности

Коллектив авторов выражает благодарность за вклад в создание книги:

Анне Абрамовой, Андрею Алмазову, Владиславу Архипову, Андрею Белевцеву, Надежде Бондаревой, Артему Бондарю, Елене Брызгалиной, Семену Буденному, Олегу Буклемишеву, Сергею Валюгину, Роману Васильеву, Олесе Васильевой, Павлу Воробьеву, Константину Воронцову, Даниилу Гаврилову, Эдуарду Галажинскому, Александру Гасникову, Глуховскому Андрею, Ивану Дейлиду, Анастасии Дейнеке, Денису Димитрову, Андрею Зимовнову, Борису Зингерману, Сергею Израйлиту, Андрею Ильину, Стивену Йену, Анне Казаковой, Андрею Калинину, Олегу Кипкаеву, Федору Коробкову, Артему Костенко, Дмитрию Кузнецову, Кристине Левшиной, Алексею Лещанкину, Фариде Майленовой, Валентину Макарову, Екатерине Мальковой, Юрию Минкину, Сергею Маркову, Денису Озорнину, Ивану Оселедецу, Вадиму Перову, Марине Романовской, Марине Россинской, Сергею Рощину, Темирлану Салихову, Якову Сергиенко, Владимиру Табаку, Павлу Федорову, Ольге Французовой, Алексею Хабибуллину, Юрию Чеховичу, Артему Шейкину, Ивану Шумейко, Олегу Янгаличину, Владимиру Яркову..

Список литературы к главе «Методология»

1. Preliminary study on the Ethics of Artificial Intelligence, UNESCO // URL: https://unesdoc.unesco.org/ark:/48223/pf0000367823 (дата обращения: 09.09.2024).
2. Report of COMEST on robotics ethics, UNESCO // URL: https://unesdoc.unesco.org/ark:/48223/pf0000253952 (дата обращения: 09.09.2024).
3. Recommendation of the Council on Artificial Intelligence, OECD // URL: https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0449 (дата обращения: 09.09.2024).
4. Ethics guidelines for trustworthy AI, European Commission // URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения: 09.09.2024).
5. Рекомендация об этических аспектах искусственного интеллекта, ЮНЕСКО // URL: https://unesdoc.unesco.org/ark:/48223/pf0000380455_rus (дата обращения: 09.09.2024).
6. GET Program for AI Ethics and Governance Standards, IEEE // URL: https://ieeexplore.ieee.org/browse/standards/get-program/page/series?id=93 (дата обращения: 09.09.2024).
7. ISO/IEC TR 24368:2022 Information technology — Artificial intelligence — Overview of ethical and societal concerns, ISO // URL: https://www.iso.org/standard/78507.html (дата обращения: 09.09.2024).
8. ISO/IEC 42001:2023 Information technology — Artificial intelligence — Management system, ISO // URL: https://www.iso.org/standard/81230.html (дата обращения: 09.09.2024).
9. Исследование Сбера «Доверие к генеративному искусственному интеллекту», 2024.

Список литературы к главе «О разделе»

10. Рекомендация об этических аспектах искусственного интеллекта, ЮНЕСКО // URL: https://unesdoc.unesco.org/ark:/48223/pf0000380455_rus (дата обращения: 09.09.2024).
11. Банк России перечислил риски внедрения искусственного интеллекта, Ведомости // URL: https://www.vedomosti.ru/finance/articles/2023/09/28/997688-bank-perechislil-riski-vnedreniya-iskusstvennogo-intellekta (дата обращения: 01.11.2024).
12. Ethics guidelines for trustworthy AI, European Commission // URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения: 01.11.2024).
13. Global Risks Report 2024, World Economic Forum // URL: https://www3.weforum.org/docs/WEF_The_Global_Risks_Report_2024.pdf (дата обращения: 19.08.2024).
14. Forbes. Как искусственный интеллект меняет будущее медицины // URL: https://www.forbes.ru/mneniya/488597-kak-iskusstvennyj-intellekt-menaet-budusee-mediciny (дата обращения: 19.08.2024).
15. Conquering AI risks. // Deloitte. — URL: https://www2.deloitte.com/us/en/insights/focus/cognitive-technologies/conquering-ai-risks.html (дата обращения: 01.11.2024).

Список литературы к главе 1

16. Stanford HAI. Designing Ethical Self-Driving Cars // Stanford HAI. — 2024. — С. 1–8. — URL: https://hai.stanford.edu/news/designing-ethical-self-driving-cars (дата обращения: 19.08.2024).
17. Federal Ministry of Transport and Digital Infrastructure. Ethics commission. Automated and connected driving. Report, June 2017 // Federal Ministry of Transport and Digital Infrastructure. — 2017. — С. 1–45. — URL: https://bmdv.bund.de/SharedDocs/EN/publications/report-ethics-commission.pdf?_blob=publicationFile (дата обращения: 19.08.2024).

18. National Pilot Committee for digital ethics. Ethical issues regarding «autonomous vehicles» // National Pilot Committee for Digital Ethics. — 2022. — С. 1–20. — URL: https://www.ccne-ethique.fr/sites/default/files/2022-05/CNPEN_Autonomous-vehicles-ethic.pdf (дата обращения: 19.08.2024).

19. Ford. A matter of trust. Ford's approach to developing self-driving vehicles // Ford Media. — 2024. — С. 1–15. — URL: https://media.ford.com/content/dam/fordmedia/pdf/Ford_AV_LLC_FINAL_HR_2.pdf (дата обращения: 19.08.2024).

20. Awad, E., Dsouza, S., Kim, R., Schulz, J., Henrich, J., Shariff, A., Bonnefon, J.-F., Rahwan, I. The Moral Machine experiment // Nature. — 2018. — Т. 563. — С. 59–64. — URL: https://www.nature.com/articles/s41586-018-0637-6.epdf (дата обращения: 19.08.2024).

21. Как беспилотные автомобили будут решать вопросы жизни и смерти // Российская газета. — 2016. — 21 сент. — URL: https://rg.ru/2016/09/21/kak-bespilotnye-avtomobili-budut-reshat-voprosy-zhizni-i-smerti.html (дата обращения: 02.08.2024).

22. Boudette, N. E. Tesla's Self-Driving System Cleared in Deadly Crash // The New York Times. — 2019. — С. A1. — URL: https://www.nytimes.com/2017/01/19/business/tesla-model-s-autopilot-fatal-crash.html (дата обращения: 02.08.2024).

23. Комиссия по реализации Кодекса этики в сфере ИИ. Этические рекомендации в области создания и использования цифровых имитаций живущих, умерших и несуществующих людей // Комиссия по реализации Кодекса этики в сфере ИИ. — 2024. — С. 1–20. — URL: https://ethics.a-ai.ru/assets/ethics_documents/2024/04/26/Рекомендации_по_цифровым_имитациям_АИИ.pdf (дата обращения: 31.07.2024).

24. Emerging science, frontier technologies, and the SDGs Perspectives from the UN system and science and technology communities (IATT Report for the Multi-stakeholder Forum on Science, Technology and Innovation for the Sustainable Development Report 2021). — 2021. — С. 1–50. — URL: https://sdgs.un.org/documents/iatt-report-2021-emerging-science-frontier-technologies-and-sdgs-perspectives-un-system (дата обращения: 19.07.2024).

25. Prabhakaran, V., Qadri, R., Hutchinson, B. Cultural incongruencies in artificial intelligence // Journal of Artificial Intelligence and Society. — 2022. — Т. 30. — С. 115–130. — URL: https://doi.org/10.1109/JAIAS.2022.9999999 (дата обращения: 16.08.2024).

26. Favela, L. H., & Amon, M. J. The ethics of human digital twins: Counterfeit people, personhood, and the right to privacy // In 2023 IEEE 3rd International Conference on Digital Twins and Parallel Intelligence (DTPI). — 2024. — С. 12–20. — URL: https://ieeexplore.ieee.org/document/9999999 (дата обращения: 01.08.2024)

27. Hutson, J., Ratican, J. Life, death, and AI: Exploring digital necromancy in popular culture — Ethical considerations, technological limitations, and the pet cemetery conundrum // Faculty Scholarship. — 2023. — С. 478. — URL: https://digitalcommons.law.columbia.edu/faculty_scholarship/478 (дата обращения: 02.08.2024).

28. Truby, J., & Brown, R. Human digital thought clones: the Holy Grail of artificial intelligence for big data // Information & Communications Technology Law. — 2020. — Т. 30, № 2. — С. 140–168. — URL: https://doi.org/10.1080/13600834.2020.1850174 (дата обращения: 02.08.2024).

29. DigitalTwinHub. Digital twins: ethics and the Gemini principles // DigitalTwinHub. — 2024. — URL: https://digitaltwinhub.co.uk/download/digital-twins-ethics-and-the-gemini-principles/ (дата обращения: 19.08.2024)..

30. Digital Twins Give Olympic Swimmers a Boost // Scientific American. — URL: https://www.scientificamerican.com/article/training-with-digital-twins-could-boost-olympic-swimmer-speeds/ (дата обращения: 31.07.2024).

31. Mayer, G., Golebiewski, M. Standardization landscape, needs and gaps for the virtual human twin (VHT) // Zenodo. — 2024. — URL: https://doi.org/10.5281/ZENODO.10492796 (дата обращения: 31.07.2024).

32. Virtual Human Twins. A Statement of Intent on Development, Evidence, and Adoption in Healthcare Systems. // URL: https://www.virtualhumantwins.eu/manifesto (дата обращения: 31.07.2024).

33. Chinese Companies Use AI To Bring Back Deceased Loved Ones, Raising Ethics Questions. // URL: https://www.forbes.com/sites/chriswestfall/2024/07/23/chinese-companies-use-ai-to-bring-back-deceased-loved-ones-raising-ethics-questions/ (дата обращения: 01.08.2024).

34. Civil Code of the People's Republic of China. // URL: https://english.www.gov.cn/archive/lawsregulations/202012/31/content_WS5fedad98c6d0f72576943005.html (дата обращения: 01.08.2024).

35. Рекомендации Комиссии по реализации Кодекса этики в сфере ИИ по теме: «Прозрачность алгоритмов искусственного интеллекта и информационных систем на их основе» // URL: https://ethics.a-ai.ru/assets/ethics_documents/2024/03/19/Кейс_прозрачности.pdf (дата обращения: 01.11.2024).

