НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ УНИВЕРСИТЕТ
«ВЫСШАЯ ШКОЛА ЭКОНОМИКИ»

Д.Э. Гаспарян, Е.М. Стырин

ПРИКЛАДНЫЕ ПРОБЛЕМЫ ВНЕДРЕНИЯ ЭТИКИ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В РОССИИ

Отраслевой анализ и судебная система

Издательский дом
Высшей школы экономики
МОСКВА, 2020

УДК 004.89:34.03
ББК 16.6+87.75
Г22

Рецензенты:

докт. филос. наук, доцент кафедры исторических и социально-философских дисциплин, востоковедения и теологии Пятигорского государственного университета *П.Н. Барышников;*
докт. филос. наук, доцент факультета общеуниверситетских дисциплин Томского государственного педагогического университета *С.Б. Куликов*

Авторы:

Гаспарян Диана Эдиковна, доцент Школы философии и культурологии факультета гуманитарных наук Национального исследовательского университета «Высшая школа экономики»;
Стырин Евгений Михайлович, ведущий научный сотрудник Института государственного и муниципального управления Национального исследовательского университета «Высшая школа экономики»

Гаспарян, Д. Э., Стырин, Е. М. Прикладные проблемы внедрения
Г22 этики искусственного интеллекта в России: отраслевой анализ и судебная система [Текст] / Д. Э. Гаспарян, Е. М. Стырин ; Нац. исслед. ун-т «Высшая школа экономики». — М.: Изд. дом Высшей школы экономики, 2020. — 112 с. — 500 экз. — ISBN 978-5-7598-2351-3 (в обл.). — ISBN 978-5-7598-2242-4 (e-book).

В книге исследуются этические проблемы, возникающие в процессе внедрения технологий искусственного интеллекта (ИИ) в государственное управление отраслями образования, здравоохранения, жилищно-коммунального хозяйства, а также в судебную систему. В начале книги приведен общий анализ прикладных проблем, дилемм, трудностей, связанных с этикой разработки и внедрения искусственного интеллекта в избранные сферы, изложены руководящие принципы и ценности этически ориентированного искусственного интеллекта и анализ этических проблем внедрения по избранным секторам. Каждая глава книги посвящена отдельной отрасли и делится на две части: в первой дан собственно обзор этических аспектов применения и внедрения ИИ, во второй — пул рекомендаций по возможному разрешению, устранению или минимизации сформулированных этических проблем.

УДК 004.89:34.03
ББК 16.6+87.75

Опубликовано Издательским домом Высшей школы экономики
<http://id.hse.ru>

doi:10.17323/978-5-7598-2351-3

ISBN 978-5-7598-2351-3 (в обл.)
ISBN 978-5-7598-2242-4 (e-book)

© Гаспарян Д.Э., Стырин Е.М., 2020

Содержание

Глоссарий 4

Введение 5

Руководящие принципы
и ценности этически-ориентированного ИИ 16

**Этические аспекты применения и внедрения ИИ
в судебной системе** 23

1. Судебная система 23
    1.1. ИИ в области права 25
    1.2. Перечень рекомендаций 43

**Этические аспекты применения
и внедрения ИИ в отраслевых сегментах** 47

2. Образование 47
    2.1. ИИ в области образования 48
    2.2. Перечень рекомендаций 56

3. Здравоохранение 60
    3.1. ИИ в области здравоохранения 63
    3.2. Перечень рекомендаций 79

4. ЖКХ. Системы «Умный дом» 83
    4.1. ИИ в области ЖКХ 87
    4.2. Перечень рекомендаций 96

Заключение 100

Библиография 103

Глоссарий

Искусственный интеллект (*ИИ*) — программная система для решения различных задач с помощью антропоразмерного интеллекта, функционирующего на автоматизированной основе.

Этическая экспертиза — тестирование технологии с точки зрения ее релевантности этическим нормам человека, причастности ценностям и нормативным предписаниям, а также психологической безопасности.

Социально-надежный ИИ — искусственный интеллект, отвечающий нормам этической релевантности, социальной предсказуемости и психологической безопасности.

Агенты ИИ — различные по своему функционалу программы, которые автономно работают по определенному расписанию на основе технологий ИИ над поставленными человеком задачами; способны получать и обрабатывать данные из внешней среды с последующей выдачей рационального результата, соизмеримого поставленным задачам.

Робот — 1) программируемый исполнительный механизм с определенным уровнем автономности для выполнения перемещения, манипулирования или позиционирования (включает систему управления и интерфейс для человека); 2) автономно действующий программный модуль с понятным для человека интерфейсом, который выполняет рутинные задачи в заданной предметной области по определенному расписанию (поиск, ответы на вопросы, сбор данных).

Государственное регулирование экономики — управляющее воздействие государства на экономическую деятельность субъектов хозяйствования, которое реализуется через различные *экономические* и *административные (внеэкономические)* механизмы.

Введение[1]

Искусственный интеллект, а точнее подразумеваемые под ним технологии машинного обучения, которые позволяют компьютерам понимать речь, разбирать тексты, классифицировать данные по заранее заданным критериям, т.е. решать ряд интеллектуальных творческих задач способами, аналогичными тем, которыми пользуется человек [Barr, 1989], прочно входит в дискурс органов исполнительной власти, но также может использоваться законодательной и судебной ветвями власти [Sun, Medaglia, 2019]. В принятой в 2019 году российской национальной стратегии развития искусственного интеллекта до 2030 года под искусственным интеллектом (далее — ИИ) понимается «комплекс технологических решений, имитирующий когнитивные функции человека (включая самообучение и поиск решений без заранее заданного алгоритма) и позволяющий при выполнении задач достигать результаты, как минимум сопоставимые с результатами интеллектуальной деятельности человека» [Указ Президента РФ № 490]. Комплекс технологических решений включает информационно-коммуникационную инфраструктуру, программное обеспечение, в котором в том числе используются методы машинного обучения, процессы и сервисы по обработке данных и выработке решений.

Поскольку спектр задач, решаемых ИИ, постоянно расширяется, мы перечислим только задачи, которые относятся к системе органов исполнительной власти, а также к судебной системе.

---

[1] Книга написана на основе прикладного исследования «Систематизация опыта ведущих стран мира в развитии технологий искусственного интеллекта и выработка предложений по нормативным и организационным мерам, направленным на опережающее развитие технологий искусственного интеллекта в Российской Федерации», проведенного ИГМУ НИУ ВШЭ в 2019 году в рамках программы прикладных исследований. В работе использованы результаты проекта «Трансцендентальный подход в философии: история и современность», выполненного в рамках Программы фундаментальных исследований НИУ ВШЭ в 2020 году.

Прежде всего, это класс задач по установлению содержательных связей между нормативно-правовыми актами или документами стратегического планирования с целью их взаимоувязки или поиска противоречий, несоответствий между целями, задачами и целевыми показателями в документах на различных уровнях управления. Известно, что нормативно-правовые акты могут содержать противоречия, ведь документы стратегического планирования в России созданы на всех уровнях управления, при этом указанные в них цели, задачи и показатели на региональном или муниципальном уровнях могут противоречить, например, ключевым документам на федеральном уровне. Работу по взаимной увязке и поиску противоречий может осуществлять программа на основе технологий искусственного интеллекта, которая в ряде документов устранит дублирование, выявит заведомо недостижимые показатели, проведет конечную классификацию типов документов стратегического планирования в зависимости от критериев классификации.

Другим классом задач для ИИ является предиктивная аналитика на основе массивов данных, которые описывают объект в отрасли, например, пациента по истории болезни в медицинской карте, школьника или студента по портфолио достижений и успеваемости, совершившего или подозреваемого в правонарушении или преступлении по материалам уголовного дела. На основе анализа аналогичных объектов одного класса рекомендательная система может классифицировать каждый новый объект относительно набора признаков, которые также могут быть априори заданы. Таким образом, можно с некоторой вероятностью получить предсказания о склонности заключенного совершить повторное преступление, о намерении ученика получить определенную профессию, о возникновении у пациента некоторой болезни или, например, послеоперационных осложнений. Другой целью рекомендательной системы может стать выработка предписаний о том, какое может быть назначено наказание подозреваемому, какие курсы в дальнейшем предпочтительно слушать школьнику или студенту, какие препараты принимать пациенту и процедуры проходить, чтобы предотвратить негативный сценарий развития болезни [ЦНТИ МФТИ, 2020].

Также у органов власти формируются данные для профиля гражданина в той роли, в которой он выступает объектом их соответствующих полномочий: налогоплательщик, подозреваемый в уголовном или административном правонарушении, предприниматель, пенсионер и т.д. На основе системы предиктивной аналитики по профилю гражданина можно выстраивать так называемые проактивные услуги, о которых гражданин мог не знать и не инициировать их самостоятельно. При этом соответствующие возможности получения поддержки, обеспечения и реализации прав были предусмотрены государством и предложены гражданину со стороны обеспечивающих органов власти в инициативном порядке. Например, проактивными услугами можно назвать предварительный расчет пенсии, начисление социальных пособий, подбор университета или места работы [Добролюбова, 2018].

Искусственный интеллект помогает обеспечивать безопасность граждан, например, сопоставляя лица граждан, попавших в камеры наблюдения, с лицами разыскиваемых нарушителей закона, тем самым идентифицируя их положение и перемещение в случае совпадения. Таким образом, органы внутренних дел получают мощный инструмент для отслеживания разыскиваемых лиц, совершивших правонарушения, выявления лиц, которые отличаются подозрительным или противоправным поведением [Faggella, 2019]. Аналогичные инструменты компьютерного зрения применяются на дорогах для выявления нарушителей правил дорожного движения. Большой потенциал для функционала компьютерного зрения наблюдается в системах «Умный дом» — это совокупность камер, датчиков и иных управляющих элементов, которые круглосуточно накапливают и предоставляют уполномоченному наблюдателю видеоданные о состоянии квартир, домов, придомовых территорий. Таким образом, для построения «умных городов», согласно соответствующей концепции Минстроя России (https://russiasmartcity.ru/), система интеллектуального видеонаблюдения является ее неотъемлемым элементом.

Отдельное направление не только в бизнесе, но и в государстве — замена рутинного труда человека на программу, функцио-

нирующую на основе технологий машинного обучения для выполнения рутинных операций. Примером такой программы является чат-бот, который может отвечать на достаточно простые вопросы граждан, связанные с государственными информационными ресурсами и заданные ему в режиме реального времени в текстовом виде, например, через официальные информационные ресурсы органов власти или судов. Вопросы могут затрагивать целый ряд тем: разъяснение отдельных положений законодательства, диагностику статусов граждан для получения социальной поддержки, консультации по получению государственных услуг. С помощью чат-ботов можно реализовать простые инструменты диагностики по различным направлениям. В основу положен механизм диалога между пользователем и чат-ботом, в котором посредством задания вопросов пользователем в виде текста или голосом может быть выявлена или решена некоторая проблема. Например, с помощью простых вопросов чат-бот может диагностировать наличие некоторого заболевания (в частности, COVID-19), определить, есть ли формальные предпосылки у кандидата пройти по конкурсу на некоторую должность на государственной службе, есть ли право у заявителя получить социальную льготу или субсидию, к примеру, в сельском хозяйстве.

Таким образом, программы на основе машинного обучения позволяют заменить труд человека при решении рутинных задач или хотя бы при их выполнении снизить нагрузку на государственных служащих, переключив усилия сотрудников на более сложные и неоднозначные проблемы.

Вообще говоря, помимо анализа и сопоставления данных, программы на основе искусственного интеллекта могут не только выдавать рекомендации, но и принимать решения в автоматическом режиме. Вопрос состоит в готовности ответственных управленцев контролировать каждое решение либо полагаться на решение, предложенное программой. Ярким примером может служить система государственного контроля и надзора, в которой применяется риск-ориентированный подход [Кнутов, Плаксин, 2019]. Данные о проверках объектов, требующих надзора (заводы, пред-

приятия, учреждения сферы образования, здравоохранения, сфера общественного питания), могут быть автоматически проанализированы и сопоставлены с нормативами, определяющими риски непроведения своевременных проверок. Таким образом, программа в состоянии присваивать риски поднадзорным объектам, автоматически устанавливая классы этих рисков, определяя в дальнейшем частоту проверок. При подобном подходе подразумевается только контроль со стороны человека, а выработка решения может оставаться за программой.

Аналогичным образом программа на основе технологий искусственного интеллекта может взять на себя функционал оценки регулирующего воздействия (ОРВ), которая проводится для целей государственного регулирования, определения возможных вариантов достижения целей, а также оценки связанных с ними позитивных и негативных эффектов [Клименко, Минченко, 2016]. При обеспечении сбора подробных цифровых данных об отрасли, подвергающейся регулированию, ИИ может просчитывать текущие и прогнозные ключевые экономические показатели государственной политики в отдельно взятой отрасли. При совершенствовании выбора решений на основе машинного обучения полученные прогнозы могут быть точнее, чем достаточно субъективные подходы, которые предлагаются отраслевыми экспертами. Тем не менее с учетом вероятностного расчета значений показателей с помощью ИИ полученные результаты также требуют контроля со стороны отраслевых экспертов и ответственных государственных служащих. Проведение ОРВ с привлечением функционала искусственного интеллекта существенно убыстряет процесс выработки и оценки альтернатив, которые возможны для совершенствования нормативно-правовой базы и основных показателей государственной политики в некоторой отрасли, а значит ИИ остается перспективной технологией для проведения ОРВ.

Осуществление государственной бюджетной и налоговой политики для стимулирования бизнеса также можно свести к задаче определения налогового режима и объема поддержки для различных компаний в зависимости от их финансово-экономических по-

казателей, сравнимых с установленными эталонными. Обученные нейронные сети могут не только классифицировать компании по критериям для определения налогового режима и бюджетной поддержки, но позволяют уточнить эти критерии на основе проанализированного множества собираемых показателей о деятельности компаний.

Абсолютно аналогично решаются задачи осуществления лицензионной и разрешительной деятельности органов власти, которые сравнимы с задачей кредитного скоринга клиента в банке. Организация, деятельность которой подлежит лицензированию или требует получения разрешений, может быть в автоматическом режиме оценена нейросетью на основе ранее изученных аналогичных данных. В таком случае организация может быть автоматически классифицирована под положительное либо отрицательное решение о выдаче лицензии или разрешения. Окончательное решение может быть принято сотрудником-специалистом, однако подавляющая часть предварительных расчетов для определения параметров выдачи лицензии или разрешения может быть проведена программой на основе технологий ИИ.

Приведенный выше обзор возможностей ИИ, призванных помочь в решении разных классов задач в государственном управлении и в судебной системе, показывает неизбежность внедрения инновационных решений на основе машинного обучения в ближайшем будущем. Искусственный интеллект дает возможность сократить издержки при осуществлении государственных функций, увеличить скорость отклика на запросы граждан, повысить качество результатов взаимодействия органов власти с внешними акторами, а также перераспределить нагрузку на государственных служащих, избавив их от решения рутинных задач. При этом изложенные выше возможности ИИ демонстрируют позитивный эффект от его внедрения в деятельность органов власти. Однако деятельность органов государственной власти сопряжена с пристальным вниманием общественности, требованиями соблюдать прозрачность и подотчетность в принятии решений и представлении результатов. Если опорой деятельности для органов власти становится ИИ, то госу-

дарству следует обеспечить важнейший аспект внедрения ИИ в операционную и стратегическую деятельность — этический.