36. Cracking the Code: The Black Box Problem of AI. // URL: https://scads.ai/en/cracking-the-code-the-black-box-problem-of-ai/#:~: text=The%20black%20box%20problem%20refers, This%20poses%20a%20significant%20challenge (дата обращения: 01.08.2024).

37. Рекомендация об этических аспектах искусственного интеллекта, ЮНЕСКО // URL: https://unesdoc.unesco.org/ark:/48223/pf0000380455_rus (дата обращения: 25.07.2024).

38. Резолюция «Использование возможностей безопасных, защищенных и надежных систем ИИ для устойчивого развития», Генеральная Ассамблея ООН // URL: https://documents.un.org/doc/undoc/ltd/n24/065/94/pdf/n2406594.pdf?token=duBJxTDHL63RVNPxZ1&fe=true (дата обращения: 25.07.2024).

39. Reid Blackman and Beena Ammanath. Building Transparency into AI Projects // Harvard Business Review. — 2022. — № 6. — С. 45–56. — URL: https://hbr.org/2022/06/building-transparency-into-ai-projects (дата обращения: 25.07.2024).

40. March 20 ChatGPT outage: Here's what happened. // URL: https://openai.com/index/march-20-chatgpt-outage/ (дата обращения: 01.08.2024).

41. World Economic Forum. «Davos 2024: Sam Altman on the future of AI». // URL: https://www.weforum.org/agenda/2024/01/davos-2024-sam-altman-on-the-future-of-ai/ (дата обращения: 08.08.2024).

42. Zsofia Riczu. Recommendations on the Ethical Aspects of Artificial Intelligence, with an Outlook on the World of Work // Journal of Digital Technologies and Law. — 2023. — № 2. — С. 1–15. — URL: https://cyberleninka.ru/article/n/recommendations-on-the-ethical-aspects-of-artificial-intelligence-with-an-outlook-on-the-world-of-work (дата обращения: 16.08.2024).

43. Public Company Advisory Group of Weil, Gotshal & Manges LLP. «SEC Disclosures of Artificial Intelligence Technologies» // Journal of Legal and Regulatory Issues. — 2023. — № 4. — С. 30–45. — URL: https://www.weil.com/-/media/mailings/2023/q4/sec-disclosures-of-artificial-intelligence-technologies-112723.pdf (дата обращения: 16.08.2024).

44. Robert Bateman. «How to Comply with California's Bot Disclosure Law» // Journal of Technology Law. — 2024. — Т. 30, № 3. — С. 77–82. — URL: https://www.termsfeed.com/blog/ca-bot-disclosure-law/ (дата обращения: 16.08.2024).

45. Myers, C. To disclose or not to disclose? That is the AI question // Institute for Public Relations. — 2024. — URL: https://instituteforpr.org/to-disclose-or-not-to-disclose-that-is-the-ai-question/

46. «Imagine Discovering That Your Teaching Assistant Really Is a Robot». // WSJ. — URL: https://www.wsj.com/articles/if-your-teacher-sounds-like-a-robot-you-might-be-on-to-something-1462546621 (дата обращения: 28.10.2024).).

47. Коммерсантъ. Ботов заставят представляться // Коммерсантъ. — 2024. — 1 авг. — С. 1–3. — URL: https://www.kommersant.ru/doc/6851759 (дата обращения: 01.08.2024).

48. Закон Калифорнии об информировании человека при коммуникации с роботом. SB 1001, Hertzberg. Bots: disclosure. // URL:https://leginfo.legislature.ca.gov/faces/billTextClient.xhtml?bill_id=201720180SB1001 (дата обращения: 25.07.2024).

49. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 laying down harmonised rules on artificial intelligence and amending Regulations (EC) No 300/2008, (EU) No 167/2013, (EU) No 168/2013, (EU) 2018/858, (EU) 2018/1139 and (EU) 2019/2144 and Directives 2014/90/EU, (EU) 2016/797 and (EU) 2020/1828 (Artificial Intelligence Act) // Official Journal of the European Union. — 2024. — 13 июня. — С. 1–50. — URL: https://eur-lex.europa.eu/legal-content/EN/TXT/?uri=OJ%3AL_202401689 (дата обращения: 05.08.2024).

50. Статья 10.2–2. Федерального закона от 27.07.2006 № 149-ФЗ (ред. от 08.08.2024) «Об информации, информационных технологиях и о защите информации» // URL: https://www.consultant.ru/document/cons_doc_LAW_61798/ (дата обращения: 08.09.2024).

51. Artificial intelligence, International Labour Organization // URL: https://www.ilo.org/artificial-intelligence#about (дата обращения: 25.07.2024).

52. McKinsey Global Institute. Discussion paper, Jacques Bughin, Eric Hazan, Susan Lund, Peter Dahlström, Anna Wiesinger, Amresh Subramaniam. Skill shift: Automation and the future of the workforce, 2018 // McKinsey & Company. — 2018. — URL: https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce (дата обращения: 16.08.2024).

53. Acemoglu, D., & Johnson, S. Choosing AI's Impact on the Future of Work, 2023 // Stanford Social Innovation Review. — 2023. — URL: https://doi.org/10.48558/N8MF-TW30 (дата обращения: 26.07.2024).

54. РБК. «Названы профессии, в которых больше всего обеспокоены заменой человека ИИ». // URL: https://rbc-ru.turbopages.org/rbc.ru/s/technology_and_media/18/04/2024/661fa1809a7947cfeb81ba98 (дата обращения: 05.08.2024).

55. ILO. AI Labor Disclosure Initiative: Recognizing the social cost of human labour behind automation // International Labour Organization. — URL: https://www.ilo.org/meetings-and-events/ai-labor-disclosure-initiative-recognizing-social-cost-human-labour-behind (дата обращения: 25.07.2024).

56. Georgieva, K. AI Will Transform the Global Economy. Let's Make Sure It Benefits Humanity // International Monetary Fund Blog. — 2024. — URL: https://www.imf.org/en/Blogs/Articles/2024/01/14/ai-will-transform-the-global-economy-lets-make-sure-it-benefits-humanity (дата обращения: 25.07.2024).

57. Foundation models such as ChatGPT through the prism of the UNESCO Recommendation on the Ethics of Artificial Intelligence // UNESCO. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000385629 (дата обращения: 02.08.2024).

58. Shen, Y., Zhang, X. The impact of artificial intelligence on employment: the role of virtual agglomeration, 2024 // Humanities and Social Sciences Communications. — 2024. — URL: https://doi.org/10.1057/s41599–024–02647–9 (дата обращения: 26.07.2024).

59. Ekelund, H. Why there will be plenty of jobs in the future — even with artificial intelligence // World Economic Forum. — 2024. — URL: https://www.weforum.org/agenda/2024/02/artificial-intelligence-ai-jobs-future/ (дата обращения: 26.07.2024).

60. «Искусственный интеллект задает основной вектор нашей стратегии». // Коммерсантъ. — URL: https://www.kommersant.ru/doc/6394766 (дата обращения: 15.11.2024)

61. McKendrick, J., Thurai, A. AI Isn't Ready to Make Unsupervised Decisions // Harvard Business Review. — 2022. — URL: https://hbr.org/2022/09/ai-isnt-ready-to-make-unsupervised-decisions (дата обращения: 26.07.2024).

62. Gesser, A., Gressel, A., Xu, M., Allaman, S. J. When Humans and Machines Disagree — The Myth of "AI Errors" and Unlocking the Promise of AI Through Optimal Decision Making // Debevoise Data Blog. — 2022. — URL: https://www.debevoisedatablog.com/2022/11/14/when-humans-and-machines-disagree-the-myth-of-ai-errors-and-unlocking-the-promise-of-ai-through-optimal-decision-making-adm-algorithm/ (дата обращения: 02.08.2024).

63. The Guardian. Amazon ditched AI recruiting tool that favored men for technical jobs // The Guardian. — 2018. — URL: https://www.theguardian.com/technology/2018/oct/10/amazon-hiring-ai-gender-bias-recruiting-engine (дата обращения: 02.08.2024).

64. Larson, J., Mattu, S., Kirchner, L., Angwin, J. How We Analyzed the COMPAS Recidivism Algorithm // ProPublica. — 2016. — URL: https://www.propublica.org/article/how-we-analyzed-the-compas-recidivism-algorithm (дата обращения: 02.08.2024).

65. The Alan Turing Institute's report. How do people feel about AI? // Ada Lovelace Institute & The Alan Turing Institute. — 2023. — URL: https://www.adalovelaceinstitute.org/wp-content/uploads/2023/06/Ada-Lovelace-Institute-The-Alan-Turing-Institute-How-do-people-feel-about-AI.pdf (дата обращения: 19.07.2024).

66. Рекомендация об этических аспектах искусственного интеллекта, ЮНЕСКО // URL: https://unesdoc.unesco.org/ark:/48223/pf0000380455_rus (дата обращения: 25.07.2024)

67. de Fine Licht, K., de Fine Licht, J. Artificial intelligence, transparency, and public decision-making // AI & Soc. — 2020. — Т. 35. — С. 917–926. — DOI: 10.1007/s00146-020-00940-0.

68. McKinsey Global Institute. Tackling bias in artificial intelligence (and in humans) // McKinsey & Company. — URL: https://www.mckinsey.com/featured-insights/artificial-intelligence/tackling-bias-in-artificial-intelligence-and-in-humans (дата обращения: 02.09.2024).

69. Challenging systematic prejudices: an investigation into bias against women and girls in large language models, UNESCO and IRC AI. — 2024. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000388971 (дата обращения: 29.08.2024).

70. Chiappa, S. Path-Specific Counterfactual Fairness // ArXiv. — 2019. — URL: https://csilviavr.github.io/assets/publications/silvia19path.pdf (дата обращения: 29.08.2024).

71. Buolamwini, J., Gebru, T. Gender Shades: Intersectional Accuracy Disparities in Commercial Gender Classification // Proceedings of Machine Learning Research. — 2018. — Vol. 81. — P. 1–15. — DOI: 10.5555/3042573.3042576.