Анализ зарубежных и российской национальной стратегии развития искусственного интеллекта демонстрирует, что в подавляющем большинстве в них включен раздел по обеспечению этичности использования ИИ в деятельности органов власти. Как правило, в стратегиях определяются этические принципы внедрения ИИ (конфиденциальность, прозрачность, подотчетность, защита гражданских прав, ответственность, доверие, устойчивое развитие, справедливость, приоритет конечного решения за человеком), декларируется необходимость защиты персональных данных граждан, принятие решений с помощью технологий ИИ согласно важным для общества ценностям и убеждениям. Важную роль в принятии этических принципов развития ИИ сыграли ОЭСР и ЕС[2]. Этические механизмы в национальных стратегиях развития ИИ – это один из механизмов контроля над результатами деятельности на основе использования ИИ. Например, декларируется право человека принимать конечные решения и вмешиваться при необходимости в деятельность, реализованную на технологиях ИИ. Использование ИИ должно вызывать доверие у общества: должны быть созданы механизмы верификации действий ИИ, механизмы защиты граждан от ошибок ИИ, которые могут нанести урон человеку, привести к дискриминации одних групп граждан перед другими при принятии решений, ущемлять моральные, религиозные ценности и достоинство граждан при реализации управленческих механизмов на основе ИИ. Отдельно государства декларируют создание консультативных органов, которые будут дискуссионными площадками для выработки политики по реализации этических принципов в продуктах и услугах на основе ИИ.

---

[2] Forty-two countries adopt new OECD Principles on Artificial Intelligence (URL: https://www.oecd.org/science/forty-two-countries-adopt-new-oecd-principles-on-artificial-intelligence.htm); Ethics Guidelines for Trustworthy AI (URL: https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai).

В Российской Федерации принята национальная стратегия развития искусственного интеллекта на период до 2030 года, создан теоретический и технологический задел в области ИИ, сформирован рынок цифровых продуктов и услуг на основе ИИ. Государственная политика России предусматривает содействие повсеместному отраслевому внедрению ИИ, включая сектор государственного управления. В качестве прикладных областей исследования мы выбрали очень чувствительные для общества сферы, в которых вопросы прозрачности, подотчетности, справедливости, соблюдения прав граждан не менее важны, чем инвестиции, внедрение цифровых технологий, создание стандартов. Это — судебная система, отрасли образования, здравоохранения, жилищно-коммунального хозяйства.

Цель данной книги заключается в разработке основанного на этике подхода к внедрению технологий ИИ в работу органов государственной власти (с акцентом на примеры из практики в Российской Федерации). Сопутствующая задача в этом исследовании — это обнаружение условий формирования, внедрения и адаптации возможностей этичного ИИ для нужд государственного управления. Отдельная задача состоит в прогнозировании и обнаружении таких нужд. Важнейшей областью участия государства в вопросах развития ИИ является разработка этического кодекса ИИ, а также контроль за его использованием. Бурное развитие ИИ в современном мире сопровождается повышением числа этических трудностей, дилемм и коллизий на пути внедрения и применения ИИ. В то же время для разработчиков и конечных пользователей существует понимание важности того, что внедрение ИИ должно соответствовать этическим нормам и ожиданиям людей, которые требуют детального рассмотрения и уточнения.

Этически ориентированный, или так называемый надежный (reliable), ИИ включает три взаимосвязанных аспекта, каждый из которых предполагает отдельный уровень экспертной оценки.

Во-первых, его разработка и применение должны соответствовать существующему законодательству и правоприменительным практикам, не нарушать прав и обязанностей граждан, не мешать

их исполнению. Успешная реализация настоящего требования предполагает проведение юридической экспертизы с привлечением советующих квалифицированных специалистов.

Во-вторых, разработка надежного ИИ предполагает соблюдение и воспроизводство ключевых этических норм, принципов и ценностей, присущих определенной культуре или региону. Для их соблюдения также необходимо привлечение соответствующих экспертов, а именно тех, кто обладает достаточными компетенциями для проведения самой этической экспертизы ИИ. В число таких экспертов могут входить люди с междисциплинарной подготовкой и разносторонней квалификацией: философы, психологи, социологи, управленцы и др. [Russell, Norvig, 2009].

В-третьих, разработка и внедрение ИИ должны быть надежными с социальной точки зрения. Под этим параметром подразумевается этическая безопасность применения ИИ, когда объектом воздействия является большое количество людей или общество в целом. Даже при успешном прохождении этической экспертизы системы искусственного интеллекта могут причинить непреднамеренный или незаметный вред в краткосрочной перспективе. Для устранения таких рисков применение ИИ должно поддерживаться не только технологиями конечного внедрения, но и процедурами долгосрочного мониторинга.

В оптимальной перспективе все три компонента должны работать слаженно и поддерживать друг друга. Если на практике между этими компонентами возникает конфликт или несогласованность экспертных сообществ, группы разработчиков и государство должны стремиться к их устранению.

Мы сосредоточимся главным образом на втором и частично на третьем принципах разработки надежного ИИ. Нашей задачей является анализ этических аспектов разработки и внедрения ИИ, а также анализ его социального и антропоразмерного потенциала. Настоящая работа преследует три цели: *описательную, аналитическую и рекомендательную.*

В рамках первой цели мы предлагаем схематичный обзор возможных применений ИИ в судебной системе и определенном от-

раслевом сегменте (образование, здравоохранение, ЖКХ). Данный обзор ни в коем случае не претендует на полный охват существующих и разрабатываемых технологий ИИ. Он нацелен на подбор иллюстративного материала для ситуаций, с которыми могут столкнуться управленцы-практики в выбранных нами отраслях, анализ которого поможет решить текущие и спрогнозировать дальнейшие этические затруднения при внедрении технологий ИИ.

Вторая цель, аналитическая, является центральной для данного руководства и указывает собственно на этически проблемные стороны внедрения и применения ИИ в определенных отраслевых сегментах и судебной системе. Согласно аналитическому замыслу раздела мы не ограничиваемся одной лишь констатацией возможных проблем, но показываем логику их формирования, что должно помочь обнаружению и прогнозированию аналогичных этических трудностей работы технологий ИИ и в других отраслевых сегментах, которые не вошли в данную книгу.

Наконец, рекомендательная цель книги заключается в описании критериев улучшения работы ИИ в случае возникновения этических затруднений и проблем при его внедрении. Реализация этой цели не предполагает формирования полного и детализированного перечня управленческих практик, в которых ИИ-технологии способствуют совершенствованию отраслевой деятельности, так как постоянные инновации в разработках ИИ быстро делают такую работу неактуальной. Нам важнее указать на общие типы проблем при этической оценке ИИ, которые можно в дальнейшем распространять на новые ситуации.

В свою очередь, для реализации рекомендательной цели на более системном уровне требуется отдельный документ, подготовленный на основе данной работы соответствующей группой специалистов. Настоящая книга состоит из четырех частей, в каждой из которых описан соответствующий блок отраслевых инноваций ИИ, дан аналитический разбор потенциальных несовершенств этического характера и пул возможных рекомендаций по их устранению.

Книга построена так, что разбор ключевых сложностей внедрения ИИ в различных отраслевых сегментах дан поблочно (на мате-

риале конкретных кейсов). В начале книги приведен общий анализ прикладных проблем, связанных с этикой разработки и внедрения искусственного интеллекта, затем изложены руководящие принципы и ценности этически ориентированного искусственного интеллекта и далее анализ этических проблем внедрения по секторам. Каждая глава посвящена отдельной отрасли и делится на две части: в первой дан собственно обзор этических аспектов применения и внедрения ИИ и, во второй, пул рекомендаций по возможному их устранению или минимизации. Таким образом, текст настоящей работы имеет выраженно прикладное и рекомендательное назначение. В книге выделяются четыре основные сферы, где использование ИИ требует решения ряда этических проблем:

1. *Судебная система. Конкретная правовая и судебная практика.*

2. *Образование. Отслеживание успехов учащихся и целевая помощь в планировании образования и трудовой карьеры.*

3. *Здравоохранение. Медицинская помощь гражданам, комплекс диагностических процедур, системы мониторинга здоровья и терапии.*

4. *ЖКХ. Системы «Умный дом».*

Мы надеемся, что помимо студентов, аспирантов, представителей академического сообщества наша книга будет полезна государственным управленцам, экспертам, которые на практике сталкиваются с различными барьерами при отраслевом внедрении технологий ИИ, а также осуществлении государственного отраслевого регулирования согласно этическим принципам и верховенству закона.

Руководящие принципы и ценности этически-ориентированного ИИ

Индустрия ИИ и робототехники в современном мире стремительно развивается, хотя многие вопросы взаимодействия ИИ с обществом остаются либо принципиально нерешаемыми, либо неуспевающими за технологическим развитием. Революция робототехники обещает массу преимуществ, но, как и в случае с другими новыми технологиями, она сопряжена с рисками и новыми вопросами, с которыми общество должно столкнуться. Использование роботов должно соответствовать законодательным нормам и этическим принципам, но сами эти нормы и принципы могут быть не до конца выяснены в области взаимодействия с ИИ. Социальная рефлексия нередко медленно «догоняет» технологические инновации, что приводит к образованию «этического вакуума» [Moor, 1985]. Такой эффект означает, что ИИ-технологии развиваются по своим собственным законам, и это приводит к ряду этических коллизий и противоречий. Чтобы их избежать, необходимо провести ревизию существующих этических лакун в области использования ИИ и наметить пути их устранения. Искусственный интеллект все еще находится в зачаточном состоянии с точки зрения сопутствующих его развитию этических исследований.

Существует *две взаимосвязанные проблемы* использования ИИ с точки зрения этики. *Первая* — согласование работы ИИ с существующими в обществе ценностными установками. *Вторая* — формализация данных ценностных установок. Большинство программ ИИ ориентированы на нейтральный анализ данных. Однако для ряда задач, связанных с оценкой человеческой деятельности, это невозможно, противоречит законодательству или неэтично. Отсутствие ценностной составляющей лишает смысла работу программы для человеческих нужд. Одним из естественных способов минимизации риска причинения вреда программами ИИ (например, роботами) является их программирование в соответствии с нашими законами или этическим кодексом. Однако здесь перед

нами встает ряд трудностей: законы могут быть расплывчатыми и контекстно-зависимыми, а сами роботы могут оказаться достаточно сложными для программирования в соответствии со всеми правовыми и этическими нормами. Более того, в силу ограниченности ценностных выборов конкретной ситуацией, бесконечность вариантов решений, с которым работает программа ИИ, может всерьез затруднить выработку этического решения.

Значительное число этических вопросов посвящено рискам дискриминационной практики алгоритмов, которые воспроизводят или даже усиливают враждебные настроения в обществе. Основное внимание должно быть уделено понятию дискриминации, как индивидуальной, так и коллективной, чтобы заложить основу измерения дискриминационной предвзятости, инструментов ее выявления и возможной коррекции. Например, постановление Европейского парламента 2016/679 от 27 апреля 2016 года [Regulation (EU), 2016/679] строго регулирует сбор персональных данных (религиозного, политического, сексуального, этнического характера и т.д.) и запрещает лицам, ответственным за алгоритмические решения, принимать их во внимание при автоматизированной обработке. Следует также уделить особое внимание уязвимым группам населения, таким как дети и инвалиды, а также другим лицам, которые исторически находились в неблагоприятном положении, подвергались риску изоляции или оказывались в ситуациях, характеризующихся асимметрией власти или информации (между работодателями и работниками, производителями и потребителями).

Для такого класса задач характерна проблема формализации собственно человеческих решений. Люди не являются полностью рациональными агентами. Вместо того чтобы рассчитывать свои действия по максимизации суммы функций полезности, мы зачастую имеем конкурирующие намерения. Иногда эмоциональные реакции не позволяют действовать рационально. Не все человеческие решения оказываются безупречными при их этической оценке. Следовательно, возникает более сложный вопрос: как выработать и формализовать этические принципы для ИИ. Некоторые

исследователи занимались вопросом их формализации [Hibbard, 2001; Yudkowsky, 2004; Muehlhauser, Helm, 2012]. В данных исследованиях ключевым вопросом было то, какие можно найти пути для преодоления противоречий между четкостью вычислительных алгоритмов ИИ и неоднозначным, непоследовательным, субъективным разнообразием человеческих ценностей. Например, иногда одни исследователи предлагают основывать системы искусственного интеллекта на нравственности, но при этом не дают точного объяснения того, как агент по искусственному интеллекту должен выбирать действия, последовательно базирующиеся на ней [Haidt, Kesebir, 2010].

Существует целый класс этических проблем, связанных с этикой предсказуемости. Множество программ ИИ пишется для решения прогностических задач. Для ряда программ (например, в области медицины) уровень их эффективности достаточно высок и продолжает расти. Однако при непосредственном взаимодействии человека и программы на стороне человека могут возникать коллизии этического характера. В частности, не ясно, в какой степени пациенты должны и желают знать исход протекающего заболевания, риски возникновения новых заболеваний, характер новых болезней и проч. Также не вполне ясно, как должно регулироваться поведение лечащих врачей в ситуации такой осведомленности и широких возможностей прогнозирования, которая обеспечивается ИИ.

Точно так же программы по тестированию психологической совместимости (например, супружеских пар) могут иметь высокую предсказательную силу. Агент ИИ на основе уже известной информации о человеке может моделировать ценности и поведение людей, которые он наблюдает в течение достаточно длительного периода времени и лучше человека может предсказывать результаты выбора разных опций. Но последствия такого взаимодействия между ИИ и человеком представляют собой этическую трудность.

Также различные программы тестирования способностей студентов или планирования профессиональных и карьерных портфолио для абитуриентов сталкиваются с проблемой готовности пользователя программы ознакомиться с результатом. В последу-

ющем могут возникать различные виды субъективного дискомфорта, связанного с «программированием» выбора или эффектами снижения мотивации и проч.

Решение данных вопросов осложняется тем, что формализовать человеческий запрос не всегда возможно. Ряд авторов [Muehlhauser, Helm, 2012] пришли к выводу, что люди не могут точно описать свои собственные ценности. Часто мы можем наблюдать у людей противоречия в их ценностной системе или отсутствие однозначных ответов на этические вопросы. Решения, принятые быстро, могут разительно отличаться от решений, которые долго продумывались или обсуждались. Ценности также могут изменяться в зависимости от пережитого опыта, под влиянием ближайшей среды и т.д.

В целом «ценностные» установки большинства программ ИИ на сегодняшний день — оптимизационные (или утилитарные). Таким образом, этический Кодекс ИИ — прагматический.