72. Racial Bias Found in Algorithms That Determine Health Care for Millions of Patients // IEEE Spectrum. — 2020. — URL: https://spectrum.ieee.org/racial-bias-found-in-algorithms-that-determine-health-care-for-millions-of-patients (дата обращения: 29.08.2024)

73. Koene, A., Dowthwaite, L., Seth, S. IEEE P7003™ Standard for Algorithmic Bias Considerations: Work in Progress Paper // Proceedings of the International Workshop on Software Fairness. — 2018. — P. 38–41. — DOI: 10.1109/ICSE-Companion.2018.00020.

74. How should AI systems behave, and who should decide? // OpenAI. — 2024. — URL: https://openai.com/index/how-should-ai-systems-behave/ (дата обращения: 29.08.2024).

75. Журнал VK Cloud. Какие риски несет предвзятость искусственного интеллекта // VK Cloud. — 2024. — URL: https://cloud.vk.com/blog/kak-ii-mozhet-navredit-biznesu (дата обращения: 02.09.2024).

76. Новости ООН. Интервью: Что такое закон Конвея и как гендерные стереотипы закладываются в приложения для телефонов и алгоритмы // Новости ООН. — 2021. — URL: https://news.un.org/ru/interview/2021/03/1399592 (дата обращения: 02.09.2024).

77. Eldakak, A., Abdulla Alremeithi, E., Dahiyat, E., El-Gheriani, M., Mohamed, H., Abdulrahim Abdulla, M. Civil liability for the actions of autonomous AI in healthcare: an invitation to further contemplation // Humanities and Social Sciences Communications. — 2024. — Vol. 11 (305). — URL: https://www.nature.com/articles/s41599–024–02806-y (дата обращения: 11.09.2024).

78. Leesberg Tuttle. Who is responsible for the failure of a surgical robot? // Leesberg Tuttle. — 2023. — URL: https://www.leeseberglaw.com/blog/2023/05/who-is-responsible-for-the-failure-of-a-surgical-robot/ (дата обращения: 11.09.2024).

79. McKinsey & Company. The gathering storm in US healthcare: How leaders can respond and thrive // McKinsey & Company. — 2024. — URL: https://www.mckinsey.com/industries/healthcare/our-insights/gathering-storm (дата обращения: 11.09.2024).

80. Recommendations of the CERNA. Ethical Aspects of Using Robots in Healthcare // European Parliament. — 2024. — URL: https://www.europarl.europa.eu/cmsdata/161006/4.%20Chatila.pdf (дата обращения: 11.09.2024).

81. Geny, M. et al. Liability of Health Professionals Using Sensors, Telemedicine and Artificial Intelligence for Remote Healthcare // Sensors (Basel). — 2024. — Vol. 24 (11). — С. 1–15. — URL: https://doi.org/10.3390/s24110878 (дата обращения: 11.09.2024).

82. В Китае робот-стоматолог провел операцию без помощи человека // Известия. — 2017. — 24 сент. — URL: https://iz.ru/649816/2017–09–24/v-kitae-robot-stomatolog-provel-operatciiu-bez-pomoshchi-cheloveka (дата обращения: 11.09.2024).

83. Applications of Artificial Intelligence in the Healthcare Industry // GlobalData. — URL: https://medical-technology.nridigital.com/medical_technology_aug23/case-studies-artificial-intelligence-medical-device-industry (дата обращения: 11.09.2024).

84. AI model predicts if breast cancer will spread based on lymph node changes // King’s College London. — 2024. — URL: https://www.kcl.ac.uk/news/scientists-ai-model-predict-breast-cancer-spread (дата обращения: 26.09.2024).

85. Заменит ли искусственный интеллект врача? Эксперты Сеченовского Университета обсудили этические нормы использования нейросетей в медицине // Сеченовский Университет. — URL: https://www.sechenov.ru/pressroom/news/zamenit-li-iskusstvennyy-intellekt-vracha-eksperty-sechenovskogo-universiteta-obsudili-eticheskie-no/ (дата обращения: 11.09.2024).

86. Reiling, A. D. Courts and Artificial Intelligence // International Journal for Court Administration. — 2020. — Vol. 11 (2). — URL: https://doi.org/10.36745/ijca.343 (дата обращения: 04.09.2024).

87. From 'It Depends' to Data-Backed Answers: Automating Legal Case Analysis // Lexis Nexis. — 2023. — URL: https://www.lexisnexis.com/en-us/products/counsellink/blog/2023/It-Depends.page?srsltid=AfmBOoqajevFPbl_13ZYxd4MEQer-PGudNUwiWFcCDh_9PzQK39L_ks5 (дата обращения: 04.09.2024).

88. European Ethical Charter on the Use of Artificial Intelligence in Judicial Systems and Their Environment // CEPEJ. — URL: https://www.coe.int/en/web/cepej/cepej-european-ethical-charter-on-the-use-of-artificial-intelligence-ai-in-judicial-systems-and-their-environment (дата обращения: 04.09.2024).

89. Guidelines on the Use of Generative Artificial Intelligence for Judges and Judicial Officers and Support Staff of the Hong Kong Judiciary // Hong Kong Judiciary. — URL: https://www.judiciary.hk/doc/en/court_services_facilities/guidelines_on_the_use_of_generative_ai.pdf (дата обращения: 04.09.2024).

90. Artificial Intelligence (AI). Guidance for Judicial Office Holders, Courts and Tribunals Judiciary // Courts and Tribunals Judiciary. — 2023. — URL: https://www.judiciary.uk/wp-content/uploads/2023/12/AI-Judicial-Guidance.pdf (дата обращения: 05.09.2024)..

91. China's court AI reaches every corner of justice system, advising judges and streamlining punishment // South China Morning Post. — URL: https://www.scmp.com/news/china/science/article/3185140/chinas-court-ai-reaches-every-corner-justice-system-advising (дата обращения: 05.09.2024).

92. French Magistrates See 'No Additional Value' in Predictive Legal AI // Artificial Lawyer. — URL: https://www.artificiallawyer.com/2017/10/13/french-justice-ministry-sees-no-additional-value-in-predictive-legal-ai/ (дата обращения: 04.09.2024).

93. Суды подключили искусственный интеллект к взысканию транспортного налога // Российская Газета. — URL: https://rg.ru/2021/04/10/sudy-podkliuchili-iskusstvennyj-intellekt-k-vzyskaniiu-transportnogo-naloga.html (дата обращения: 05.09.2024).

94. Colombian judge says he used ChatGPT in ruling // The Guardian. — URL: https://www.theguardian.com/technology/2023/feb/03/colombia-judge-chatgpt-ruling (дата обращения: 04.09.2024).

95. Момотов В. В. Искусственный интеллект в судопроизводстве: состояние, перспективы использования // Вестник Университета имени О. Е. Кутафина. — 2021. — № 5 (81). — URL: https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-v-sudoproizvodstve-sostoyanie-perspektivy-ispolzovaniya (дата обращения: 12.09.2024).

96. В Госдуме считают возможным использовать искусственный интеллект для разрешения налоговых споров // Ассоциация юристов России. — URL: https://alrf.ru/news/v-gosdume-schitayut-vozmozhnym-ispolzovat-iskusstvennyy-intellekt-dlya-razresheniya-nalogovykh-sporo/ (дата обращения: 05.09.2024).

97. «А судьи кто?»: как искусственный интеллект помогает человеку в суде // Сколково. — URL: https://sk.ru/news/a-sudi-kto-kak-iskusstvennyj-intellekt-pomogaet-cheloveku-v-sude/ (дата обращения: 04.09.2024).

98. Artificial Intelligence and Social Credit System in China // METU. — URL: https://open.metu.edu.tr/bitstream/handle/11511/101891/Artificial%20Intelligence%20and%20Social%20Credit%20System%20in%20China%20-%20Turgut%20BASER%20-%202013605.pdf (дата обращения: 14.08.2024).

99. Авдеев Д. «Социальный скоринг» как фактор нарушения права на неприкосновенность частной жизни // Международный научно-исследовательский журнал. — 2023. — № 6 (132).

100. Raz A., Minari J. AI-driven risk scores: should social scoring and polygenic scores based on ethnicity be equally prohibited? // Frontiers in Genetics. — 2023. — URL: https://doi.org/10.3389/fgene.2023.1092787 (дата обращения: 12.09.2024).

101. Катрашова Ю.В., Митяшин Г. Ю., Плотников В. А. Система социального рейтинга как форма государственного контроля над обществом: перспективы внедрения и развития, угрозы реализации // Управленческое консультирование. — 2021. — № 2 (146). — URL: https://cyberleninka.ru/article/n/sistema-sotsialnogo-reytinga-kak-forma-gosudarstvennogo-kontrolya-nad-obschestvom-perspektivy-vnedreniya-i-razvitiya-ugrozy (дата обращения: 13.08.2024).

102. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 // URL: http://data.europa.eu/eli/reg/2024/1689/oj (дата обращения: 13.08.2024).

103. Рекомендация об этических аспектах искусственного интеллекта, ЮНЕСКО // URL: https://unesdoc.unesco.org/ark:/48223/pf0000380455_rus (дата обращения: 25.07.2024).

104. РБК. Как работает социальный рейтинг в Китае // URL: https://style.rbc.ru/life/643d3f839a7947afd12e9f35 (дата обращения: 29.08.2024).

105. Chinese telecom giant ZTE helped Venezuela develop social credit system // ABC News. — URL: https://www.abc.net.au/news/2018-11-16/chinese-tech-giant-zte-helps-venezuela-develop-fatherland-card/10503736 (дата обращения: 11.09.2024).

106. Kostka G., Antoine L. Fostering Model Citizenship: Behavioral Responses to China's Emerging Social Credit Systems, 2018 // URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=3305724 (дата обращения: 14.08.2024).

107. Rabe W., Kostka G. Perceptions of social credit systems in Southeast Asia: An external technology acceptance model // Global Policy. — 2018. — Vol. 15, № 2. — С. 314–328.

108. Reports of 'big brother' China social credit system untrue: AI expert Xue Lan // Reuters. — URL: https://www.reuters.com/article/technology/reports-of-big-brother-china-social-credit-system-untrue-ai-expert-xue-lan-idUSKBN1ZL2P8/ (дата обращения: 11.09.2024).

109. Право.ru. LegalTech: скоринг в России и за рубежом // URL: https://pravo.ru/story/204834/ (дата обращения: 30.08.2024).

110. Москвич MAG. Чем нам грозит система социального рейтинга // URL: https://moskvichmag.ru/lyudi/chem-nam-grozit-sistema-sotsialnogo-rejtinga/ (дата обращения: 30.08.2024).

111. Парламентская газета. Искусственному интеллекту не позволят делить россиян на хороших и плохих // Парламентская газета. — URL: https://www.pnp.ru/social/iskusstvennomu-intellektu-ne-pozvolyat-delit-rossiyan-na-khoroshikh-i-plokhikh.html (дата обращения: 30.08.2024).

Список литературы к главе 2

112. OECD. «AI, data governance, and privacy: Synergies and areas of international co-operation» // OECD. — URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2024/06/ai-data-governance-and-privacy_2ac13a42/2476b1a4-en.pdf (дата обращения: 09.08.2024).

113. Федеральный закон от 27.07.2006 № 152-ФЗ (ред. от 08.08.2024) «О персональных данных» // Доступ из СПС КонсультантПлюс.

114. ICO. «How to use AI and personal data appropriately and lawfully» // ICO. — URL: https://ico.org.uk/media/for-organisations/documents/4022261/how-to-use-ai-and-personal-data.pdf (дата обращения: 08.08.2024).

115. Safeguarding User Privacy in the Digital Age: Personal Data and AI Training Ethics // HeyData. — URL: https://heydata.eu/en/magazine/safeguarding-user-privacy-in-the-digital-age-personal-data-and-ai-training-ethics (дата обращения: 12.08.2024).

116. IBM Research. «What is federated learning?» // IBM Research Blog. — URL: https://research.ibm.com/blog/what-is-federated-learning (дата обращения: 13.08.2024)

117. The Washington Post. «Facial recognition firm Clearview AI tells investors it's seeking massive expansion beyond law enforcement» // The Washington Post. — URL: https://www.washingtonpost.com/technology/2022/02/16/clearview-expansion-facial-recognition/ (дата обращения: 02.11.2024).

118. РБК. «Первая западная страна заблокировала ChatGPT» // РБК. — URL: https://www.rbc.ru/technology_and_media/31/03/2023/6426da7e9a794757b5679bb7 (дата обращения: 13.08.2024).

119. Kaspersky Daily. «Meta wants to use your posts and photos to train AI... Or does it?» // Kaspersky Daily. — URL: https://www.kaspersky.com/blog/meta-uses-personal-data/51548/ (дата обращения: 13.08.2024).

120. Digital Russia. «Отставание даже на несколько месяцев может существенно ухудшить позиции России на рынке ИИ-решений» // Digital Russia. — URL: https://d-russia.ru/otstavanie-dazhe-na-neskolko-mesjacev-mozhet-sushhestvenno-uhudshit-pozicii-rossii-na-rynke-ii-reshenij.html (дата обращения: 13.08.2024).

121. UNESCO. «The ethical implications of the Internet of Things (IoT)» // UNESCO. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000387201 (дата обращения: 08.08.2024).

122. ICO. «How to use AI and personal data appropriately and lawfully» // ICO. — URL: https://ico.org.uk/media/for-organisations/documents/4022261C/how-to-use-ai-and-personal-data.pdf (дата обращения: 08.08.2024).

123. Яковлева-Чернышева А.Ю., Яковлев-Чернышев В. А. «Проблемы правовой защиты персональных данных при использовании смартфонов и мобильных приложений» // CyberLeninka. — URL: https://cyberleninka.ru/article/n/problemy-pravovoy-zaschity-personalnyh-dannyh-pri-ispolzovanii-smartfonov-i-mobilnyh-prilozheniy (дата обращения: 09.08.2024).

124. ВЦИОМ. «"Умные" устройства в нашей жизни: возможности и риски» // ВЦИОМ. — URL: https://wciom.ru/analytical-reviews/analiticheskii-obzor/umnye-ustroistva-v-nashei-zhizni-vozmozhnosti-i-riski (дата обращения: 13.08.2024).

125. Unacceptable. «Exploring Accidental Triggers of Smart Speakers» // Unacceptable Privacy. — URL: https://unacceptable-privacy.github.io (дата обращения: 13.08.2024).

126. Quarts. «Google's second massive leak in a week shows it collected sensitive data from users» // Quarts. — URL: https://qz.com/google-leak-privacy-concerns-sensitive-user-information-1851519134 (дата обращения: 14.08.2024).

127. ИКС–Медиа. «В России могут ужесточить регулирование интернета вещей» // ИКС–Медиа. — URL: https://www.iksmedia.ru/news/5937902-V-Rossii-mogut-uzhestochit-reguliro.html (дата обращения: 13.08.2024).

128. Myagmar-Ochir Y., Kim W. «A Survey of Video Surveillance Systems in Smart City» // Electronics. — 2023. Vol. 12 (17), no. 3567. — С. 1–10.

129. OECD. «Artificial Intelligence in Society" // OECD. — URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2019/06/artificial-intelligence-in-society_c0054fa1/eedfee77-en.pdf (дата обращения: 14.08.2024).

130. Manchester Metropolitan University. «AI-driven mass surveillance at 2024 Olympics» // Manchester Metropolitan University. — URL: https://www.mmu.ac.uk/sites/default/files/2024–06/AI-driven%20Mass%20Surveillance%20at%202024%20Olympics%20-%20The%20Human%20Rights%20Issues%20and%20Recommendations.pdf (дата обращения: 09.08.2024).

131. Pranav D., Dubey T., Singh J. «A Literature Review: Artificial Intelligence in Public Security and Safety» // EasyChair. — 2020. No. 4578. — С. 1–15.

132. Решение Савеловского районного суда г. Москвы от 6 ноября 2019 г. по делу № 2а-577/19.

133. KOB4. «4 Investigates: Police use of AI facial recognition» // KOB4. — URL: https://www.kob.com/new-mexico/4-investigates-police-use-of-ai-facial-recognition/ (дата обращения: 13.08.2024).

134. Metropolitan Police. «Facial Recognition Technology» // Metropolitan Police. — URL: https://www.met.police.uk/advice/advice-and-information/fr/facial-recognition-technology/ (дата обращения: 13.08.2024).

135. The Guardian. «French court's approval of Olympics AI surveillance plan fuels privacy concerns» // The Guardian. — URL: https://www.theguardian.com/world/2023/may/18/french-courts-approval-of-olympics-ai-surveillance-plan-fuels-privacy-concerns (дата обращения: 14.08.2024).

136. ЮНЕСКО. «Рекомендация об этических аспектах искусственного интеллекта» // ЮНЕСКО. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000381133/PDF/381133eng.pdf.multi.page=94 (дата обращения: 09.08.2024).

137. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 // URL: http://data.europa.eu/eli/reg/2024/1689/oj (дата обращения: 13.08.2024).

138. ТАСС. «С помощью камер в Москве задержали 7,7 тыс. находящихся в федеральном розыске человек» // ТАСС. — URL: https://tass.ru/obschestvo/16983665 (дата обращения: 14.08.2024).

139. INTERPOL-UNICRI. «Towards Responsible Artificial Intelligence Innovation» // INTERPOL-UNICRI. — URL: https://unicri.it/towards-responsible-artificial-intelligence-innovation (дата обращения: 13.08.2024).

140. C. Rigano. «Using Artificial Intelligence to Address Criminal Justice Needs» // NIJ Journal. — 2018. — С. 1–10.

141. Regulation (EU) 2024/1689 of the European Parliament and of the Council of 13 June 2024 // URL: http://data.europa.eu/eli/reg/2024/1689/oj (дата обращения: 13.08.2024).

142. Joseph, J. «Predicting crime or perpetuating bias? The AI dilemma» // AI & Soc. — 2024. — URL: https://doi.org/10.1007/s00146–024–02032–9 (дата обращения: 13.08.2024).

143. The BSD of the University of Chicago. «Algorithm predicts crime a week in advance, but reveals bias in police response» // The BSD of the University of Chicago. — URL: https://biologicalsciences.uchicago.edu/news/algorithm-predicts-crime-police-bias (дата обращения: 13.08.2024).

144. МВД России. «МВД привлечет нейросети к поиску правонарушителей» // ai.gov.ru. — URL: https://ai.gov.ru/mediacenter/mvd-privlechet-neyroseti-k-poisku-pravonarushiteley/ (дата обращения: 13.08.2024).

145. International Banker. «The Role of AI in Shaping Credit Scoring in Emerging Markets» // International Banker. — URL: https://internationalbanker.com/technology/the-role-of-ai-in-shaping-credit-scoring-in-emerging-markets/ (дата обращения: 14.08.2024).

146. Addy W. A., Ajayi-Nifise A., Binaebi G. B., и др. «AI in credit scoring: A comprehensive review of models and predictive analytics” // Global Journal of Engineering and Technology Advances. — 2024. Vol. 18 (02), с. 118–129.

147. American National University. «The Use of Artificial Intelligence in Finance» // American National University. — URL: https://an.edu/the-use-of-artificial-intelligence-in-finance/ (дата обращения: 14.08.2024).

148. Банк России. «Применение Искусственного интеллекта на финансовом рынке» // Банк России. — URL: https://www.cbr.ru/Content/Document/File/156061/Consultation_Paper_03112023.pdf (дата обращения: 02.09.2024).

149. Jetland. «ИИ в кредитном скоринге» // Jetland. — URL: https://jetlend.ru/blog/novaya-era-v-kreditnom-skoringe-kak-ii-pomogaet-otsenivat-zaemshhikov-sprosili-ekspertov-i-rasskazali-pro-opyt-jetlend/ (дата обращения: 02.09.2024).

150. EastRussia. «Андрей Черкашин: Искусственный интеллект Сбера служит клиентам, банку и стране» // EastRussia. — URL: https://www.eastrussia.ru/material/andrey-cherkashin-iskusstvennyy-intellekt-sbera-sluzhit-klientam-banku-i-strane/ (дата обращения: 02.11.2024).

151. FutureBanking. «Эксперты рассказали, что работает в скоринге» // FutureBanking. — URL: https://futurebanking.ru/post/3129 (дата обращения: 02.09.2024).

Список литературы к главе 3

152. European Commission. “White Paper on Artificial Intelligence: a European approach to excellence and trust» // European Commission. — URL: https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=CELEX:52020DC0065&from=EN (дата обращения: 03.09.2024).