В связи с этими установками возникает масса проблем. Например, при выполнении инструкций по максимизации прибыли «Максимально увеличить прибыль, но никому не навредить» — для ИИ остается неясным, что имеется в виду под вредом. Идет ли речь о том, что не стоит заключать жесткие сделки, или не следует продавать никому то, что им не нужно? «Размышляя» над этими вопросами, ИИ может прийти к выводу, что в случае соблюдения всех этих правил, акционеры, сотрудничавшие с компанией, могут потерять свои деньги, что, действительно, причинит им вред [Raphael, 2009].

Есть и другой аспект. Поскольку рациональный агент выбирает действия для максимизации ожидаемой суммы функций полезности, он может выбирать действия, которые не увеличивают полезность в краткосрочном периоде, но в целом увеличивают его способность повысить полезность в будущем. Такие действия могут включать в себя самозащиту, увеличение собственных ресурсов и возможностей агента.

Если углубляться в детали этих компромиссов, то придется пересматривать этическую модель ИИ. Однако проблема не только в том, что недостаточно провести изменения в инструкции высокого

уровня (такой как «максимизация прибыли»). Не совсем ясно, как сформулировать инструкции для ИИ так, чтобы сбалансировать двусмысленность максимизации прибыли с минимизацией вреда. Прибыль — это ценность, связанная с результатами, но существуют и другие способы определения ценности результатов, которые в более общем плане влияют на благосостояние человека [Arkin, 2013].

Процесс человеческой оценки тех или иных ситуаций зависит от слишком многих правил, условий и влияния среды, чтобы его можно было сформулировать через перечень конкретных аналитических высказываний. Существует мнение, что в силу отсутствия таких правил или точных способов их поиска [Tasioulas, 2017], более успешной стратегией является использование статистических данных для решения различных дилемм (как люди и с какой частотой поступают в конкретных ситуациях). Это было бы более эффективным способом взаимодействия с ИИ, однако потребовалась бы огромнейшая база данных и сложные алгоритмы ИИ для такого масштабного исследования человеческих ценностей статистическим путем. Это создает проблему курицы и яйца для этического ИИ: изучение человеческих ценностей требует мощного ИИ, но этический ИИ требует знания человеческих ценностей.

Таким образом, в масштабах проблемы этической оценки ИИ появляется дилемма. Если мы могли бы запрограммировать кодекс этики для регулирования поведения роботов, какую этическую теорию мы использовали бы? Или же роботы должны рассматриваться исключительно инструментально (как оружие, компьютеры и т.д.) и регулироваться соответствующим образом?

На эти вопросы предстоит ответить как широкой общественности, так и узким специалистам, занимающимся разработкой ИИ. Важную роль в экспертной подготовке решений поставленных проблем должны играть специальные комитеты по этической экспертизе разработки и внедрения того или иного продукта ИИ, куда будут входить специалисты из междисциплинарных областей. В ходе открытого обсуждения и свободной дискуссии ими будут вырабатываться решения по тем или иным этическим вопросам внедрения и успешной реализации ИИ.

Ключевые этические рекомендации, лежащие в основе настоящего документа, выстроены на основе определенных принципов и ценностей, которые мы последовательно артикулируем в этой части. Таким образом, настоящий документ основан на следующих этических принципах и ценностях.

1. *Социальные принципы*:

а. Уважение автономии человека на основе равенства прав и доступа к реализации потребностей в рамках закона.

б. Справедливость в распределении благ, социальная и правовая справедливость.

в. Право знать, как устроены те или иные социальные механизмы, в том числе включающие ИИ-технологии.

г. Отсутствие любых форм дискриминации или ущемления прав граждан, внимание к наиболее уязвимым группам граждан.

2. *Экологические принципы*:

Сохранение природы, разумное использование ресурсов, техническая надежность и безопасность.

3. *Правовые принципы*:

Верховенство закона, неприкосновенность частной жизни, надзор за законностью применения новых технологий и использованием старых.

4. *Информационные принципы*:

Прозрачность, объяснимость, доступность и открытость информации.

Системы ИИ должны функционировать в соответствии с этическими ценностями обществ, в которых они будут реализованы. Несовпадение этих значений между человеком и машиной может препятствовать их эффективному взаимодействию.

Вопрос о том, какие ценности машины должны использовать и как внедрить эти ценности в научном поле, до сих пор остается дискуссионным. В настоящее время рассматривается несколько влиятельных этических теорий, регулирующих взаимодействие человека и ИИ (деонтологические, утилитарные, этики добродетели и т.д.).

На этом уровне мы находим два типа взаимосвязанных проблем:

1. Каким образом решать этические проблемы, возникающие при использовании ИИ?

2. Как формализовать выработанные этические стратегии для самого ИИ?

Способствовать решению данных проблем может разработка руководства ИИ по этике, которое может быть адаптировано к конкретным профессиям и реальным сценариям. В нем будут изложены основные принципы и ценности, которые должны функционировать в системах искусственного интеллекта. Они могут основываться на его способности усваивать примеры и динамически адаптироваться к реальным ситуациям, с которыми придется столкнуться в определенной области.

Отдельно обозначим фундаментальные этические принципы работы ИИ:

1. ИИ не должен вредить людям своими действиями или бездействием.

2. ИИ обязан выполнять приказы людей, только если это не вступает в противоречие с п. 1.

3. ИИ должен стремиться продолжать свое существование и поддерживать его, кроме тех случаев, когда это противоречит п. 1 или п. 2 [Pei, 2018].

ЭТИЧЕСКИЕ АСПЕКТЫ ПРИМЕНЕНИЯ И ВНЕДРЕНИЯ ИИ В СУДЕБНОЙ СИСТЕМЕ

1. Судебная система

Активное применение и распространение цифровых технологий не обошло стороной и такие сферы, как юриспруденция и судопроизводство. Применение искусственного интеллекта в данных сферах вызывает не меньшее число опасений, чем в области здравоохранения или образования. Традиционно в данной сфере человеческий фактор представляет собой наибольшую трудность. Риски подкупа судей, предвзятости следователей, различного рода уловки защиты, равно как и обвинения, психологические факторы воздействия на суд присяжных и/или свидетелей и проч., все это служило поводом для движения в сторону формализации судебных решений, ограничения роли человеческого вмешательства. В рамках подобных начинаний развитие и применение ИИ в правовой сфере было встречено с определенным энтузиазмом. Передача судебных решений в компетенцию ИИ может позволить значительно уменьшить риски недобросовестных рассмотрений дел, а также различного рода воздействия на участников судопроизводства. Применение так называемого «электронного правосудия», в котором решение выносит «ИИ-судья» (AI-judge), упрощает и ускоряет весь процесс принятия решения, уменьшает финансовые и временные издержки сторон, облегчает ознакомление с материалами дела. При этом ИИ-судья олицетворяет идеал абсолютной непредвзятости, бесстрастности и неподкупности. Как иронизируют некоторые современные правоведы, «антропоморфный судья всегда уже был создан по образу ИИ-судьи» [Bhorat, 2017].

Кроме того, машинизация обработки большого объема материалов должна значительно оптимизировать процесс рассмотрения дел и вынесения судебных решений. По мнению специалистов, внедрение ИИ должно улучшить качество судебных приговоров, уменьшить сроки судопроизводства. Внедрение электронного документооборота и цифровизация механизмов передачи и анализа данных позволит оптимизировать работу судебных органов [Нагорная, 2019].

Отметим, что в России созданы серьезные предпосылки для цифровизации различных процессов в судебной системе. В настоящее время максимальный эффект цифровизации в российской судебной системе обеспечивается при использовании систем электронной подачи документов и организации онлайн-взаимодействия сторон спора и судей, видеоконференц-связи, автоматического протоколирования судебных заседаний, а также при автоматизации деятельности аппарата суда. В ближайшем будущем развитие информационных технологий с высокой вероятностью позволит отказаться от проведения очных слушаний по широкому перечню дел на небольшие суммы, а также по бесспорным делам [Кашанин и др., 2020].

Цифровизация судебной системы будет осуществляться путем предоставления гражданам и организациям равных возможностей по использованию цифровых решений, позволяющих найти в электронной базе нужное дело; дистанционно знакомиться с материалами дела; участвовать в судебных заседаниях в режиме онлайн; получать и отправлять в электронном виде необходимые документы. Это обеспечит вклад судебной системы в экономическое развитие страны, поскольку позволит своевременно разрешать споры экономических агентов в арбитражных судах, а также рассматривать иные категории дел, включая уголовные.

Между тем, несмотря на перечисленные преимущества в отношении применения ИИ в судебной системе, специалистами высказывается также ряд опасений. Большинство из них носит этический характер. Судебные ошибки, фактор дискриминации, ущемление свободы воли человека при прогнозировании совер-

шенного им правонарушения, неэтичность принятия во внимание биографических данных субъекта правонарушения, проблема иерархии отношений между человеком-судьей и ИИ-судьей, а также многие другие трудности этического характера дают основания для сомнений в некритическом применении ИИ в правовой сфере.

1.1. ИИ в области права

1. Одним из ключевых способов цифровизации судебной системы является введение компьютеризированных узлов обработки и принятия решений на всех уровнях судебно-правовой сферы. Речь идет о механизмах встраивания и инкорпорирования ИИ в цепочки судебных процессов или правовых споров. При этом подразумевается не линейный, но интерактивный тип взаимодействия между компьютеризированными и антропоморфными элементами системы. Отношения между людьми в данной сфере также предполагается регулировать с точки зрения цифрового обмена. В частности, после того как адвокат соберет материалы, они могут быть автоматически сгенерированы в исковое заявление, которое будет направленно в вышестоящие инстанции. Искусственный интеллект может быть подключен также на этапе отбора материалов. Адвокатские запросы могут формироваться с помощью соответствующего программного обеспечения. Между одним человеком и другим будет заложен механизм компьютеризированного оформления запроса и ответа. Ключевым узлом взаимодействия в заданных пределах будет отбор и анализ материалов, а также различные аспекты принятия решения, в частности подбор наиболее подходящего прецедента, решение о приобщении документов к материалам дела, подборка типовых решений по тому или иному делу и т.д.

Кейс. *В апреле 2019 года проходил VI Московский юридический форум «Российская правовая система в условиях четвертой промышленной революции». В рамках мероприятия Председатель Совета судей России Виктор Момотов в докладе об особенностях работы судебной власти в цифровизированном мире указал на то, что*

важнейшей особенностью процесса современной диджитализации является его повсеместность для социальной, экономической и политической сфер жизни. Право не становится исключением — так, по мнению Момотова, апробация новых технологий может позитивно сказаться на институте судопроизводства в отношении законности, независимости судей, гласности и других его основополагающих принципах [Нагорная, 2019].

Как правило, проекты моделирования ИИ для формализации решения правовых вопросов построены по принципу алгоритмической матрицы оптимизации решений. Это значит, что система нацелена на так называемый «честный», или «холодный», поиск минимально затратного решения при максимальной его полезности. Мораль или справедливость подобных суждений не определена заранее и зависит от способа усвоения правил (объективный критерий, который был выбран), а также от того, каким образом была сформирована выборка для принятия решения. Кроме того, как правило, в человеческой коммуникации столкновение идей позволяет человеку прояснить свою собственную позицию, осознать ошибки, даже если изначально неверный выбор был сделан осознанно. Но большинство машинизированных решений на сегодняшний день не инициирует самостоятельный пересмотр решения. В связи с этим возникает деликатный вопрос: как определить или с помощью каких «измеримых» характеристик перевести понятия лояльности, доверия, ответственности (справедливости, достоверности, подотчетности), применяемые к алгоритмическим решениям. Нерешенные вопросы, подразумеваемые в данном сегменте цифровизации правовой сферы, вытекают из трудности формализации целого спектра, собственно ценностных категорий. Смысловые и ценностные (семантические и аксиологические) аспекты коммуникации сопровождаются неопределенностью и при взаимодействии людей. Однако в случае человеческой коммуникации они многократно уточняются и заменяются в процессе интеракции. Компьютеризированное взаимодействие не предполагает оперативной коррекции в процессе уже заданного общения.

Отсюда возможны погрешности коммуникации с точки зрения представления интересов той или иной стороны. В частности, машинизированное решение может заключаться в оптимальном решении приобщения к исковому заявлению всех материалов дела, в то время как защищающий интересы подопечного адвокат-человек может ограничиться частичной выборкой. Нарушения в коммуникации могут возникать на протяжении всей цепочки взаимодействия между разными участниками процесса.

2. В настоящее время в судебной системе активно обсуждается вопрос о возможности полной замены судей системами ИИ. Аргументом в пользу такой замены, как было сказано выше, является избегание предвзятости, уменьшение коррупционной компоненты судопроизводства, устранение фактора дискриминационной погрешности, которая может зависеть от расы, пола, внешности, типа занятости или биографии человека. Идея замены человека-судьи на компьютеризированного исполнителя могла бы наиболее совершенным образом воплотить идею бесстрастности и независимости суда. Судебное решение устремится к максимальной формализации и объективности, согласно существующим нормам закона. Субъективный фактор в судейском решении должен быть сведен к минимуму. При таких характеристиках судебное решение в идеале должно сводиться к подведению любого частного случая (спора или коллизии) под реализацию общего закона. Судебное решение в этом случае будет четко формализованным предписанием. Фактически оно будет стремиться к четкому алгоритмическому выводу, где все значения пошагово прояснены (подкреплены законом и нормами права) и прозрачны для воспроизведения другим судьей при тех же вводных. При таком определении статуса судебных решений их полноценная компьютеризированная эмуляция должна оказаться делом техники. В идеале ИИ будет со стопроцентной вероятностью воспроизводить решение человека-судьи и наоборот. Наличие погрешностей же будет либо указывать на несовершенства человеческих решений (не принявших во внимание ряд объективных факторов или допустивших примесь субъективной оценки), либо на несовершенства самого ИИ, кото-

рый ошибся на семантическом (при установлении значения), но не логическом участке работы [Miller, Wolf, Grodzinsky, 2019].

Кейс. *В декабре 2018 года была принята Европейская этическая хартия о использовании искусственного интеллекта в судебных системах. Она стала первым международным актом, регулирующим использование ИИ в судопроизводстве [Нагорная, 2019].*

Кейс. *Специалисты из Университетского колледжа Лондона и Университета Шеффилда создали программное обеспечение, которое может предсказать решения по реальным судебным процессам с вероятностью, превышающей 70%. С помощью данной разработки удалось спрогнозировать ряд решений ЕСПЧ с точностью до 79%. В 2019 году с помощью этого же обеспечения ИИ-судья вынес решения по шести делам, пять из которых полностью совпали с решениями ЕСПЧ. В процессе анализа существа данных дел ИИ обработал около 600 дел с вынесенными и обнародованными судебными решениями, пока не пришел к собственному приговору [Taylors, 2020].*

Вопрос о полноценной замене решений человека-судьи с помощью ИИ является сейчас очень обсуждаемым и дискуссионным. Сама идея об ИИ в правосудии задевает важные этические вопросы, связанные с автономией и юридической ответственностью самих ИИ.