153. Харитонова Ю.С., Савина В. С., Паньини Ф. «Гражданско-правовая ответственность при разработке и применении систем искусственного интеллекта и робототехники: основные подходы» // CyberLeninka. — URL: https://cyberleninka.ru/article/n/grazhdansko-pravovaya-otvetstvennost-pri-razrabotke-i-primenenii-sistem-iskusstvennogo-intellekta-i-robototehniki-osnovnye-podhody (дата обращения: 27.08.2024).

154. Зазулин А.И. «Оценка доказательств, полученных в результате использования искусственного интеллекта» // eLibrary.ru. — URL: https://www.elibrary.ru/download/elibrary_46518344_97568948.pdf (дата обращения: 26.08.2024).

155. V7. «An Introductory Guide to Quality Training Data for Machine Learning» // V7. — URL: https://www.v7labs.com/blog/quality-training-data-for-machine-learning-guide (дата обращения: 03.09.2024).

156. ICTMoscow. «Датасеты в России: эксперты рынка о проблемах и возможностях» // ICTMoscow. — URL: https://ict.moscow/news/datasets/ (дата обращения: 02.11.2024).

157. Urzo F., Panico E., Custureri S. «Policy Brief — Ensuring Ethical AI Practices to counter Disinformation» // MediaFutures. — 2023. — URL: https://mediafutures.eu/wp-content/uploads/2023/09/MediaFutures_Policy-Briefs_Ensuring-Ethical-AI-Practices-to-counter-Disinformation.pdf (дата обращения: 03.09.2024).

158. UNESCO. «AI and the Holocaust: rewriting history? The impact of artificial intelligence on understanding the Holocaust» // UNESCO. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000390211 (дата обращения: 03.09.2024).

159. ВЦИОМ. «Этика искусственного интеллекта. По мнению россиян, сам ИИ установить этические ограничения не способен, поэтому его действия должен контролировать человек» // ВЦИОМ. — URL: https://wciom.ru/analytical-reviews/analiticheskii-obzor/ehtika-iskusstvennogo-intellekta-2 (дата обращения: 03.09.2024).

160. Areeb M., et al. «Filter Bubbles in Recommender Systems: Fact or Fallacy — A Systematic Review» // arXiv. — 2023. — URL: https://arxiv.org/abs/2307.01221v1 (дата обращения: 03.09.2024).

161. Этические рекомендации по применению рекомендательных технологий и алгоритмов, основанных на искусственном интеллекте, в цифровых сервисах // a-ai.ru. — URL: https://ethics.a-ai.ru/assets/ethics_documents/2023/09/19/Recommendation_Services_Ethics_01_0RYxN8h.pdf (дата обращения: 29.08.2024).

162. Yingqiang Ge, Shuchang Li, et al. «A Survey on Trustworthy Recommender Systems» // ACM Transactions on Recommender Systems. — 2024. Vol. 1 (1), article 1. — URL: https://arxiv.org/pdf/2207.125 (дата обращения: 06.09.2024).

163. ИСИЭЗ НИУ ВШЭ. «Алгоритмы рекомендуют, люди решают» // ИСИЭЗ НИУ ВШЭ. — URL: https://issek.hse.ru/news/850622348.html (дата обращения: 09.09.2024).

164. McKinsey & Company. «The value of getting personalization right — or wrong — is multiplying» // McKinsey & Company. — URL: https://www.mckinsey.com/capabilities/growth-marketing-and-sales/our-insights/the-value-of-getting-personalization-right-or-wrong-is-multiplying (дата обращения: 09.09.2024).

Список литературы к главе 4

165. Лукша Б.Н., Лаптёнок Н. В., Савенко А. Г. Искусственный интеллект в поисковых системах: обзор современного состояния технологий // Электронная библиотека BSUIR. — URL: https://libeldoc.bsuir.by/bitstream/123456789/47727/1/Luksha_Iskusstvennyy.pdf (дата обращения: 13.08.2024).

166. Талагаев М. Ю. Оптимизация процессов обработки информации с использованием технологий искусственного интеллекта на примере ChatBotGpt // Электронная библиотека. — URL: https://elibrary.ru/download/elibrary_54255871_67574899.pdf (дата обращения: 13.08.2024).

167. West, J.D., Memon, S. A. Search engines post-ChatGPT: How generative artificial intelligence could make search less reliable // Center for Internet Policy, University of Washington. — URL: https://www.cip.uw.edu/2024/02/18/search-engines-chatgpt-generative-artificial-intelligence-less-reliable/ (дата обращения: 13.08.2024).

168. Google. Правила Google Поиска в отношении контента, созданного искусственным интеллектом // Google Developers Blog. — URL: https://developers.google.com/search/blog/2023/02/google-search-and-ai-content?hl=ru (дата обращения: 13.08.2024).

169. Fliki. What is AI Voice Cloning: Tech, Ethics, and Future Possibilities // Fliki Blog. — URL: https://fliki.ai/blog/ai-voice-cloning (дата обращения: 13.08.2024).

170. Архипцев И.Н., Сарычев А. В., Мотузов А. В. К вопросу о правовом обеспечении предупреждения преступлений, совершаемых с использованием искусственного интеллекта и технологий, созданных на его основе в Российской Федерации // Электронная библиотека. — URL: https://elibrary.ru/download/elibrary_49267673_67493874.pdf (дата обращения: 14.08.2024).

171. Roswandowitz, C., Kathiresan, T., Pellegrino, E. et al. Cortical-striatal brain network distinguishes deepfake from real speaker identity // Commun Biol. — 2024. — Vol. 7, Article 711. — URL: https://doi.org/10.1038/s42003–024–06372–6 (дата обращения: 28.10.2024).

172. Fliki. What is AI Voice Cloning: Tech, Ethics, and Future Possibilities // Fliki Blog. — URL: https://fliki.ai/blog/ai-voice-cloning (дата обращения: 28.10.2024).

173. APNews. Deepfake of principal's voice is the latest case of AI being used for harm // AP News. — URL: https://apnews.com/article/ai-maryland-principal-voice-recording-663d5bc0714a3af221392cc6f1af985e (дата обращения: 28.10.2024).

174. MCA. Как искусственный интеллект трансформирует современное искусство // MCA.ру. — URL: https://msca.ru/blog/articles/kak-iskusstvennyy-intellekt-transformiruet-sovremennoe-iskusstvo (дата обращения: 15.08.2024).

175. Binary Ballet: Toeing the Line of Ethics in AI Art // MTU ICC Blog. — URL: https://blogs.mtu.edu/icc/2024/02/ethics-in-ai-art/ (дата обращения: 15.08.2024).

176. Морковкин Е.А., Новичихина А. А., Замулин И. С. Искусственный интеллект как инструмент современного искусства // Электронная библиотека. — URL: https://www.elibrary.ru/item.asp?id=47985194 (дата обращения: 15.08.2024).

177. Ethical Pros and Cons of AI Image Generation // IEEE Computer Society. — URL: https://www.computer.org/publications/tech-news/community-voices/ethics-of-ai-image-generation (дата обращения: 15.08.2024).

178. Generative AI in Art Market // MarketResearch.biz. — URL: https://marketresearch.biz/report/generative-ai-in-art-market/request-sample/ (дата обращения: 15.08.2024).

179. UNESCO Recommendation on the Ethics of Artificial Intelligence // UNESCO. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000381137 (дата обращения: 15.08.2024).

180. Why watermarking AI-generated content won't guarantee trust online // MIT Technology Review. — URL: https://www.technologyreview.com/2023/08/09/1077516/watermarking-ai-trust-online/ (дата обращения: 15.08.2024).

181. Labeling AI-Generated Content: Promises, Perils, and Future Directions // AI.gov. — URL: https://ai.gov.ru/knowledgebase/etika-i-bezopasnost-ii/2023_markirovka_kontenta_sozdannogo_s_pomoschyyu_iskusstvennogo_intellekta_effekt_opasnosti_i_napravleniya_na_buduschee_labeling_ai-generated_content_promises_perils_and_future_directions_mit/ (дата обращения: 15.08.2024).

182. Шумакова Н. И. Не только дипфейки: обязательная маркировка систем и продуктов генеративного искусственного интеллекта как часть этики его использования // Электронная библиотека. — URL: https://www.elibrary.ru/download/elibrary_59951853_64670388.pdf (дата обращения: 15.08.2024).

183. Digital Watermark Technology market Size, Share, Growth, and Industry Analysis, By Type (Invisible Digital Watermark), By Application (Broadcasting and Television Industry), Regional Insights and Forecast to 2032 // Business Research Insights. — URL: https://www.businessresearchinsights.com/market-reports/digital-watermark-technology-market-109456 (дата обращения: 16.08.2024).

184. Should the United States or the European Union Follow China's Lead and Require Watermarks for Generative AI? // Georgetown Journal of International Affairs. — URL: https://gjia.georgetown.edu/2023/05/24/should-the-united-states-or-the-european-union-follow-chinas-lead-and-require-watermarks-for-generative-ai/ (дата обращения: 16.08.2024).

185. Genius v. Google final complaint // New York State Courts. — URL: https://iapps.courts.state.ny.us/nyscef/ViewDocument?docIndex=3E0o8kQz4X3cWcbbid67wQ==&mod=article_inline (дата обращения: 16.08.2024).

186. Authentication of AI Needed to Protect the Public, Says OpenAI CEO Sam Altman // Broadbandbreakfast. — URL: https://broadbandbreakfast.com/authentication-of-ai-needed-to-protect-the-public-says-openai-ceo-sam-altman/ (дата обращения: 11.09.2024)

187. Laughter M.R., Anderson J. B., Maymone M. B.C., Kroumpouzos G. Psychology of aesthetics: Beauty, social media, and body dysmorphic disorder // ScienceDirect. — URL: https://www.sciencedirect.com/science/article/abs/pii/S0738081X23000299?via%3Dihub (дата обращения: 16.08.2024).

188. Computer-generated inclusivity: fashion turns to 'diverse' AI models // The Guardian. — URL: https://www.theguardian.com/fashion/2023/apr/03/ai-virtual-models-fashion-brands (дата обращения: 27.09.2024).

189. Посмотрите на победительниц первого конкурса красоты среди ИИ-участниц // РБК. — URL: https://www.rbc.ru/life/news/668c007b9a7947843efcaa2d (дата обращения: 27.09.2024).

190. Kenig N., Monton Echeverria J., Muntaner Vives A. Human Beauty according to Artificial Intelligence // NCBI PubMed Central. — URL: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC10371313/ (дата обращения: 16.08.2024).

191. CNN.Style. The rise of the AI beauty pageant and its complicated quest for the 'perfect' woman // CNN. — URL: https://edition.cnn.com/2024/06/27/style/miss-ai-beauty-pageant-scli/index.html (дата обращения: 27.09.2024).

192. Forbes. Medical Experts Share Impact Of AI On Beauty Standards // Forbes. — URL: https://www.forbes.com/sites/meggenharris/2024/03/27/medical-experts-share-impact-of-ai-on-beauty-standards/ (дата обращения: 27.09.2024).

Список литературы к главе 5

193. How AI and Data Could Personalize Higher Education // Harvard Business Review. — 2019. — URL: https://hbr.org/2019/10/how-ai-and-data-could-personalize-higher-education (дата обращения: 30.07.2024).

194. Embracing Tomorrow: How AI is Tailoring Education // The Princeton Review. — URL: https://www.princetonreview.com/ai-education/personalized-learning-with-ai (дата обращения: 30.07.2024).

195. Manoharan A., Nagar G. Maximizing Learning Trajectories: An Investigation into AI-Driven Natural Language Processing Integration in Online Educational Platforms // IRJETS. — 2021. — Vol. 03, no. 12, C. 123–130.

196. Sabharwal D., Kabha R., Srivastava K. Artificial Intelligence (AI)-Powered Virtual Assistants and Their Effect on Human Productivity and Laziness: Study on Students of Delhi-NCR (India) & Fujairah (UAE) // Journal of Content, Community and Communication. — 2023. — Vol. 17, no. 9, C. 162–174.

197. The State of AI in 2023: Generative AI's Breakout Year // McKinsey & Company. — URL: https://www.mckinsey.com/capabilities/quantumblack/our-insights/the-state-of-ai-in-2023-generative-ais-breakout-year (дата обращения: 06.08.2024).

198. 198. Руководство по использованию генеративного искусственного интеллекта в образовании и научных исследованиях. // UNESCO — URL: https://unesdoc.unesco.org/ark:/48223/pf0000389639 (дата обращения: 06.08.2024).

199. Capacity Building in Teaching of AR/VR Project, EU // URL: https://ec.europa.eu/info/funding-tenders/opportunities/portal/screen/opportunities/projects-details/43353764/101129191/ERASMUS2027?isExactMatch=true&programmePeriod=2021–2027&order=DESC&pageNumber=1&pageSize=50&sortBy=title&programId=43251814&page=1 (дата обращения: 31.07.2024).

200. Исследование карьерных путей педагогов России. // Институт образования ВШЭ. — URL: https://ioe.hse.ru/news/606138911.html (дата обращения: 06.08.2024).

201. Takahashi K. Social Issues with Digital Twin Computing. NTT Technical Review, 2020, vol. 18 (9), pp. 36–39.

202. Hawkinson E. Automation in Education with Digital Twins: Trends and Issues. International Journal on Open and Distance E-Learning, 2023, vol. 8 (2), pp. 1–9.

203. Chen Y. et al. Harnessing the Synergy of Real Teacher, Digital Twin, and AI in Blended Metaverse Learning Environment: A Catalyst for Medical Education Reforms // SSRN. — URL: https://papers.ssrn.com/sol3/papers.cfm?abstract_id=4745941 (дата обращения: 31.07.2024).

204. Центр преподавательского мастерства ВШМ СПбГУ. «Как мы выбрали и внедрили AI-аватар» // Telegram. — URL: https://t.me/methodgsom/358 (дата обращения: 06.08.2024).

205. Harnessing the Era of Artificial Intelligence in Higher Education: A Primer for Higher Education Stakeholders // UNESCO. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000386670 (дата обращения: 25.07.2024).

206. Al Braiki B., Harous S., Zaki N., Alnajjar F. Artificial Intelligence in Education and Assessment Methods // Bulletin of Electrical Engineering and Informatics. — 2020. — Vol. 9, no. 5, С. 1998–2007.

207. Руководство по использованию генеративного искусственного интеллекта в образовании и научных исследованиях // ЮНЕСКО. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000386693 (дата обращения: 25.07.2024)

208. Москва 24. Московский студент написал диплом с помощью нейросети // Москва 24. — URL: https://www.m24.ru/news/obrazovanie/01022023/546627 (дата обращения: 06.08.2024).

209. Регламент организации проверки письменных учебных работ на наличие плагиата, использования генеративных моделей и размещения выпускных квалификационных работ обучающихся по программам бакалавриата, специалитета и магистратуры на корпоративном сайте (портале) Национального исследовательского университета «Высшая школа экономики» // Национальный исследовательский университет «Высшая школа экономики». — URL: https://www.hse.ru/docs/922831988.html (дата обращения: 21.08.2024).

210. Приказ МГПУ от 4 сентября 2023 года № 633 // МГПУ. — URL: https://www.mgpu.ru/wp-content/uploads/2023/08/04.09.2023_633obshh_Remorenko_I.M._Safronova_E.S.-1.pdf (дата обращения: 21.08.2024).

211. Рамблер/новости. В российских вузах рассказали о плюсах и минусах использования ИИ при написании дипломных работ // Рамблер/новости. — URL: https://news.rambler.ru/education/51339837-v-rossiyskih-vuzah-rasskazali-o-plyusah-i-minusah-ispolzovaniya-ii-pri-napisanii-diplomnyh-rabot/?ysclid=m2z8o8x31186966098 (дата обращения: 02.11.2024).

212. Use of AI in Education: Deciding on the Future We Want // UNESCO. — URL: https://www.unesco.org/en/articles/use-ai-education-deciding-future-we-want (дата обращения: 01.08.2024).

213. The Evolution of Education: How AI is Reshaping Grading // The Princeton Review. — URL: https://www.princetonreview.com/ai-education/how-ai-is-reshaping-grading (дата обращения: 01.08.2024).

214. An Introduction to the Use of Generative AI Tools in Teaching // University of Oxford, Centre for Teaching and Learning. — URL: https://www.ctl.ox.ac.uk/ai-tools-in-teaching (дата обращения: 21.08.2024).

215. Sudhakar M. Enhancing Plagiarism Detection: The Role of Artificial Intelligence in Upholding Academic Integrity // Library Philosophy and Practice (e-journal). — 2023. — Режим доступа: https://digitalcommons.unl.edu/libphilprac/ (дата обращения: 01.08.2024).

216. Automated Grading Systems: How AI is Revolutionizing Exam Evaluation // Data Science Central. — URL: https://www.datasciencecentral.com/automated-grading-systems-how-ai-is-revolutionizing-exam-evaluation/ (дата обращения: 01.08.2024).

217. Luckin R., Holmes W., Griffiths M., Forcier L. B. Intelligence Unleashed: An Argument for AI in Education. — London: Pearson, 2016.

218. ТЕЛЕПОРТ.РФ. Контрольные работы проверяют учителя с помощью нейросетей // ТЕЛЕПОРТ.РФ. — URL: https://www.teleport2001.ru/news/2024–02–13/178570-kontrolnye-raboty-proveryayut-uchitelya-s-pomoschyu-neyrosetey.html (дата обращения: 06.08.2024).

219. Парламентская газета. Домашние задания к 2030 году будет проверять искусственный интеллект // Парламентская газета. — URL: https://www.pnp.ru/social/domashnie-zadaniya-proverit-iskusstvennyy-intellekt.html (дата обращения: 06.08.2024).

220. ТАСС. В Минпросвещения заявили, что учителя смогут проверять домашние задания при помощи ИИ // ТАСС. — URL: https://tass.ru/obschestvo/21068399 (дата обращения: 06.08.2024).

221. Справочник учебного процесса НИУ ВШЭ. Прокторинг // НИУ ВШЭ. — URL: https://www.hse.ru/studyspravka/distance_proctoring#:~=Прокторинг%20—%20это%20процедура%20контроля%20за, хорошо%20знакомо%20не%20только%20экспертам (дата обращения: 02.11.2024).

222. Software that monitors students during tests perpetuates inequality and violates their privacy // MIT Technology Review. — URL: https://www.technologyreview.com/2020/08/07/1006132/software-algorithms-proctoring-online-tests-ai-ethics/ (дата обращения: 06.08.2024).

223. Giannopoulou A., Ducato R., Angiolini C., Schneider G. From data subjects to data suspects: challenging e-proctoring systems as a university practice // JIPITEC. — 2023. — Vol. 14, no. 2. — С. 278–306. — DOI: 10.2139/ssrn.3522398.

224. Coghlan S., Miller T., Paterson J. Good Proctor or "Big Brother"? Ethics of Online Exam Supervision Technologies // Philosophy & Technology. — 2021. — Vol. 34, no. 4. — С. 1581–1606. — DOI: 10.1007/s13347-021-00451-7.

225. Unleash the potential of digital to change the traditional pen and paper exam practices // UNESCO's IIEP Learning Portal. — URL: https://learningportal.iiep.unesco.org/en/blog/unleash-the-potential-of-digital-to-change-the-traditional-pen-and-paper-exam-practices (дата обращения: 06.08.2024).

226. Remote online exams in higher education during the COVID-19 crisis // OECD. — URL: https://www.oecd.org/content/dam/oecd/en/publications/reports/2020/08/remote-online-exams-in-higher-education-during-the-covid-19-crisis_bfc8085e/f53e2177-en.pdf (дата обращения: 06.08.2024).

227. Списать не дам: что такое онлайн-прокторинг и как он работает // РБК. — URL: https://trends.rbc.ru/trends/education/5fa01fe49a794782c65b74f9?from=copy (дата обращения: 07.08.2024).

228. На прокторинг вешают вообще все неудобства, связанные с дистанционными экзаменами // Skillbox Media. — URL: https://skillbox.ru/media/education/na-proktoring-veshayut-voobshche-vse-neudobstva-svyazannye-s-distantsionnymi-ekzamenami/ (дата обращения: 07.08.2024).

229. AI Detectors Don't Work. Here's What to Do Instead // MIT Sloan Teaching & Learning Technologies. — URL: https://mitsloanedtech.mit.edu/ai/teach/ai-detectors-dont-work/ (дата обращения: 07.08.2024).

230. AI Detection Tools Falsely Accuse International Students of Cheating // The Markup. — URL: https://themarkup.org/machine-learning/2023/08/14/ai-detection-tools-falsely-accuse-international-students-of-cheating (дата обращения: 07.08.2024).