Серьезная проблема в сфере алгоритмизации правовых решений — баланс между принятой человеком ответственностью и решениями ИИ. Человека можно судить за нарушение чужих прав и несоблюдение возложенных на него обязанностей. В ситуации, когда судейские решения выносятся человеком, на судью также возлагается и определенная ответственность, а также необходимость соблюдения прав [Sharkey, 2019]. В случае нарушения судьей возложенных на него обязательств он должен понести определенное наказание. Например, судья несет ответственность за совершение судебной ошибки, однако неясно, в какой степени мы можем возлагать ответственность ИИ на судью. В связи с риском ошибок

ИИ может быть неясно, кто несет ответственность за любой возникающий ущерб. Законы об ответственности за продукцию на текущий момент часто возлагают ответственность на производителей (как это происходит через лицензионные соглашения с конечными пользователями в области программного обеспечения). Так, в случае производства военных роботов по всей цепочке поставок существует список субъектов, которые могут быть привлечены к ответственности: программист, производитель, группа правовой экспертизы оружия, офицер по военным закупкам, куратор робота и т.д. Некоторые специалисты полагают, что по мере того как роботы станут более автономными, ответственность можно будет возложить на них. Если ИИ способен продемонстрировать достаточно функций, которые обычно определяют индивидуальность человеческого субъекта, то его можно будет считать ответственным лицом. Однако другие специалисты указывают, что ответственность не может возлагаться на ИИ в силу отсутствия у последнего спонтанной свободы и внутреннего сознания [Warwick, 2011]. Как следствие, ситуация, в которой человека, наделенного ответственностью, обязанностями и правами, судит лицо, предположительно не обладающее такими характеристиками, представляется не справедливой. Процедура, в которой человека судит машина, является этически нерелевантной по причине того, что оба участника судопроизводства, как судья, так и обвиняемый, должны быть одинаково подотчетны закону. Закон должен быть един и для обвиняемого, и для судьи. Согласно концептуальной рамке суда как процесса, все лица, включенные в него, должны обладать равным доступом к правам и возможностям реализовать свою ответственность. Иными словами, все участники судебного процесса должны быть субъектами права. Однако на сегодняшний день нет ясности в том, можно ли считать ИИ-судью субъектом права или, скорее, техническим помощником. Нет ясности в том, считаем ли мы подотчетным закону сам ИИ, его создателей или пользователей. В условиях отсутствия ясности в этом вопросе отношения между обвиняемым и ИИ-судьей выглядят неопределенными, а сам акт вынесения судебного решения ИИ неэтичным с точки зрения нарушения равен-

ства и паритета гражданских отношений. Возникает риск того, что инструменты ИИ, используемые в принятии судебных решений, могут идти вразрез с правом на доступ к судье, правом на равенство сторон в судебном споре и состязательность судопроизводства. В таком случае было бы правильнее говорить о вспомогательной роли ИИ-судей в работе судей-людей, которые единственно и остаются субъектами права. В последнем случае ИИ-судья может лишь помогать судье-человеку выносить судебные решения, но не может окончательно выносить их сам. Как вариант — ИИ может помогать людям или заменять их в судах низшей инстанции, но люди-судьи должны сохранять свое место в качестве конечных арбитров в любой правовой системе.

3. Одним из привлекательных принципов внедрения ИИ в судебную систему является его объективность. Ряд программ ИИ в судебной системе составлен из расчета нейтрализации дискриминационных ошибок. ИИ-судьи могут обойти известные факторы дискриминационного риска, использованные при человеческой оценке ситуации, а именно принятие во внимание возраста, расы, пола, происхождения, занятости, истории арестов, а также того, где человек живет.

> ***Кейс.*** *При тестировании одной из программ по оценке вероятности совершения террористического акта из совокупности личных данных были удалены сведения о поле, расе, национальной принадлежности, возрасте, религиозном вероисповедании и т.д. Вместе с тем в доступных для программы данных была сохранена информация о родном языке подозреваемых. Программа показала высокий риск совершения террористического акта у всех тестируемых лиц, родным языком которых оказался фарси [Floridi et al., 2018].*

Одна из этических проблем вынесения решений ИИ-судьями состоит в том, что в самых разнообразных базах данных по судебным приговорам практически всегда можно обнаружить дискриминационные компоненты. Очевидно, что в программах ИИ могут быть предрассудки, разделяемые их создателями-програм-

мистами и другими людьми, которые обслуживают функционал ИИ [Richardson, 2019]. Однако многие более усовершенствованные программы, основанные на нейросетях, способны к обучению и самостоятельному извлечению информации из массива данных. Хотя такие программы могут частично воспроизводить определенные человеческие предрассудки (косвенно через обработку данных, содержащих тот или иной градиент дискриминационных значений), важной особенностью нейросетей является то, что они могут вести себя спонтанно и непреднамеренно. По мере своего обучения они создают собственную сетку смыслов и ценностей. Однако нельзя быть уверенным, что при спонтанном самообучении машина не выработает дискриминационный алгоритм. Это может быть как результатом случая, так и закономерным итогом обработки некоторых данных. Например, если при обработке данных машина заметит, что чернокожих правонарушителей осуждали чаще или решения об оформлении опекунства чаще принимались в пользу женщин, то для ИИ это может сработать в качестве паттерна, по которому он обучится. Данный паттерн ляжет в основу будущих решений нейросетевого ИИ и будет определять его дальнейшее самообучение. Следует понимать, что дискриминация как социальный и этический феномен содержит сложный смысловой и ценностный компонент, который едва ли можно формализовать. Семантический анализ предполагает изначальный ввод ценностной матрицы, однако если данный ввод сам по себе является предвзятым, «прописанным» под определенный интерес или нерепрезентативным по отношению к населению, то это будет источником деформации анализа и прогнозов. Алгоритмизированные системы проверки данных при этом будут воспроизводить решение или даже усиливать предвзятость и, следовательно, дискриминацию. Для ИИ любой принцип выделения может прочитываться как дискриминационный и, наоборот, ни один дискриминационный аспект может не противоречить обычной процедуре выделения. Например, фильтрация дел по годам или странам ничем не отличается от фильтрации дел по возрасту или по полу [Burr, Cristianini, Ladymann, 2018]. Блокировка фильтрации по определенным пара-

метрам, однако, не исключает косвенного извлечения фильтрующей информации из других, разрешенных источников. В частности, высокий процент правонарушений в определенный год или в определенный стране за фиксированный промежуток времени, скорее всего, будет использован ИИ против обвиняемого, который будет иметь отношение к данной стране или году. Поэтому искоренение риска дискриминационных решений становится наиболее трудоемкой этической задачей на пути применения ИИ в рамках правовых решений [Asada, 2019]. Поскольку избежать любой дискриминационной предвзятости по отношению к меньшинствам и уязвимым группам сложно в связи с непредсказуемостью процесса самообучения ИИ, то следует уделять особое внимание механизму обратной связи. Если некое статистическое или вероятностное решение было принято системой, то она должна уметь отчитываться об этом и объяснять понятным образом.

4. Один из принципов использования ИИ в судебной системе — оценка рисков. Так, одна из систем ИИ, используемая судьями США, называется оценкой общественной безопасности. Поскольку ИИ в состоянии обрабатывать огромное количество статистических данных, как и способен применять сложные расчетные операции, недоступные человеку, то обработку любых данных целесообразнее поручать программам. Одним из важных и полезных результатов такой обработки является возможность предсказаний на основе анализа данных. В судебной системе оценка рисков применяется в целом ряде случаев — для профилактики правонарушений, оценке степени опасности правонарушителя с точки зрения будущего, оценка степени возможного ущерба, а также при определении степени наказания. Один из таких инструментов по оценке рисков был разработан частным благотворительным фондом Лауры и Джона Арнольдов, базирующимся в Техасе. На основании обработки большого массива статистических данных указанная программа предоставляет судьям информацию о потенциальных рисках рецидивов со стороны обвиняемых, а также выдает рекомендации для принятия справедливых вердиктов в отношении заключенных [Brożek, Jakubiec, 2017].

Кейс. *В ряде американских судов успешно внедрена ИИ система освобождения под залог. Согласно традиционному применению этой меры, судья устанавливает достаточно высокую сумму залога для обвиняемых во избежание возможного их уклонения от явки в суд.*

ИИ стал применяться для расчета залоговой суммы, при установлении которой ИИ проводит оценку риска неявки. Для этого ИИ использует данные о прошлых судимостях и/или арестах, сведения о различных штрафах, административных нарушениях, просрочке кредитов, а также уровне образования, доходов, престижности района проживания, семейном статусе и проч. На основании большого числа вводных биографических данных и личных сведений о человеке производится оценка риска его неблагонадежности и степени законопослушности. В соответствии с полученным коэффициентом законопослушности ИИ и устанавливает размер залоговой суммы. В рамках данной системы, например, в судах Нью-Джерси оценка рисков активно применяется судьями для вынесения собственных вердиктов. На первом слушании, где из тюрьмы обвиняемые предстают перед судом при помощи видеоконференции, оценка риска предоставляется судье. Обвиняемые с более низкими оценками освобождаются прямо из зала суда до следующего судебного заседания.

Программа компьютеризированной оценки рисков правонарушений при принятии судебных решений является одной из самых этически болезненных. Ведутся споры о том, является ли этичным оценивать активность человека в будущем исходя из его прошлых или настоящих действий. Согласно одному из аргументов против такой оценки мы не можем рассчитывать на полноту данных о действиях и поступках человека, чтобы иметь право прогнозировать его дальнейшее поведение. Та выборка, которой может располагать соответствующее лицо, выносящее оценку (например, судья или его помощник), будет либо не полной, либо предвзятой, либо составленной субъективно. Подобного рода оценки признаны не слишком информативными в силу того, что разные судьи на свое

усмотрение принимают во внимание разные факторы. Как правило, при составлении подобных прогнозов человеком во внимание принимается только совершение или не совершение того или иного проступка. Совершению или не совершению приписывается определенный балл. Противники подобного подхода указывают на его неэтичность — не вполне справедливо оценивать только совершение дурных поступков или их несовершение, но не принимать во внимание совершение положительных поступков [Coeckelbergh, 2016]. Создание компьютеризированных программ оценки рисков нацелено на устранение данной погрешности, поскольку отныне в силу большой подборки данных программа может принимать во внимание практически все действия человека, совершенные им от рождения до текущего момента [Giubilini, Savulescu, 2018]. Согласно мнению некоторых из сторонников внедрения подобных программ, обширная подборка релевантных данных об активности человека в прошлом должна способствовать наиболее объективной оценке его законопослушности, равно как наиболее объективно оценивать вероятность новых правонарушений в будущем. В подборку по оценке рисков может быть помещен анализ генетических и социальных предрасположенностей к совершению преступлений, равно как потенциальной склонности к противоправному поведению. Сюда же могут включаться данные о наличии социопатических черт, анализ биографических и психологических особенностей личности с возможностью моделирования и предсказания возможных рисков противоправных действий, к совершению которых может быть склонно то или иное лицо.

Подход к формированию подобной оценки представляется не вполне этичным. Для этики в высшей степени проблематичным представляется оценка поступков человека на основании математической статистики. Презумпция свободной воли предписывает закреплять за человеком возможность совершить свободное, а значит, непредсказуемое действие [Holm, Voo, 2011]. Если человеческое действие будет рассчитываться на основе одних только статистических вычислений, это будет равнозначно тому, что он не обладает свободой воли. В свою очередь, наличие свободной воли

необходимо для того, чтобы считать человека субъектом морального и правового действия. Лицо, не наделенное свободной волей и не способное к совершению свободного поступка, не может осуждаться или порицаться, подпадать под юридические или моральные оценки. Применение юридических санкций к лицу, не обладающему субъектными чертами, важнейшей из которых является свобода действия, абсурдна. Соответственно, использование любых расчетов, которые могут прямо или косвенно указывать на отсутствие такой свободы, нарушает требования этики, согласно которым мы должны следовать презумпции существования свободной воли, даже если у нас имеются данные в пользу ее ограниченности. Применение прогнозов, основанных на анализе большой выборки данных о действиях человека, нарушает его право на то, чтобы рассчитывать на данную презумпцию [Livingston, Risse, 2019]. Если в ходе анализа большой выборки фактов о деятельности того или иного лица, ИИ действительно сможет предложить высоковероятный сценарий его последующей активности, то применение данных результатов будет научно релевантным, но этически неоправданным.

5. Еще одной технологией ИИ в области судебной системы являются автоматизированные сбор, аккумуляция и сопутствующая юридическая экспертиза данных, касающихся граждан. Управление данными, обрабатываемыми ИИ, — это определенная этическая и юридическая проблема. Преимущество ИИ заключается в том, чтобы собирать большие объемы данных, используя различные устройства и приложения, как в случае с беспилотниками, роботами-помощниками и роботами по безопасности. Неясно, однако, как оценивать технологии обработки данных, связанные с судебным разбирательством и судебным решением. Согласно данной технологии, о каждом гражданине можно будет сделать централизованный запрос, в котором будут указаны специфические особенности, например, история уплаты налогов, история погашения штрафов, история и обстоятельства погашения возможных административных или иных правонарушений. Система также будет аккумулировать и хранить также данные о возможных обращениях

гражданина в суд, полицию или иные правоохранительные органы или органы власти. В данном профиле могут быть сведения об обращениях в банк по вопросу хищения денег с карты, потери карты и проч. Система будет оснащена встроенным аналитическим модулем, который не просто собирает информацию, но и аналитически обрабатывает с точки зрения различных запросов. Например, система может предоставлять информацию о статистике тех или иных ситуаций, возникших в определенный период жизни, или сообщать о росте/снижении административной ответственности в определенный период времени. Данные технологии могут оказать посильную помощь при возникновении правовых споров.

Кейс. *Программы VapeNews, Focus.kontur.ru, Avinfo.co в России предоставляют данные о гражданах в открытом доступе по различным тематическим запросам (предпринимательская деятельность, сделки по автомобилям и недвижимости, активность в Интернете и т.д.).*

Кейс. *В 2015 году в США программа сбора данных о гражданах Америки на основе пользования мобильной связью, интернетом, электронной почтой прекратила свое функционирование. Данное решение было принято на основании признания несоответствия данной программы положениям Закона США «О свободе». В рамках данной программы Агентство национальной безопасности (АНБ) США собирало записи сообщений миллионов своих граждан, независимо от того, подозревались ли они в совершении правонарушений или нет. Весной 2015 года Федеральный апелляционный суд США признал существующую программу сбора сведений о гражданах незаконной. Вскоре ее заменили на более легкую версию, которая работает по настоящее время, однако, представляет собой предмет неутихающих споров и дебатов этического свойства.*

Закрепленное законом право на неприкосновенную частную жизнь любого гражданина вступает в явное противоречие со сбором и анализом личных данных в случаях, если это происходит без

ведома и согласия владельца. Существующий в настоящее время закон регулирует процедуры обработки личных сведений о гражданах, однако, неурегулированным остается ряд вопросов. Интернет-программы, осуществляющие автоматический поиск данных, работают в условиях открытой информации, на доступ к которой физические или юридические лица не давали своих согласий. Например, стандартная база данных публикационной активности ученых выдает автоматизированный наукометрический коэффициент научной значимости публикаций, а также личный научный рейтинг исследователя (индекс Хирша). Данный показатель рассчитывается автоматически и не предполагает получения письменного согласия ученого о разрешении на поиск данной информации в открытом доступе. Между тем данный индекс сообщает о степени квалификации ученого, его профессиональном уровне, научных достижениях и карьерных возможностях. Данный показатель активно используется потенциальными работодателями, решающими вопросы научного рекрутинга. Этически открытым остается вопрос, допустимо ли размещение сведений о профессиональном уровне ученых в открытом доступе.

1. Еще одной из ключевых этических проблем, связанных с технологией сбора данных, является степень расширения доступности конфиденциальных сведений и их переход в область неконфиденциальных. Неясным является вопрос алгоритмизации отделения конфиденциальной информации от неконфиденциальной. Конфиденциальная информация может косвенно содержаться, даже без намерения ее искать, в неконфиденциальной информации и тем самым способствовать предвзятости, к примеру, судебного решения [Myungsik, 2018]. Потребительские привычки, мнения в социальных сетях, геолокационные данные дают информацию о привычках и склонностях человека и позволяют косвенно реконструировать конфиденциальные данные [Galliott, Reed, 2016].

Техническое определение границ конфиденциальной и неконфиденциальной информации представляет собой отдельную проблему. Доступ к информации лежит в основе функционирования агентов ИИ. Модели и алгоритмы прогнозирования разрабаты-

ваются на основе наблюдений за окружающей средой и анализа больших массивов данных. Чем больше данных будет проанализировано, тем более точные прогнозы будут подготовлены и большую полезность получат дальнейшие прогнозы [Rouvroy, 2010]. В этом случае агенты ИИ будут вести наблюдения, даже если все другие считают секретной информацию, которую ИИ может получить. Например, в некоторых интернет-программах цензура рассматривается как препятствие, и чтобы его устранить следует искать обходные пути [Totschnig, 2019].

Следствие подобной ситуации то, что сбор и обработка данных даже в случаях, когда их организация будет проводиться с соблюдением закона, может порождать этически напряженные ситуации, в которых о людях «всплывает» информация, которой не предполагалось делиться. В контексте судебных споров и коллизий подобная ситуация может служить причиной как недоразумений и судебных ошибок, так и определенных злоупотреблений. В любом случае непредсказуемость и непреднамеренность разглашения конфиденциальных данных может вредить гражданам не только в повседневной или профессиональной жизнедеятельности, но особенно в контексте судебно-правовой деятельности.

2. Если ИИ-технологии в судебной системе не закладываются как полноценные эмуляторы судопроизводственной деятельности, то они могут внедряться на правах технических помощников профессионалов (судей, прокуроров, адвокатов или следователей). Квалифицированные сотрудники, работающие в системе правосудия, могут обращаться к решениям ИИ-судей как к своего рода рекомендательной инстанции. С решениями компьютеризированных помощников можно сравнивать решение, принятое судьей-человеком. Отбор аналогичных случаев или подбор соответствующей правовой документации, осуществляемой в таких программах, как «Гарант-Плюс», можно дополнить предварительной экспертизой дела, а также вынесением рабочего решения. Профессиональный сотрудник, работающий с данной системой, может принять во внимание или отклонить предложенное программой решение. Однако в любом случае, учитывая рекоменда-

тельный характер предварительной информации, она должна быть полезной в оптимизации и ускорении приятия решения человеком-профессионалом.

> ***Кейс.*** *Такие программы, как «Гарант-Плюс» за последние 10 лет претерпели серьезное расширение и серьезную трансформацию с точки зрения запроса конечного пользователя. Если в 2000 году 80% запросов при использовании программы касались нахождения того или иного законодательного акта, постановления или нормативного акта, то в 2016 году около 60% запросов касаются поисков прецедентных решений или помощи в решении того или иного правового вопроса. Данная трансформация позволяет говорить о возрастании запроса юристов и прочих специалистов правоведов на эмуляцию решений конкретных правовых вопросов и ситуаций [Шевелёва, 2018].*

При применении данной технологии контроль за окончательным решением закрепляется за человеком. Специалисты, работающие в правоохранительной области и судопроизводстве, имеют возможность принимать или отклонять решения, принятые системой ИИ, и контролировать доступ к данным, используемым ИИ. Для сотрудника-пользователя организован прозрачный доступ к тому, каким образом было выработано решение ИИ. В связи с этим возникает вопрос, какова должна быть степень человеческого вмешательства в определение границ решений ИИ? Наиболее демократичным для человека-специалиста является режим полного контроля вплоть до наложения вето на те решения, которые принимает ИИ. Однако даже в таком случае открытым остается вопрос о степени влияния решений, полученных ИИ, на решения, которые может принять человек. Этически неоднозначным представляется возможность скрытого манипулирования человеческим вмешательством при отсылке к авторитетности компьютеризированного решения. Поскольку решения, принимаемые ИИ, имеют доступ к большим массивам данных, наиболее репрезентативной выборке прецедентов, статистической обработке и многофакторному ана-

лизу, то его решения могут представляться более квалифицированными. Поскольку вердикты, выносимые ИИ, подпадают под требования максимальной прозрачности и алгоритмической открытости шагов, то подобные решения могут выглядеть проработанными намного лучше. Это может приводить к тому, что судья-человек будет осознанно или неосознанно склоняться к редупликации решения, предложенного ИИ. Этическая дилемма будет связана не только с психологической инертностью и тенденцией следовать за авторитетом, но и нежеланием подвергать свою репутацию риску дополнительной проверки. Пересмотр решения, предложенного ИИ-судьей, будет означать своего рода тест на квалификационное соответствие человека-судьи своему званию. Во избежание дополнительных проверок и публичного подтверждения собственного профессионализма судьи могут некритически следовать за судейскими решениями ИИ. Еще одной этической проблемой в связи с этим будет риск формирования предвзятых эмоций с помощью ИИ. Социальные сети уже экспериментируют со способами манипулирования эмоциями пользователей [Goel, Joyner, 2014]. Возникает вопрос, может ли судебное решение ИИ эмоционально повлиять на судью или, к примеру, на суд присяжных? От ИИ в судебном процессе потребуется не просто принятие решения, но и его аргументация. При самообучении ИИ по типу нейронных сетей программы будут иметь доступ ко всем судебным процессам, а именно линиям защиты или обвинения по сходным делам. Процесс аргументации в суде содержит известный корпус логических уловок и умышленных ошибок в рассуждении, таких как софизмы, ложные парадоксы, подмена тезиса, отсылка к авторитету, поспешное обобщение, гиперболизация, объяснение непонятного посредством другого непонятного, различные ошибки в умозаключениях по типу построения неверных силлогизмов. Весь корпус данных ошибок может использоваться ИИ для обоснования выбранной позиции, поскольку целевой регулятив программы заключается в максимальной оптимизации своих решений. Хотя предполагается, что ИИ функционирует в условиях ценностно нейтральной организации, большинство логических уловок могут быть использова-

ны в выбранной им линии аргументации. При отсутствии высококвалифицированной подготовки судей и прочих специалистов, задействованных в судопроизводстве, люди, которые причастны к контролю за решениями ИИ, будут подпадать под различные тактики, использованные ИИ. В связи с этим следует отметить специфические проблемы российской судебно-правовой системы. Так, с 2010 года из подавляющего большинства образовательных программ, готовящих будущих юристов, был исключен курс по логике, а также в ряде вузов курсы по риторике или ораторскому искусству. Это стало причиной того, что судебные специалисты не всегда профессионально владеют техниками построения аргументации. В данной ситуации осуществление экспертного контроля за той частью программы ИИ, в которой решение обосновывается и выбирается, может оказаться непосильным для человека. Сложной этической дилеммой в таком случае будет вопрос о степени, в которой человек сохраняет экспертный контроль и нейтралитет к решениям, предложенным ИИ. Реалистичным может оказаться такой сценарий, в котором человек эмоционально и интеллектуально подпадает под логику, представленную ИИ-судьей.

3. Аспект объективности, который обычно связывают с преимуществом внедрения ИИ, позволяет предложить технологию, в которой ИИ-судьи и сопутствующие их решениям механизмы сортировки, обработки и анализа данных максимально ориентированы на формальный индекс юридических понятий и принципов. Программы в состоянии отфильтровать культурные и лингвистические особенности рассмотрения дел, наиболее распространенные в данной культуре, стране или регионе. Применение подобных технологий должно упростить рассмотрение дел в международных судах, а также не ограничивать ИИ-судейство рамками отдельных стран. Цель данных программ — максимальная формализация судопроизводства, упрощающая его использование в различных ситуациях.

Кейс. *Группа ученых из Университетского колледжа в Лондоне, которой руководил Николас Алетрас, пришла к выводу, что судьи ЕСПЧ чаще обращают внимание на описание фактов, а не на спе-*

циальные юридические термины в текстах обращений. При анализе дела, связанного с возможным нарушением прав человека, судьи обращают внимание на слова «душность», «переполненность», «прогорклость», «нагноение» и т.д. Эти выводы соотносятся с предыдущими исследованиями практики судов других уровней, в том числе Верховного суда США. После анализа судебных приговоров исследователи сделали вывод, что судей чаще волнует моральная сторона дела [Aletras et al., 2016].

При использовании данных технологий возникают некоторые этические осложнения.

Их применение отягощается в случае такого типа судейства, при котором этические и моральные нормы так или иначе вводятся в закон. В частности, ряд современных американских судей полагают, что при рассмотрении дел должны учитываться текущие моральные нормы, в том числе религиозные и церковные. Однако программирование такого ИИ, который был бы чувствителен к подобного рода ценностным величинам, является крайне сложной задачей. Кроме того, анализ тех значений, которые не типичны для словаря юридических терминов, многократно усложняет применение формализованных судейских решений. В свою очередь, отказ от ценностных и «обыденных» слов или способов описания ситуаций может привести к невосприимчивости ИИ к собственно этической природе правовых споров. Между тем в том типе судопроизводства, который применяется сейчас в большинстве европейских государств, так или иначе сохраняется значение ключевых моральных кодексов и таких философско-этических категорий, как справедливость, честность, сострадание, милосердие, благородство и проч. Однако формат «обучение» ИИ, который закладывается в большинство компьютеризированных систем судопроизводства, основан на принципе «lex искусственный». Такие программы не воспринимают не только трудно формализуемые понятия этического словаря, но и абстрагируются от культурных норм и особенностей истолкования этических понятий в различных языках и странах. В качестве решения данной проблемы предлагается

и в этом случае рассматривать ИИ-судей как помощников, которые могут заменять людей только в судах низшей инстанции. В качестве конечных арбитров на вершине любой правовой системы должны оставаться судьи-люди. Последние будут продолжать ориентироваться на этические и моральные ценности как релевантные нормы права. С технической точки зрения ИИ окажется полезным для актуализации собственно инструментальной части закона, однако именно люди будут ответственны за конечный выбор наилучшего решения.

1.2. Перечень рекомендаций

Принимая во внимание указанные выше трудности можно отметить, что программа судейства, основанная на технологиях ИИ, должна проходить экспертную проверку на различные этические недоразумения и дилеммы. Правовая сфера — одна из труднейших для цифровизации, поскольку эта область практически полностью построена на ценностной основе. Понятия справедливости, достоинства, ценности человеческой жизни, благополучия, морального ущерба и проч. представляют собой сложнейшие комплексы философских смыслов и идей. Их формализация — большая трудность. Возможно, на начальной стадии внедрения использование ИИ в области судопроизводства может быть ограничено оказанием помощи судьям в выявлении их собственных предубеждений и поддержании последовательности. По мере дальнейшего распространения внедрения ИИ в судебной сфере данные полномочия могут расширяться, однако при условии экспертного контроля со стороны общественности, профессиональных сообществ и, отчасти, государства. Необходимо также иметь в виду, что все более широкое использование автоматизации в судопроизводстве неизбежно повлияет на процесс принятия решений человеком.

Передача формирования судебных решений в ведение ИИ является неоднозначным шагом, что предполагает устранение не только технических вопросов, таких как информационная безопасность и защита от возможных взломов, но и решение вопро-

сов обучения ИИ с точки зрения этических требований. Этическая настройка технологий ИИ в судебной сфере требует нормативного контроля, социально-правовой экспертизы, общественного наблюдения и сотрудничества с узкими специалистами.

Отдельное внимание потребуется уделить параметрам защиты от избыточной формализации. При этом подразумеваются различные дискриминационные риски, а именно алгоритмизация оценки таких факторов, как этническое происхождение, расовая принадлежность, социальное и экономическое положение, политические взгляды, интенсивность и характер социальных связей, возраст, пол, уровень образования и проч. По-видимому, следует отдельно устанавливать определенные дискриминационные запреты в алгоритмах ИИ, используемых в судебной и правоохранительной системах. К ним относятся применение генетических и биометрических данных, данных, связанных с религиозными и политическими взглядами, сексуальной ориентацией. Если подобная дискриминация формируется самой системой ИИ, необходимо усовершенствование алгоритмов или подключение антропоморфного интеллекта. Принимая во внимание имманентно присущую технологиям ИИ (в особенности на базе нейронных сетей) возможность самообучаться, а именно фиксировать таксономию новых данных на основе библиотеки ранее обнаруженных данных, необходимо, чтобы общественные институты, специализированные исследовательские центры или государственные органы, еще в процессе создания ИИ-программ ограничивали возможность ущемления отдельных социальных групп или граждан в ходе применения программы. В целом важно держаться равновесия между частной инициативой разработчиков ИИ (их интеллектуальной собственности) и независимыми экспертными группами (как привлеченных, так и действующих независимо). Данное требование могло быть реализуемо каждым конкретным судом, однако контроль за его исполнением следовало бы возложить на специальные группы независимых судей, которые определялись бы выборным путем и регулярно переизбирались. Такая практика должна способствовать сохранению беспристрастности и прозрачности в разработках ИИ для су-

дебных и правоохранительных органов, а также их открытости для возможной проверки.

Создателям технологий самообучения программ необходимо создать все условия для использования профессионального опыта сотрудников правовых органов и учреждений. Помимо сотрудничества с профессионалами из тематически релевантной группы, такими как судьи, адвокаты, прокуроры, следователи и др., необходимо подключить специалистов различных гуманитарных профилей, в частности философов, в том числе аксиологов, социологов, психологов и др. Для успешной разработки ИИ в судебной сфере имеет смысл заняться организацией интегрированных проектных команд, ориентированных на междисциплинарный подход. В частности, при формировании госзаказов на разработку и внедрение программ ИИ для правовой области важным критерием выбора должно быть обучение по курсам этики ИИ с получением соответствующих сертификатов.