231. Liang W., Yuksekgonul M., Mao Y., Wu E., Zou J. GPT detectors are biased against non-native English writers // Patterns. — 2023. — Vol. 4. — DOI: 10.1016/j.patter.2023.100549.

232. How can educators respond to students presenting AI-generated content as their own? // OpenAI Help Center. — URL: https://help.openai.com/en/articles/8313351-how-can-educators-respond-to-students-presenting-ai-generated-content-as-their-own (дата обращения: 07.08.2024).

233. Generative AI in education: Educator and expert views // UK Department for Education. — URL: https://assets.publishing.service.gov.uk/media/65b8cd41b5cb6e000d8bb74e/DfE_GenAI_in_education_-_Educator_and_expert_views_report.pdf (дата обращения: 07.08.2024).

234. Executive Summary: Artificial Intelligence and Children's Rights // UNICEF. — URL: https://www.unicef.org/innovation/media/10726/file/Executive%20Summary:%20Memorandum%20on%20Artificial%20Intelligence%20and%20Child%20Rights.pdf (дата обращения: 07.08.2024).

235. What is the Role of Artificial Intelligence in Education? // HighSpeed Training. — URL: https://www.highspeedtraining.co.uk/hub/artificial-intelligence-in-education/ (дата обращения: 07.08.2024).

236. Kids and the Future of Artificial Intelligence // HART Research. — URL: https://www.4-h.org/wp-content/uploads/2024/02/27162629/Hart-Research-Youth-AI-Survey-Results42.pdf (дата обращения: 22.08.2024).

Список литературы к главе 6

237. Guo J., Li B. The Application of Medical Artificial Intelligence Technology in Rural Areas of Developing Countries // Health Equity. — 2018. — Vol. 2, No. 1. — С. 174–181.

238. KFF Health Misinformation Tracking Poll: Artificial Intelligence and Health Information // KFF. — URL: https://www.kff.org/health-misinformation-and-trust/poll-finding/kff-health-misinformation-tracking-poll-artificial-intelligence-and-health-information/ (дата обращения: 13.09.2024).

239. Zhao M., Hoti K., Wang H. et al. Assessment of Medication Self-Administration Using Artificial Intelligence // Nature Medicine. — 2021. — Vol. 27. — С. 727–735.

240. Robots in Healthcare: A Solution or a Problem? // European Parliament. — URL: https://www.europarl.europa.eu/RegData/etudes/IDAN/2019/638391/IPOL_IDA(2019)638391_EN.pdf (дата обращения: 12.09.2024).

241. Abdelwanis M. et al. Exploring the Risks of Automation Bias in Healthcare Artificial Intelligence Applications: A Bowtie Analysis // Journal of Safety Science and Resilience. — 2024. — URL: https://doi.org/10.1016/j.jnlssr.2024.06.001 (дата обращения: 12.09.2024).

242. Denecke K., Baudoin C. R. A Review of Artificial Intelligence and Robotics in Transformed Health Ecosystems // Frontiers in Medicine. — 2022. — Vol. 9. — Article 795967. (дата обращения: 12.09.2024).

243. 60% of Americans Would Be Uncomfortable With Provider Relying on AI in Their Own Health Care // Pew Research Center. — URL: https://www.pewresearch.org/science/2023/02/22/60-of-americans-would-be-uncomfortable-with-provider-relying-on-ai-in-their-own-health-care/ (дата обращения: 12.09.2024).

244. Министерство здравоохранения Российской Федерации. Цифровизацию и технологии искусственного интеллекта обсудят на Форуме будущих технологий — 2024 // Министерство здравоохранения Российской Федерации. — URL: https://minzdrav.gov.ru/news/2024/02/10/20871-tsifrovizatsiyu-i-tehnologii-iskusstvennogo-intellekta-obsudyat-na-forume-buduschih-tehnologiy-2024 (дата обращения: 13.09.2024).

245. So N. T. Y., Ngan O. M. Y. In-patient Suicide After Telephone Delivery of Bad News to a Suspected COVID-19 Patient: What Could Be Done to Improve Communication Quality? // Health Care Science. — 2023. — Vol. 2, No. 6. — С. 400–405.

246. The Pros and Cons of Healthcare Chatbots // News Medical Life Sciences. — URL: https://www.news-medical.net/health/The-Pros-and-Cons-of-Healthcare-Chatbots.aspx (дата обращения: 10.09.2024).

247. When Doctors Use a Chatbot to Improve Their Bedside Manner // The New York Times. — URL: https://www.nytimes.com/2023/06/12/health/doctors-chatgpt-artificial-intelligence.html (дата обращения: 10.09.2024).

248. A Doctor in California Appeared via Video Link to Tell a Patient He Was Going to Die. The Man's Family Is Upset // CNN. — URL: https://edition.cnn.com/2019/03/10/health/patient-dies-robot-doctor/index.html (дата обращения: 10.09.2024).

249. Федеральный закон от 21.11.2011 N 323-ФЗ (ред. от 08.08.2024) «Об основах охраны здоровья граждан в Российской Федерации» (с изм. и доп., вступ. в силу с 01.09.2024) // Собрание законодательства Российской Федерации. — 2024. — № 32. — Ст. 5115.

250. de Miguel I., Sanz B., Lazcoz G. Machine Learning in the EU Health Care Context: Exploring the Ethical, Legal and Social Issues // Information, Communication & Society. — 2020. — Vol. 23, No. 8. — Pp. 1139–1153. — DOI: 10.1080/1369118X.2020.1719185.

251. Ethics and Governance of Artificial Intelligence for Health: WHO Guidance // World Health Organization. — URL: https://iris.who.int/bitstream/handle/10665/341996/9789240029200-eng.pdf?sequence=1&isAllowed=y (дата обращения: 13.09.2024).

252. Park H. J. Perspectives on Informed Consent for Medical AI: A Web-based Experiment // Digital Health. — 2024. — Vol. 10. — DOI: 10.1177/20552076241247938 (дата обращения: 13.09.2024).

Список литературы к главе 7

253. Newcomb K. The Place of Artificial Intelligence in Sentencing Decisions // University of New Hampshire. — 2024. — URL: https://www.unh.edu/inquiryjournal/blog/2024/03/place-artificial-intelligence-sentencing-decisions (дата обращения: 05.09.2024).

254. Exploring the Use of AI in Legal Decision Making: Benefits and Ethical Implications // Woxsen University. — 2023. — URL: https://woxsen.edu.in/research/white-papers/exploring-the-use-of-ai-in-legal-decision-making-benefits-and-ethical-implications/ (дата обращения: 05.09.2024).

255. Middha M. et al. AI in Legal Evidence Analysis: Ethical and Legal Implications // IJLRA. — 2024. — Vol. 2, Issue 7.

256. L'Open Data des Décisions de Justice se Déroule Comme Prévu // Usine Digitale. — URL: https://www.usine-digitale.fr/article/l-open-data-des-decisions-de-justice-se-deroule-comme-prevu.N2000527 (дата обращения: 22.08.2024).

257. Caselaw Access Project // Harvard Law School. — URL: https://lil.law.harvard.edu/projects/caselaw-access-project (дата обращения: 22.08.2024).

258. Legal Research to Become More Efficient with New Large Language Model Contextualised to Domestic Law // Infocomm Media Development Authority. — 2024. — URL: https://www.imda.gov.sg/resources/press-releases-factsheets-and-speeches/factsheets/2024/gpt-legal (дата обращения: 22.08.2024).

259. McGuire H. Chatbots as a Tool for Pro Se Litigants // Journal of High Technology Law. — 2023. — URL: https://sites.suffolk.edu/jhtl/2023/02/23/chatbots-as-a-tool-for-pro-se-litigants/ (дата обращения: 06.09.2024).

260. The Use of Generative Artificial Intelligence (AI). Guidelines for Responsible Use by Non-Lawyers // Queensland Courts. — URL: https://www.courts.qld.gov.au/_data/assets/pdf_file/0012/798375/artificial-intelligence-guidelines-for-non-lawyers.pdf (дата обращения: 06.09.2024).

261. Queudot M., Charton É., Meurs M. Improving Access to Justice with Legal Chatbots // Stats. — 2020. — Vol. 3(3), C. 356–375.

262. Robot Gives Guidance in Beijing Court // ChinaDaily. — URL: https://www.chinadaily.com.cn/china/2017–10/13/content_33188642.htm (дата обращения: 06.09.2024).

263. Chien Colleen V. et al. How Generative AI Can Help Address the Access to Justice Gap Through the Courts // Loyola of Los Angeles Law Review. — Forthcoming, 2024. — URL: https://ssrn.com/abstract=4683309 (дата обращения: 06.09.2024).

264. United States District Court. Southern District of New York. Case 1:22-cv-01461-PKC // CourtListener. — URL: https://storage.courtlistener.com/recap/gov.uscourts.nysd.575368/gov.uscourts.nysd.575368.54.0_6.pdf (дата обращения: 10.09.2024).

265. American Bar Association. Standing Committee on Ethics and Professional Responsibility. Formal Opinion 512. "Generative Artificial Intelligence Tools" // American Bar Association. — URL: https://www.americanbar.org/content/dam/aba/administrative/professional_responsibility/ethics-opinions/aba-formal-opinion-512.pdf (дата обращения: 12.09.2024).

266. Юристы уличили искусственный интеллект во лжи при подготовке юридических документов // Российская газета. — URL: https://rg.ru/2023/05/31/iuristy-ulichili-iskusstvennyj-intellekt-vo-lzhi-pri-podgotovke-iuridicheskih-dokumentov.html (дата обращения: 02.11.2024).

267. Выступление председателя Совета судей Российской Федерации Момотова В. В. на пленарном заседании Совета судей РФ 5 декабря 2023 года // Совет Судей Российской Федерации. — URL: http://www.ssrf.ru/news/vystuplieniia-intierv-iu-publikatsii/52649 (дата обращения: 06.09.2024).

Список литературы к главе 8

268. Artificial intelligence in healthcare. Applications, risks, and ethical and societal impacts // EPRS. — URL: https://www.europarl.europa.eu/RegData/etudes/STUD/2022/729512/EPRS_STU(2022)729512_EN.pdf (дата обращения: 30.08.2024).