Полезной рекомендацией также является государственная сертификация и экспертная оценка производителей ИИ для судебной и правоохранительной сферы. В этом же ключе можно рассмотреть возможность организации внешнего аудита технологий ИИ. Мы предлагаем установить порядок, при котором будет соблюдена прозрачность алгоритмической базы и понятность всех решений, принятых ИИ. Данная мера важна для установления защиты от дискриминационных решений, любых типов злоупотреблений, а также возможности выявления ошибки на любой стадии принятия решения. Наиболее эффективной экспертами признается модель неограниченного доступа людей в решения ИИ, когда используется открытая документация и экспликативная модель алгоритмизации. Уже на стадии обучения ИИ программу следует снабдить навыком объяснения своих решений на метаязыке, отличном от объект-языка самого решения. Необходимость данного требования вполне объяснима — все участники судебного процесса (адвокаты, обвиняемые и проч.), в котором судьей выступает ИИ, должны иметь возможность оспорить решения, принятые ИИ-судьей. В дополнение к этому следует позаботиться о том, чтобы все участ-

ники судебного процесса получили доступ ко всем входным данным. Кроме того, участникам должно быть предоставлено право на апелляцию не только при участии судьи-человека, но и другого ИИ-судьи. Использование данной услуги может потребоваться, чтобы установить, насколько независимыми являются рекомендации ИИ-судьи, поддерживающего одну из сторон процесса (к примеру, сторону обвинения). Суммируя данные рекомендации, можно сказать, что в них подразумевается своего рода организация системы внешнего аудита ИИ-решений для судебной и правоохранительной сфер. Наиболее предпочтительно, чтобы подобная система аудита функционировала в качестве общественного контроля (независимых правозащитных организаций) и не зависела от производителей ИИ.

ЭТИЧЕСКИЕ АСПЕКТЫ ПРИМЕНЕНИЯ И ВНЕДРЕНИЯ ИИ В СУДЕБНОЙ СИСТЕМЕ В ОТРАСЛЕВЫХ СЕГМЕНТАХ

2. Образование

Образовательные системы различных стран, в том числе России, ориентированы на отбор и выявление молодых людей с определенными способностями и дарованиями. Эта задача стоит локально перед различными образовательными учреждениями, потенциальными работодателями, глобально — в масштабах страны и государства. Одним из приоритетных направлений применения ИИ в данной области является обработка больших данных при анализе наследуемости интеллекта и способностей. В образовании активно внедряются новые технологии для мониторинга эффективности его процессов и результатов. В настоящее время измеряются не только hardskills (знания, результаты использования знаний, которые проще всего квантифицировать), но также и softskills (социальные, коммуникативные, лидерские навыки и т.д.). Одним из результатов применения данной практики ИИ является создание персональных цифровых портфолио, отражающих карту наследования тех или иных компетенций [Nilsson, 1998]. Искусственный интеллект может анализировать успехи обучающихся и их образовательную траекторию.

В России уже реализуется ряд значимых проектов в сфере цифрового образования (подготовка к ЕГЭ, изучение английского языка, цифровые библиотеки учебников, заданий и тестов). Как правило, данные проекты ориентированы на создание платформ образовательных услуг, сводя в едином виртуальном пространстве преподавателей и обучающихся. Платформы анализируют

профиль обучающихся, предлагая им индивидуальные траектории обучения (Parla, Examer, Skyeng). Ряд платформ интегрировали большое количество электронных учебников и тестовых заданий (Yandex.учебник и Московская электронная школа). Также отметим проект по прокторингу Proctor.Edu, который позволяет с помощью камеры верифицировать сдающего тест или экзамен, а также формирует оценку доверия к действиям экзаменуемого в реальном времени, непрерывно наблюдая за действиями человека на компьютере перед камерой в режиме реального времени.

2.1. ИИ в области образования

> ***Кейс.*** *В 2014 году группа британских ученых провела исследование образования с использованием ИИ и выявила, что различия в успеваемости между учащимися связаны с их происхождением и семьей. Это наблюдается с ранних школьных лет и до момента завершения среднего образования. По окончании средней школы британские абитуриенты сдают по всей стране стандартные тесты (общий сертификат среднего образования GCSE) [Vincent, 2016].*

Такая технология, безусловно, должна способствовать повышению эффективности управления системами образования. Однако этические риски произвольного внедрения подобных технологий достаточно высоки. Они заключают в себе значительный дискриминационный потенциал, поскольку способствуют ослаблению концепции социальных лифтов, которая стимулирует социальную активность. Это ведет к негибкой стратификации общества, стигматизации определенных групп и членов общества, вульгаризации биологического языка в описании социальных процессов.

Работодатели или образовательные институции могут стремиться применять данные технологии, когда требуется принятие решения в отношении выбора сотрудников или отбора учащихся. Опора на данную технологию для проведения отбора содержит этические нарушения. В их основе лежит грубое нарушение пре-

зумпции свободы воли человека и его свободного потенциала к саморазвитию. Идея жесткой биологической детерминации всерьез девальвирует гуманистические принципы и обедняет представления о человеческой природе.

1. В условиях стремительного развития цифровых технологий, особенно в образовании, интегральная составляющая институтов образования все более развивается в направлении децентрализации [Coppin, 2004]. Это ведет к отказу от жестких образовательных стандартов, унифицированного разделения образовательного времени на семестровые и модульные курсы, а также унификации в преподавании самих курсов. Выбор делается в пользу коротких курсов, выполнения самостоятельной проектной работы, проработки кейсов, освоению краткосрочных программ, прохождению тренингов и т.д. Немаловажным аспектом цифровизации современного образования является объединение возможностей онлайн- и оффлайн-пространства при осуществлении многих курсов. Именно в этом направлении активно применяются ИИ-технологии. Жесткая закрепленность за конкретным образовательным учреждением и обязательность фактического нахождения в нем теряют свою актуальность. Процесс обучения становится виртуальным и глобальным, образование будет развиваться на основе индустрии онлайн-курсов.

Также развитие ИИ способствует изменению отношений между ключевыми субъектами образования: влияние конкретного педагога на отдельного студента уменьшается, и ключевым актором в образовательном процессе становится организатор образования. Консультанты заменяются онлайн-консультантами, те, в свою очередь, сменяются ботами или виртуальными тьюторами. Отдельные исследователи на основании социологических опросов приходят к выводу, что ИИ способен заменить репетитора или преподавателя в ближайшем будущем. Так, согласно данным опроса PwC, 58 % менеджеров и технологических экспертов уверены, что к 2022 году ИИ сможет заменить репетиторов в будущем [Rahwan, 2018]. Сложным этическим вопросом при этом является необходимость педагога не только для прикладных навыков, но и для транслиро-

вания человеческих ценностей. В случае такого радикального изменения его социального статуса и модификации программ обучения по оцифрованным (blended) направлениям гуманистический потенциал образования ставится под вопрос.

2. Еще одной технологией ИИ в области современного образования является геймификация образования. Данная технология — частный случай интеграции образования и развлечения. Согласно этой стратегии эффективность образовательного процесса повышается в условиях отсутствия четких границ между обучением и досугом. Включение игровых компонентов в образовательный цикл предполагается за счет технологий ИИ, а именно виртуальной реальности и игровой среды. Прохождение тестов проходит в форме разнообразных нетривиальных игровых заданий с большим содержанием визуального и акустического контента.

> ***Кейс.*** *В рамках запущенного на Нацплатформе курса по «Философии» были апробированы тестовые задания в форме игр, в которых требовалось найти правильный ответ после прослушанной онлайн-лекции. Студенты различных факультетов НИУ ВШЭ неплохо справлялись с разработанным блоком заданий, и он получил высокий рейтинг при отборе заданий к курсу. В итоговом опросе студентов им был задан вопрос: почему они выбирали один, а не другой ответ; на каком основании формировался выбор. Подавляющее большинство указало, что ответ выбирался только на основании сличения дословно произнесенной лектором информации с той, которая задавалась в тестовом задании. Таким образом, слушатели успешно справлялись с заданием, не понимая «смысла» вопроса и ответа.*

Активное внедрение игровых ИИ-технологий, основанных на различных виртуальных эффектах, обсуждается различными специалистами из области психологии образования. Определенные успехи и высокая результативность применения данных технологий (скорость и легкость освоения материала, положительное отношение слушателей к процессу научения, привлекательность об-

разовательных продуктов и многое другое) способствует поддержке и продвижению данных технологий на рынке образовательных услуг. Вместе с тем данные технологии сопровождаются критическим обсуждением, касающимся этической стороны. В частности, поднимаются вопросы об оправданности идеи стирания границ между игрой и обучением. Подобная идея может быть истолкована с точки зрения дидактической полезности любой игры. Если с помощью компьютерных игр можно обучать, будет ли это означать, что обучает любая компьютерная игра? Кроме того, аспекты геймификации характеризуются высоким потенциалом формирования толерантности [Feil-Seifer, Matarić, 2010]. Специалисты отмечают проблематичность присущей игровым технологиям обучения высокой соревновательности, которая может формировать легковесный или узкоцелевой образ получения образовательных навыков. Дискуссионным также остается вопрос о том, в каком соотношении находятся игровые компоненты и способность к обучению. Игровое обучение построено на спонтанных реакциях, автономных и нерефлексивных процессах запоминания и формирования навыков. Процесс освоения цифровых и компьютерных технологий на стороне пользователя является интуитивным. Однако подобные методы освоения не заменяют рефлексивных и сознательных процедур направленного поиска решений, в котором обучаемый прилагает осознанное усилие. Этический аспект описанных трудностей сводится к тому, что обучаемый пассивно вовлекается в процесс, в котором не может самостоятельно дифференцировать игровой компонент от содержания, попадает в определенную группу риска по формированию игровых зависимостей или усиления уже имеющихся.

3. Разновидностью геймификации образовательного и воспитательного процесса является виртуальное моделирование ситуаций, в которых ребенок отрабатывает определенные социальные навыки. Зачастую виртуализация данных компонент связана с некоторым высоким уровнем стресса от события и избеганием его в реальной жизни. Виртуальный помощник призван создать воображаемую и игровую среду, в которой тот или иной социальный навык ребенка может быть либо протестирован, либо сформирован.

Кейс. *В 2018 году некоторые школы США внедрили образовательный продукт, комплект программ FearNot, в котором создавались виртуальные миры с различными персонажами. Ресурс нацелен на формирование социальных навыков у детей в ходе общения с виртуальными персонажами. Программа инсталлирует различные повседневные коллизии, требующие эмоциональной и социальной вовлеченности. Целью программы является формирование у детей навыков сопереживания, взаимовыручки и коллективной поддержки в ситуациях, когда по разным причинам в реальном пространстве ребенок с подобными вызовами не сталкивается. В частности, даже если ребенок не был жертвой или свидетелем буллинга, программа имитатор инсталлирует подобную ситуацию, в которую ребенок принудительно интегрируется как актор. Одним из побочных мотиваторов создания подобных программ является также идея апробации способов помощи подросткам в определенной ситуации (например, жертвам буллинга), а также разработки лучших способов взаимодействия в межподростковой среде (например, противодействию буллингу).*

Аспекты геймификации сопровождаются широким резонансом и обсуждениями. В них этические споры занимают не последнее место. Ключевой момент, вызывающий тревогу, — психологические последствия внедрения виртуальных помощников. Под последними понимаются как персонифицированные элементы, так и смоделированные жизненные ситуации. Открытым остается вопрос об оправданности виртуального моделирования процессов, в которые ребенка принудительно погружают и которые могут причинить вред его психике. В частности, при моделировании ситуаций с буллингом система развивает социальные навыки ребенка, равно как тестирует социальный потенциал выработанных рекомендаций и норм поведения. Однако оправданность нагрузок, которые при этом испытывает ребенок, может вызывать вопросы. Превентивность формирования определенных социальных навыков в условиях стресса может вызвать сопротивление как родителей, так и самих детей. Кроме того, некоторые из разыгранных ситуаций могут вызывать неоднозначные последствия. Результатом

общения с виртуальным помощником может стать проактивное поведение, а именно решение воспроизвести в реальных условиях проигранный в виртуальном сеансе игры сценарий. Подобные намерения и их реализация могут указать на провоцирующий характер виртуальных помощников, моделирующих социально опасные и нежелательные ситуации. Как следствие, не стоит ожидать большой поддержки таких программ со стороны общества.

4. На качество образования могут повлиять и особые консультативные программы, призванные сгенерировать персональные и индивидуально релевантные тактики обучения, выбор профессии, перспективность приобретения определенной квалификации отдельно взятым лицом. Данные стратегии нацелены на оптимизацию выбора пути профессионального развития человека. Они строятся на основе тестирования и анализа его природных и приобретенных способностей. Программы ИИ направлены на выполнение данных функций и активно генерируются в современные консультативные центры и методики индивидуального обучения. Концепция персонифицированных траекторий образования активно развивается уже сейчас, однако, согласно оценке специалистов, будущее образование будет в значительной степени ориентировано на уход от массовости образования в сторону специально подобранных карт обучения и повышения квалификации, разработанных под специфические потребности отдельных людей. Немаловажную роль в создании таких подходов играют комплексы программ по тестированию индивидуальных особенностей молодых людей на наличие врожденных способностей и дарований, равно как их обучаемости необходимым навыкам. Определенные типы ИИ, подготовленные для этих целей, активно работают последние десятилетия в виде консультативных программ подбора будущих профессий, рекомендуемых школьникам, абитуриентам или студентам.

Кейс. *Большинство обзоров рынка наиболее перспективных и высокооплачиваемых индустрий будущего называют услуги индивидуального образования людей, согласно их нуждам и потребностям, а также с точки зрения оптимальной самореализации и раскрытия талантов. Данный сектор услуг будет предоставлять расчет*

траекторий карьерного успеха и предлагать индивидуальный план развития и обучения. Образование станет одной из компонент данной услуги. Одним из важнейших диагностических параметров данной услуги станет тестирование человека с точки зрения его природных дарований и наиболее выраженных способностей.

Консультативные программы тестирования профессиональных потенций сталкиваются с определенными этическими дилеммами. Определенный скепсис вызывает прогностический уклон данных программ. Программа, тестирующая навыки абитуриента, показывает в процентном выражении наличие тех или иных способностей: например, около 10% в области математики и около 40% в музыке. Данные сведения могут серьезно демотивировать тестируемых. В силу высокой степени доверия экспертному уровню программ ИИ тестируемые могут воспринимать рекомендации программы относительно некритически [Hakli, Mäkelä, 2019]. Несмотря на рекомендательный характер итогов тестирования и высокий процент погрешности результатов, а также не решенный с научной точки зрения вопрос о соотношении врожденных способностей и их развития в ходе обучения, абитуриенты могут буквально интерпретировать результаты тестов. По итогам консультирования они могут либо полностью отказаться от тех профессий, которые планировали рассмотреть для будущей карьеры, либо целиком ориентироваться на те сферы, которые порекомендует программа. Данные программы отличаются также высокой психологической травматичностью. Абитуриенты могут быть не готовы к тому, чтобы потенциал их способностей в любимом деле оказался недостаточно высоко или низко оцененным. Одним из ключевых и еще недостаточно хорошо изученных эффектов применения, консультирующих ИИ в образовании, является феномен нефальсифицируемости предсказаний системы. Информирование абитуриента о потенциале его способностей может способствовать тому, что он действительно перестанет развиваться в одном направлении и начнет двигаться в другом. Рост развития его потенциала может подтвердить прогноз программы. Однако в действительности указан-

ные последствия станут результатом целенаправленных действий самого испытуемого на основе сделанного прогноза, а не успешностью прогностической функции самой программы. Таким образом, проверить эффективность работы самой программы не представляется возможным. Активный вклад тестируемого на основе доверия к программе будет способствовать ложному подтверждению ее правоты. И в этом случае тестируемый оказывается не только введенным в заблуждение, но и лишается способа проверки сделанного программой прогноза.