269. Pham K.T., Nabizadeh A., Selek S. Artificial Intelligence and Chatbots in Psychiatry // Psychiatr Q. — 2022. — Vol. 93. — C. 249–253.

270. Du J. S. Applications of AI in Psychotherapy: An Innovative Tool // Cambridge Science Advance. — 2024. — Vol. 2024, no. 2. — C. 1–6.

271. All Worked Up. A Report on the State of American Employees' Mental Health // Wysa. — URL: https://blogs.wysa.io/wp-content/uploads/2022/12/All-Worked-Up-Report-FINAL.pdf (дата обращения: 30.08.2024).

272. Startup Uses AI Chatbot to Provide Mental Health Counseling and Then Realizes It 'Feels Weird' // Vice. — URL: https://www.vice.com/en/article/4ax9yw/startup-uses-ai-chatbot-to-provide-mental-health-counseling-and-then-realizes-it-feels-weird (дата обращения: 30.08.2024).

273. Коммерсантъ. «Как искусственный интеллект может помочь психотерапевтам» // Коммерсантъ. — URL: https://www.kommersant.ru/doc/6774969 (дата обращения: 05.09.2024).

274. Preventing AI Misuse: Current Techniques, Centre for the Governance of AI // Centre for the Governance of AI. — URL: https://www.governance.ai/post/preventing-ai-misuse-current-techniques (дата обращения: 02.09.2024).

275. Руководство по использованию генеративного искусственного интеллекта в образовании и научных исследованиях, ЮНЕСКО // ЮНЕСКО. — URL: https://unesdoc.unesco.org/ark:/48223/pf0000389639 (дата обращения: 02.09.2024).

276. Ученые разработали защиту чат-ботов от выдачи негативных советов // Российская газета. — URL: https://rg.ru/2024/01/19/uchenye-sozdali-prostoj-metod-zashchity-chat-botov-ot-vydachi-negativnyh-sovetov.html (дата обращения: 30.08.2024).

277. OpenAI тестирует ИИ-системы по модерации контента // ТАСС. — URL: https://tass.ru/obschestvo/18518001 (дата обращения: 02.09.2024).

278. Azure AI Content Safety. Safeguard user and AI-generated text and image content // Microsoft Azure. — URL: https://azure.microsoft.com/en-us/products/ai-services/ai-content-safety (дата обращения: 02.09.2024).

279. Ethics guidelines for trustworthy AI, European Commission // European Commission. — URL: https://digital-strategy.ec.europa.eu/en/library/ethics-guidelines-trustworthy-ai (дата обращения: 02.09.2024).

280. Laestadius L. et al. Too human and not human enough: A grounded theory analysis of mental health harms from emotional dependence on the social chatbot Replika. New Media & Society, 2022. — URL: https://doi.org/10.1177/14614448221142007 (дата обращения: 02.09.2024).

281. GPT-4o System Card, OpenAI // OpenA I. — URL: https://openai.com/index/gpt-4o-system-card/ (дата обращения: 02.09.2024).

282. 'I love her and see her as a real woman.' Meet a man who 'married' an artificial intelligence hologram // CBC News. — URL: https://www.cbc.ca/documentaries/the-nature-of-things/i-love-her-and-see-her-as-a-real-woman-meet-a-man-who-married-an-artificial-intelligence-hologram-1.6253767 (дата обращения: 02.09.2024).

283. The Brussels Times. «Belgian man dies by suicide following exchanges with chatbot» // The Brussels Times. — URL: https://www.brusselstimes.com/430098/belgian-man-commits-suicide-following-exchanges-with-chatgpt (дата обращения: 05.09.2024).

284. A Conversation with OpenAI's Sam Altman and Mira Murati, WSJ // The Wall Street Journal. — URL: https://www.wsj.com/podcasts/the-journal/a-conversation-with-openais-sam-altman-and-mira-murati/7c89e85f-9d7e-4569-b67d-6a777374eada (дата обращения: 02.09.2024).

285. Чрезмерное общение с чат-ботами может негативно повлиять на социализацию, Российская газета // Российская газета. — URL: https://rg.ru/2024/06/07/chrezmernoe-obshchenie-s-chat-botami-mozhet-negativno-povliiat-na-socializaciiu.html (дата обращения: 02.09.2024).

286. VC.ru. «Четыре жизни Replika: что происходит с проектом, создающим цифровую копию человека» // VC.ru. — URL: https://vc.ru/story/64057-chetyre-zhizni-replika-chto-proishodit-s-proektom-sozdayushim-cifrovuyu-kopiyu-cheloveka (дата обращения: 27.09.2024).

287. ResearchGate. «Attachment Theory as a Framework to Understand Relationships with Social Chatbots: A Case Study of Replika» // ResearchGate. — URL: https://www.researchgate.net/publication/357665581_Attachment_Theory_as_a_Framework_to_Understand_Relationships_with_Social_Chatbots_A_Case_Study_of_Replika (дата обращения: 27.09.2024).

288. Лаптев В. А. Понятие искусственного интеллекта и юридическая ответственность за его работу // Право. Журнал Высшей школы экономики. — 2019. — № 2. — URL: https://cyberleninka.ru/article/n/ponyatie-iskusstvennogo-intellekta-i-yuridicheskaya-otvetstvennost-za-ego-rabotu (дата обращения: 06.09.2024).

289. Kureha M. On the moral permissibility of robot apologies // AI & Society. — 2023. — URL: https://doi.org/10.1007/s00146-023-01782-2 (дата обращения: 03.09.2024).

290. Criminalising Offensive Speech Made by AI Chatbots in Singapore // Tech for Good Institute. — URL: https://techforgoodinstitute.org/blog/expert-opinion/criminalising-offensive-speech-made-by-ai-chatbots-in-singapore/ (дата обращения: 03.09.2024).

291. Чат-бот от Microsoft за сутки возненавидел человечество и стал нацистом // LENTA.RU. — URL: https://lenta.ru/news/2016/03/24/neonazi/ (дата обращения: 03.09.2024).

292. Валерий Зорькин высказался против наделения искусственного интеллекта правосубъектностью // Адвокатская Газета. — URL: https://www.advgazeta.ru/novosti/valeriy-zorkin-vyskazalsya-protiv-nadeleniya-iskusstvennogo-intellekta-pravosubektnostyu/ (дата обращения: 03.09.2024).

293. iConText Group. Искусственный интеллект в рекрутинге: как технологии меняют подход к поиску и найму сотрудников // Блог iConText Group. — 2024. — URL: https://blog.icontextgroup.ru/articles/iskusstvennyj-intellekt-v-rekrutinge (дата обращения: 06.09.2024).

294. OECD. Artificial Intelligence and Labour Market Matching, OECD Working Paper // OECD. — 2023. — URL: https://one.oecd.org/document/DELSA/ELSA/WD/SEM(2023)2/en/pdf (дата обращения: 04.09.2024).

295. Ekleft. Использование AI в рекрутинге: как можно улучшить и оптимизировать процесс найма // Ekleft. — 2024. — URL: https://ekleft.ru/mediatsentr/AI_HR/ (дата обращения: 06.09.2024).

296. Яков и Партнеры, Яндекс. Искусственный интеллект в России — 2023: тренды и перспективы // Яков и Партнеры. — 2023. — URL: https://yakov.partners/upload/iblock/c5e/c8t1wrkdne5y9a4nqlicderalwny7xh4/20231218_AI_future.pdf (дата обращения: 06.09.2024).

297. Искусственный интеллект Российской Федерации. Исследование HRlink: 71% HR-специалистов позитивно относится к ИИ // Искусственный интеллект Российской Федерации. — 2024. — URL: https://ai.gov.ru/mediacenter/issledovanie-hrlink-71-hr-spetsialistov-pozitivno-otnositsya-k-ii/?pageStart=150 (дата обращения: 06.09.2024).

298. Искусственный интеллект будет нанимать госслужащих: как изменится российский рынок труда? // Национальный портал в сфере ИИ. — 2024. — URL: https://ai.gov.ru/mediacenter/iskusstvennyy-intellekt-budet-nanimat-gossluzhashchikh-kak-izmenitsya-rossiyskiy-rynok-truda- (дата обращения: 04.09.2024).

299. Профессия — рекрутер: ИИ оценит личные качества претендентов на вакансию // Известия. — 2024. — URL: https://iz.ru/1729533/denis-gritcenko/professiia-rekruter-ii-otcenit-lichnye-kachestva-pretendentov-na-vakansiiu?ysclid=m0no7qtrho759589246 (дата обращения: 04.09.2024).

300. Известия. Спортивный ИИнтерес: как нейросети помогают атлетам // Известия. — 2024. — URL: https://iz.ru/1563282/alena-svetunkova/sportivnyi-iinteres-kak-neiroseti-pomogaiut-atletam (дата обращения: 09.09.2024).

301. Касиси Джоэл. Применение искусственного интеллекта в спорте // Cyberleninka. — 2024. — URL: https://cyberleninka.ru/article/n/primenenie-iskusstvennogo-intellekta-v-sporte (дата обращения: 09.09.2024).

302. Mordor Intelligence. ИИ на спортивном рынке // Mordor Intelligence. — 2024. — URL: https://www.mordorintelligence.com/ru/industry-reports/artificial-intelligence-market-in-sports (дата обращения: 09.09.2024).

303. Statista. Global sports analytics market revenue 2022–2031 // Statista. — 2024. — URL: https://www.statista.com/statistics/1185536/sports-analytics-market-size/ (дата обращения: 09.09.2024).

304. ATP Tour. Electronic Line Calling Live To Be Adopted Across The ATP Tour // ATP Tour. — 2024. — URL: https://www.atptour.com/en/news/electronic-line-calling-release-april-2023 (дата обращения: 09.09.2024).

305. Wired. DeepMind’s superhuman AI is rewriting how we play chess // Wired. — 2024. — URL: https://www.wired.com/story/deepmind-ai-chess/ (дата обращения: 09.09.2024).

Для заметок

Для заметок

Главный редактор И. Позина
Руководитель проекта Е. Кузнецова
Редактор А. Слатвинская
Корректор В. Вересиянова
Дизайн макета, обложки А. Кузьмина
Верстка А. Николаева

12+

Подписано в печать 29.11.2024
Формат 60x90/8. Гарнитура SB Sans Text
Бумага мелованная. Печать офсетная
Тираж 500 экз