Еще один этически проблемный участок внедрения консультативных программ — это риск различных, в том числе незапланированных, дискриминаций. При использовании систем ИИ для оценки того, к какому виду профессиональной деятельности подходят абитуриенты, система может рекомендовать специальность по полу, а также может разделять студентов по регионам и расам. Поэтому следует уделять больше внимания этическому обучению искусственного интеллекта.

5. Еще одна технология ИИ, так называемый комплекс интерактивных образовательных программ, отражает аспекты персонификации обучения. Согласно данной технологии, уровень требований при подготовке обучаемого должен соответствовать уровню его возможностей. Анализируя разные комбинации методик и успехи учеников по различным показателям, система гибко подстраивается под уровень подготовки ученика и предлагает посильные ему задания и материал для освоения. Оценивая мотивацию и эмоциональное состояние ученика, программа составляется так, что тестовые задания подстраиваются под уровень обучаемого. В рамках обучения по данным программам успешность освоения материала и прохождения курса, как правило, характеризуется высокими показателями.

Кейс. *Обучающая система Smart Sparrow была внедрена в некоторых школах США и уже зарекомендовала себя как пользующаяся большим успехом как у учителей, так и учеников. В рамках данной программы обучения за каждым учеником закрепляется своя история успеха, а также личный профиль прогресса при освоении курса. Задания строго*

индивидуализированы и гибко подстраиваются под средний коэффициент успеваемости отдельно взятого ученика в определенный отрезок времени. Ни у кого, кроме самого ученика и преподавателя, ведущего курс, нет доступа к профилю каждого ученика.

Методики, подстраивающиеся под уровень обучаемых, пользуются большим спросом в силу демократичности процесса обучения, гуманистического замысла проекта, высокой доли учета индивидуальных особенностей каждой отдельной личности в глазах самих обучаемых и высоким мотивационным потенциалом. Вместе с тем данные программы вводят самих обучающихся в некоторое заблуждение в отношении их истинных успехов в процессе обучения [Bankins, Formisa, 2019]. Система персональной настройки оценки качества освоения материала могла бы полностью оправдать себя только в условиях, когда общество в целом отказывается от стандартизированных процедур оценки на всех уровнях. Хотя шаблонные критерии оценки качества труда или успеха обучения пока не применяются повсеместно, использование их в отдельно взятом сегменте может способствовать разрыву в оценках и искаженному представлению обучающегося о собственном прогрессе. Успешность освоения той или иной программы конкретным абитуриентом может не найти подтверждения в его последующем карьерном росте или продолжении обучения. Данное обстоятельство дает основания для этических споров о том, вводят ли системы индивидуальной оценки успешности обучения в заблуждение тех, кто по ним обучается [Danaher, 2019]. В какой степени можно говорить о соответствии оценок, которые обучаемый получает внутри системы «подстраивающегося» обучения, и широком признании результатов его персонифицированного успеха?

2.2. Перечень рекомендаций

Внедрение многих образовательных технологий на основе ИИ требует особой щепетильности на всех стадиях реализации. Во-первых, прогностические технологии должны иметь статус реко-

мендаций, обязательна предварительная информированность всех участников о специфике работы программ, а также исключительно добровольное их участие. Важна также экспертная оценка эвристического потенциала технологий, что может быть поручено соответствующим специалистам как по дидактическому профилю, так и по профилю цифровой грамотности. При создании и внедрении ИИ-технологий в сфере образования обязательно консультирование психологов, в том числе специалистов по возрастной психологии. В комплексе с другими измерительными процедурами, необходимо лонгитюдное исследование эффективности инструментов, а также отсутствие ущерба для обучающихся пользователей.

Контроль за разработкой и внедрением ИИ-технологий в образовании может быть возложен частично на руководство самих университетов, различные научно-исследовательские группы, общественный и государственный контроль. Важно избегать централизации, в том числе государственной. Цели и критерии надзора должны соответствовать разным жизненным контекстам и окружению, способствовать развитию навыков коллективной работы, в том числе быть готовыми к межкультурному диалогу при работе над проектами.

В общем и целом распространение технологий ИИ в области начального, среднего и высшего образования должно сопровождаться особым социально-гуманитарным экспертным консультированием. При отсутствии общественной поддержки и пристального коммерчески незаинтересованного социального мониторинга, существует опасность понимания эффективности как сугубо количественной характеристики. Сохранение ценностной и гуманистической основы возможно только в случае учета культурных и общечеловеческих ценностей, не всегда подлежащих наукометрическим стандартам. Именно поэтому контроль за применением ИИ в области образования должен быть возложен на общественные институты, частично поручен государственным контролирующим органам, но непременно при условии открытого общественного обсуждения. Оправданным представляется внедрение специального курса, обучающего этике ИИ, который преподавал-

ся бы отдельным специалистам, в том числе специалистам в области госуправления.

Безусловно, ИИ-технологии в сфере образования становятся источником совершенствования возможностей обучения посредством использования доступа к обширным цифровым данным. Однако следует следить за тем, чтобы ориентация на децентрализацию и мобильный доступ к огромным информационным потокам не сопровождалась принесением в жертву аналитических и коммуникативных навыков, формируемых в личном общении, корпоративного «духа» университетов. Следует сохранить в образовании функцию социальных лифтов и формирование первичного коллегиального окружения, возникающего, как правило, в стенах учебных заведений. Кроме того, важно не допустить того, чтобы использование ИИ лишь подстраивало образование под такие стартовые условия, как природные дарования, уровень интеллекта, социальная среда или сила мотивации, поскольку в этом случае образование может утратить функцию воспитания, а именно формирования в культуре того, чего нет, привнесения нового, собственно, развития. Следует стремиться также к тому, чтобы ставка на просчитанные параметры не порождала явные или скрытые ограничения для социальной мобильности.

Можно резюмировать, что бум применения ИИ в сфере образования — это одновременно и замечательная возможность обновления и перехода на совершенно новый уровень, и источник серьезных социальных рисков и угроз. При этом потребуется ценностное обоснование и определение границ внедрения различных версий оценки человеческого потенциала, в котором будут участвовать не только органы власти, но также вузы, школы и отраслевые эксперты. Кроме того, необходимо наличие обратной связи от преподавателей и обучающихся, а также понимание того, как на их эмоциональное состояние и мотивацию влияет цифровая среда. Отдельного экспертного решения с привлечением высококвалифицированных специалистов требует проблема конфиденциальности информации и установление разных режимов доступа к ним, ситуаций и целей, когда эти данные могут быть использова-

ны. Также внимания экспертов к себе требует проблема ошибок ИИ в сфере образования и ответственности за них. Поскольку в образовании сохранение идеалов гуманизма и ценности человеческого достоинства является главным критерием эффективности, то ИИ-программы должны быть снабжены различными антидискриминационными защитами. В частности, предстоит тщательно продумать, какие могут быть «зашиты» в ИИ для того, чтобы избежать такие недопустимые формы поведения, как сексизм, расизм, эйджизм, буллинг и проч. Разработку данных мер в России можно было бы возложить на независимые ассоциации школ, а последующее внедрение и контроль на отдельный экспертный орган при Министерстве науки и высшего образования или Министерстве просвещения по вопросам этического регулировании внедрения ИИ. Важным требованием здесь было бы наличие обязательной двусторонней связи между независимыми ассоциациями и органами государственной власти. Следует не допускать единоличных решений органами власти, обеспечить заинтересованное и открытое к изменениям обсуждение с низовыми уровнями управления, а также представителями граждан и самими гражданами.

Наконец, важная задача при цифровизации образования, которую предстоит решить государству в ближайшем будущем, связана с проблемами справедливости распределения общественных благ. Наряду с этим еще одной наиболее прикладной рекомендацией будет организация курсов по этике для специалистов-разработчиков ИИ в области образования. Общественная инициатива могла бы состоять в том, чтобы ввести и расширить модуль преподавания по этике ИИ и не ограничиваться базовыми знаниями. Чтобы обучение студентов компьютерных специальностей было современным и полноценным, им следует изучать различные курсы по этике, например, такие как «Философские проблемы этики», «Философские проблемы этики ИИ», «Культура, этика и цифровые технологии» и проч. Контроль за этим процессом можно поручить общественным институтам и государственным органам.

3. Здравоохранение

Применение искусственного интеллекта в области здравоохранения во много раз повышает эффективность развития медицины, позволяет лучше и быстрее внедрять результаты научных открытий в практическую сферу помощи пациентам. Благодаря возможности работы с большими объемами данных, программы ИИ в состоянии предложить многократно усовершенствованные методы диагностики. Последние 10 лет программы ИИ уже активно используются в расшифровке данных объективных обследований, таких как УЗИ, рентгенография КТ- и МРТ-снимки (загруженное в систему изображение проходит полный цикл заключения — от расшифровки данных до постановки диагноза и выдачи рекомендаций). Программы ИИ активно применяют на предприятиях при мониторинге профессиональных заболеваний, при оценке развития хронических состояний, автоматизированных решений о госпитализации людей в случае вызова скорой помощи, а также в высокотехнологичных операциях с привлечением роботов-ассистентов в хирургии. Еще одним бурно развивающимся сегментом развития программ ИИ является область фармацевтики как в части подбора лекарственных средств по самым разным параметрам, так и помощи в создании аналоговых препаратов, а также при разработке оригинальных рецептур. В связи с применением ИИ в области фармацевтики и медицинского администрирования, эксперты прогнозируют значительное сокращение доли рабочих мест для людей-фармацевтов, специалистов по УЗИ, КТ- и МРТ-диагностике, рентгенологическим, а также лабораторным исследованиям. К середине текущего десятилетия, как утверждают эксперты BIS Research, доля ИИ-технологий в сфере здравоохранения приблизится к 30 млрд долларов [Council of Europe, 1990].

Наряду с этими успехами, применение искусственного интеллекта в медицине поднимает самые серьезные вопросы этического характера. Определенная их часть перекликается также с юридическими проблемами, в частности, вопросами, касающимися конфи-

денциальности персональных данных граждан. Этические проблемы применения ИИ в здравоохранении касаются главным образом дискриминации, душевного дискомфорта пациентов, риска нанесения им психологического ущерба и неоднозначности отношений между врачом и пациентом. В области здравоохранения технологии ИИ используют очень большие объемы данных для выявления закономерностей, которые затем применяются для прогнозирования вероятности будущих событий. В медицине такие наборы данных будут интегрироваться из электронных медицинских карт, заявлений о медицинском страховании, данных профилактических осмотров по месту работы, различных источников по самодиагностике пациентов, предоставляемых различными компьютеризированными источниками, а также из других источников. Формируя профиль здоровья, ИИ может опираться на информацию о частоте посещений поликлиники, пребывания в стационаре, данных амбулаторных обследований, частоте обращения к врачам-специалистам, а также частоте и длительности отпусков, доходах, судимостях и, в ряде случаев, информацию из социальных сетей.

Проекты с элементами технологий ИИ уже начались в России. Так, в Якутии 23 июля 2019 года врачи продемонстрировали коллегам из Минздрава России ИИ, разработанный группой молодых ученых Северо-Восточного федерального университета имени М.К. Аммосова. Молодые исследователи научили ИИ на основе анализа данных рентгеновских снимков распознавать патологию и даже оценивать степень вероятности наличия той или иной патологии, что является важной поддержкой в принятии врачебных решений. Показательно, что аналогичные разработки в области цифровой медицины имеются и в России, и за рубежом, однако все они — коммерческие проекты. Таким образом, в Якутии появился первый ИИ, разработанный государственным медицинским учреждением. Однако, несмотря на доминирование в РФ государственной медицины, этот случай является, скорее, исключением, чем правилом. Пилотные проекты по применению ИИ в совершенствовании радиологической диагностики реализованы российскими коммерческими компаниями в Новгородской и

Тульской областях. Однако аналитики считают эти проекты разовыми, а инвестиции «несистемными». Регионы платят за пилотные проекты 3–7 млн рублей, что не покрывает и 10% инвестиций в создание продуктов на базе ИИ.

Производители для продвижения продуктов на базе ИИ формируют ассоциации. Так, в 2018 году при Ассоциации Заслуженных врачей России была основана некоммерческая ассоциация «Национальная база медицинских знаний» (здесь и далее — НБМЗ), задачей которой является продвижение в медицинском сообществе РФ решений на базе ИИ. В настоящее время, несмотря на реализацию пилотных проектов, продукты на базе ИИ не пользуются высоким спросом в российском здравоохранении. Основными препятствиями являются отсутствие доверия у медицинского сообщества к результатам диагностики, проводимой с помощью ИИ, барьеры законодательства, препятствующие внедрению решений на базе ИИ, поддерживающих принятие решений медиками. Тем не менее ИИ представляет собой либо полноценный помощник медика, либо вспомогательный инструмент в различных направлениях медицины.

На сегодняшний день различные группы исследований уже демонстрируют, как можно использовать ИИ для прогнозирования возникновения и развития различных заболеваний. К заболеваниям, которые уже так или иначе мониторятся в наши дни с помощью различных технологий ИИ, относятся кардиологические и ишемические заболевания и болезни сосудов, инсульт, диабет, снижение когнитивных функций, в том числе возрастной этиологии: болезни Альцгеймера и Паркинсона, а также онкологические заболевания. Эта прогностическая способность искусственного интеллекта поднимает серьезные вопросы этического характера в сфере здравоохранения. Если ИИ генерирует прогнозы о здоровье, однажды эта информация может быть не просто включена в электронные медицинские карты, но и использована на всех уровнях и этапах жизнедеятельности индивида, может повлиять на качество его жизни, а также состояние здоровья и уровень психологического комфорта. В рамках программы внедрения ИИ в сферу здравоохранения предполагается, что автоматизированный умный мониторинг здо-

ровья граждан станет средством сопровождения состояния их здоровья в течение всей жизни. Внедрение подобных программ ориентировано на продление длительности жизни граждан и улучшение ее качества. Самые сильные стороны ИИ в области здравоохранения — это автоматизированные сбор и анализ данных, мониторинг и диагностика пациентов, оценка рисков возникновения заболеваний, назначение лечения и терапевтических мероприятий, а также проведение высокотехнологичных операций. Эти и другие технологии поддержания здоровья, включающие новейшие разработки ИИ, могут привести к целому ряду этических, юридических, социальных и даже философских проблем. Уже сейчас многие эксперты обеспокоены технологической зависимостью врачей-людей от машин [Council of Europe, 1997]. Например, проводимые по программам ИИ высокотехнологичные операции людей могут приводить к постепенной утрате навыков и квалификации людей-медиков, а также потере знаний, поскольку программы ИИ могут продолжительное время существовать на самообучении и автоматизированном поддержании своих рабочих навыков. Специалисты отмечают, что данная проблема серьезней традиционной проблемы о вытеснении людей роботами. Потеря высокотехнологичных знаний и навыков вдвойне опасна, поскольку человечество может разучиться создавать и контролировать ИИ, фактически уже инициировав его производство.

3.1. ИИ в области здравоохранения

1. Ключевой и относительно широко внедряемой технологией ИИ в сферу здравоохранения является автоматизированный сбор данных о состоянии здоровья и расчет вероятности возникновения заболеваний у пациентов. Предполагается ввести прогнозы, формируемые ИИ, в электронные медицинские карты. Согласно идее создателей данной технологии, доступ к автоматизированному расчету вероятностей заболевания должен облегчить работу лечащего врача. Также при реализации подобной технологии должна возрасти эффективность решений о назначении того или иного лечения.

Считается, что с расчетом прогностических рисков заболеваний автоматизированные системы справляются намного лучше. Такой успех объясняется тем, что разрабатываемый для сферы здравоохранения ИИ формирует эмпирическую базу случаев несоизмеримо большую, чем любой врач за всю свою практику. В данных разработках также сведена к минимуму погрешность человеческой психологии. Врачи-люди, как правило, ориентируются на наиболее типовые случаи лечения (поскольку они легки и понятны в применении), они также лучше помнят случаи из недавней практики и те экстраординарные случаи, которые увенчались успешным лечением. Напротив, необычные и неудачные случаи лечения люди-врачи реже принимают во внимание. В отличие от избирательной памяти врачей-людей, программы ИИ «помнят» все случаи и не отдают психологического предпочтения одним или другим.

Кейс. *Система AI Clinician, разработанная британскими учеными, на основании анализа ста тысяч историй болезней сепсисом, определила стратегию лечения пациентов с сепсисом лучше, чем врачи-люди [Strickland, 2018]. Программа работала примерно с пятьюдесятью переменными, включая детали анамнеза иных заболеваний, которыми болели пациенты в течение всей жизни, возраст, общие показатели здоровья, вредные привычки и рацион питания. Выводы по результатам использования системы были опубликованы в журнале «Nature Medicine» и вызвали большой прилив энтузиазма у врачей. Результаты исследования отчетливо показали, что искусственный интеллект предложил более достоверный прогноз, чем врачи-люди. Авторы-создатели данной программы утверждают, что подобные технологии должны стать рутиной при поиске наилучших стратегий лечения пациентов. Практикующим врачам следует включать данные автоматизированного прогноза в личные медицинские карты, чтобы оценка рисков заболеваний стала неотъемлемой частью диагностики и терапии.*

В России в 2016 году советник Президента РФ Герман Клименко анонсировал, что в области медицины началась разработка перво-

го проекта фонда Института развития интернета (URL: https://thirdopinion.ai/ru). Система ИИ «Третье мнение» должна стать аналогом наиболее крупной ИИ разработки в области здравоохранения на Западе IBM Watson. Она будет распознавать отклонения и патологии организма, анализируя данные рентгена, УЗИ, КТ, МРТ.

***Кейс.** В 2016 году в связи с внедрением программ ИИ, разработанных корпорацией DeepMind Health, разразился громкий скандал. Программы, разработанные данной корпорацией, были внедрены в Лондонской Королевской больнице, однако вскоре сотрудничество было прекращено. Как оказалось, практически сразу после начала работы автоматизированных систем, один отдел корпорации Google получил доступ ко всей цифровой информации больницы за последние несколько лет: записям болезней, вызовам скорой медицинской помощи, данных радиологиии т.д. Информация об этой утечке попала в прессу, и на корпорацию DeepMind Health был подан судебный иск. Согласно законодательству, данные пациентов подпадают под конфиденциальную информацию и всегда передаются в зашифрованном виде без указания имен. Однако DeepMind Health получил доступ к именам, которые при передаче данных оказались не зашифрованы. На сегодняшний день разбирательство до конца не прекращено, однако возникновение подобных инцидентов может стать кейсом, который даст толчок для разработки специальных нормативных актов, направленных на защиту персональных медицинских данных от систем ИИ.*

Данная технология вызывает ряд опасений этического характера. Скептики высказываются за то, чтобы ограничить доступ к медицинской карте. Если данные оценки ИИ-рисков будут доступны потенциальным работодателям, страховым компаниям, кредитным и заемным организациям и даже будущим супругам, то они могут быть использованы не в интересах человека. Например, если прогнозы о снижении когнитивных функций сотрудника станут доступными работодателю, то человек может лишиться выгодного

контракта. Также информация о возникшей частичной толерантности к опиоидам может всерьез повредить репутации человека. Любые сведения о перенесенных ранее инфекционных заболеваниях, выкидышах, абортах или то, что человек является носителем генетических мутаций, могут послужить источником нежелательных последствий. В процессе лечения личные карточки пациентов попадают в руки не только врачей, но также администраторов и многих клиницистов. В ряде случаев доступ к медицинским данным требуется по условиям сотрудничества при найме на работу или оформлении страховых контрактов [McDougall, 2019]. Утечка данных возможна в самых разных случаях, несмотря на меры предосторожности и политику конфиденциальности, которой придерживаются медицинские учреждения. Наконец, сами пациенты могут передавать сведения о себе третьим лицам или доверять им доступ к результатам анализов или обследований. Вся эта информация может попадать в открытый доступ или передаваться по месту работы, в кредитные организации, страховые компании, судебные органы, иные инстанции. Если при этом данные о здоровье будут содержать высокоточные сведения о долгосрочных прогнозах по здоровью граждан, то этическая сторона дела будет выглядеть весьма тревожной. Представим, что деловые партнеры оказываются информированными о 70%-й вероятности смерти от инфаркта в ближайшие 5 лет у одного из потенциальных агентов сделки. Данные сведения могут быть использованы как в пользу отклонения сделки, что может расцениваться как дискриминация, или, напротив, поспешного заключения на условиях, предполагающих их невыполнение по причине смерти агента, что будет откровенным злоупотреблением. Так или иначе, высокоточные прогнозы ИИ в отношении весьма деликатных ситуаций приводят к серьезным этическим дилеммам. Также информация о 60%-й вероятности развития заболеваний, вызывающих временную или пожизненную нетрудоспособность в ближайшие 10 лет, могут побуждать работодателей к немотивированным увольнениям сотрудников, входящих в группу риска. Долгосрочные и точные прогнозы могут всерьез влиять на решения людей в отношении друг друга.

Информированность о вероятностях развития различных патологий беременности могут сказаться на решении на решения людей о замужестве и рождении детей. Очевидно, что любое раскрытие врачебной тайны или данных о состоянии здоровья граждан может приводить к дискриминации и злоупотреблениям. Работодатели, заинтересованные в эффективности своих сотрудников, будут нанимать кандидатов с минимальным риском развития болезней. Напротив, потенциальные кандидаты с определенными хроническими или врожденными заболеваниями, вероятнее всего, будут отклоняться. Также кредитные агенты или страховые компании могут принимать неблагоприятные решения в отношении лиц с ослабленным здоровьем. Такое положение дел не зависит напрямую от технологий ИИ, однако многократно усиливает эти риски, а также создает новые прецеденты и сферы распространения. Высокая тонкость прогнозов ИИ в отношении здоровья, более всесторонний мониторинг и анализ данных, а также долгосрочность прогноза при условиях сохранения точности неизбежно меняют отношение людей к подобным прогнозам. В случае их высокой точности и долгосрочности доверие к ним возрастает, а значит, растет и соблазн делать более прямые расчеты экономической целесообразности на основе настолько точных сведений о здоровье и продолжительности жизни людей. Ситуация отягощается также тем, что прогнозы ИИ не просто носят вероятностный характер, но и могут содержать в корне ошибочный прогноз. Некоторые пациенты могут вовсе не быть подвержены риску прогнозируемых ИИ заболеваний. Это означает, что пациенты могут страдать от дискриминации и возможного психологического ущерба в результате ошибки или ложных сведений.

2. На сегодняшний день доказано, что с формированием прогнозов программы ИИ справляются лучше, чем люди-врачи. При тестировании успешности людей и машин люди были правы в 60% случаев, а машины — в 72%. Программы ИИ демонстрируют успешность как в прогнозах по протеканию болезней, так и по рискам возникновения тех или иных заболеваний, появлению различных осложнений и т.д. Результаты недавнего исследования по риску возникновения родового кровотечения у рожениц, прини-

мающих кроворазжижающие препараты при высоком D-димере, показали, что чаще всего технология ИИ соответствовала или была лучше, чем решение акушеров-людей. В исследовании также было обнаружено, что решения ИИ об отмене препаратов были наиболее целесообразны и обоснованы лучше, чем решения людей. Успешность ИИ в подобных случаях связана с доступом к большому массиву данных, не ограниченных страной, временем или определенным учреждением. Известно, что стратегии лечения зачастую не совпадают в различных странах, где сертифицированы разные подходы и официальные медицинские стандарты. В различных разделах медицины наблюдаются несовпадения большей или местной интенсивности. К примеру, в области стоматологии рынок медицинских услуг и технологии лечения стандартизированы в высокой степени. Еще одна область, где прослеживается значительное совпадение в стратегиях лечения, — офтальмология. К сферам с невысокой степенью совпадений относится акушерство и гинекология. Здесь разброс по странам в способах диагностики и методах лечения существенно разнится. Как следствие, люди-врачи в работе придерживаются тех технологий, которые распространены в их странах. Напротив, библиотека практических случаев и методик лечения в программах ИИ не ограничивается определенной страной. Как следствие системы ИИ имеют доступ к гораздо более обширному и разнообразному арсеналу случаев и средств лечения, чем и объясняется их успех.

Кейс. *В 2016 году «умный» сервис диагностики IBM Watson выявил у 60-летней пациентки редкую форму лейкемии [Monegain, 2016]. Для этого система за 10 минут «изучила» 20 миллионов научных статей о раке. Лечащими врачами пациентке был поставлен другой, неверный диагноз. Программа лучше справилась с задачей диагностики и фактически исправила ошибку врачей-людей.*

Несмотря на очевидные преимущества прогностического потенциала ИИ, сложную этическую проблему представляет сам статус решений на основании вероятностей. В основе этой проблемы лежит философский вопрос о том, где именно существуют вероят-

ности — в реальности или в системах наблюдений (сознании наблюдателей). Согласно одной из точек зрения, сама реальность жестко детерминирована и событийно определена, поэтому никаких вероятностей в ней не существует. Вероятность может появиться только в ситуации неопределенности. Если это действительно так, то решения, принимаемые на основе вероятностей, обладают гораздо меньшей предсказательной силой, чем принято думать. Очевидно, что когда на пренатальном скрининге высчитывается риск наличия генетических поломок у плода (например, мутация, вызывающая синдром Дауна), то в реальности наличие или отсутствие данного синдрома уже определено, оно есть. Соответственно, высчитывается вероятность не самого события, а эффективного описания события, скорее вероятность самой ошибки. Описание никогда не равняется 100% и в этом смысле всегда не может быть точно определено. Расчет вероятностей по модели 100% является результатом логического и психологического заблуждения. Это положение подтверждается тем фактом, что прогнозы ИИ далеко не безошибочны. Кроме того, многие внешние факторы могут способствовать ошибкам в прогнозах. Например, если данные, использованные для написания самой программы, ошибочны, то результаты анализа также будут ошибочны. Точно так же, если в программу загружаются неверные параметры или нестабильные показатели, то результаты прогноза могут быть нерелевантными [Vallverdú, 2010]. Между тем прогнозы ИИ зачастую приравниваются к решениям, которые обладают абсолютным приоритетом. Не только пациенты, но и врачи склонны наделять решения машины высоким доверием и зачастую некритически относятся к самому процессу принятия решения, а также тем вводным, которые находились на входе. Некоторые медики отмечали тот факт, что они, скорее, стараются предсказать решение системы, чем сформировать собственный прогноз [Lamanna, 2018]. Этически проблематичным является непрозрачность формирования доверия к врачебным прогнозам ИИ. Если пациенты склонны демонстрировать высокую степень доверия врачам, то, скорее всего, они окажутся не готовыми к тому, что сами врачи полностью полагаются на прогнозы ИИ.

3. Еще одна из технологий, основанных на предсказательной функции ИИ в здравоохранении, — это прогнозы в отношении здоровья, к которым у пациента есть прямой доступ. Если в неоднозначных и сложных случаях пациенты обращаются к врачу-человеку как посреднику между машиной и человеком, то ряд программ предполагает аутодиагностику или аутомониторинг. Часто пациентам с хроническими заболеваниями нужно ежедневно контролировать динамику показателей своего здоровья. Они могут использовать технологии для измерения пульса, давления или дыхания. Полученные в автоматическом режиме данные необходимо собрать, обработать и правильно интерпретировать. Для этого используются различные мобильные приложения, настроенные на портативный доступ и легкость в эксплуатации. Такие программы мобильно аккумулируют данные о параметрах здоровья и автоматически передают их лечащему врачу. Они также могут оперативно предложить наиболее простые рекомендации по облегчению состояния пациента. Данные программы также оснащены модулем долгосрочного прогнозирования и расчетом ключевых рисков на основании реальных данных пациента и ежедневного мониторинга за динамикой его здоровья. Такая программа может просчитать риск инфаркта в ближайший год исходя из данных работы сердца за последний месяц. Другой пример — риск онкологического заболевания легких на основании количества выкуренных сигарет за истекший месяц, а также объективных данных по состоянию легких. Наконец, программа может выдать динамику снижения когнитивных способностей на основе годовых МРТ-обследований и данных по энцефалопатии. Данные, полученные в таком режиме, передаются лечащим врачам и во много раз ускоряют возможность получить лечение. При посещении врача нет необходимости заново проходить обследование. Кроме того, долгосрочное наблюдение за состоянием здоровья в динамике повышает точность и эффективность назначенного лечения.

Кейс. *Мобильное приложение, созданное компанией AliveCor, обрабатывает данные с устройства для снятия кардиограмм, ко-*

торое имеет дома пациент (URL: https://www.alivecor.com/press/press_release/alivecor-unveils-medical-grade-mobile-app/). ИИ каждый день расшифровывает их и может предсказать ухудшение ситуации. Все проанализированные данные также отправляются врачу. Приложение может обнаружить приближение инфаркта и попросить пользователя обратиться к врачу. Медсестра в приложении Sense.ly может говорить с пациентом о самочувствии, давлении, жалобах на здоровье, она может комментировать симптоматику пациента, оповещать о необходимости принять лекарства или связывать с врачом по видеосвязи.

***Кейс.** Самым известным примером внедрения ИИ-мониторинга стал сервис Healthtap, который использовал массив скриптов более тысячи консультаций. По ним был натренирован чат-бот Doctor A.I. Он умеет отвечать на вопросы, сформулированные на простом языке пациентов, а не на специальном медицинском языке. Подключив умную колонку с Amazon, можно получить консультацию от этого бота [Wiggers, 2018].*

Несмотря на удобство в эксплуатации программ по автоматизированному мониторингу здоровья, они могут причинить пользователям определенный психологический ущерб. Ознакомление с данными прогнозирования может оказаться травматическим для пациентов. Например, многие люди могут сильно расстроиться, если узнают, что в будущем они, вероятно, будут страдать от снижения когнитивных способностей. Оценка рисков возникновения инфаркта в ближайшие несколько лет может всерьез травмировать пациента, страдающего от сердечно-сосудистой дистонии. Оценка рисков появления диабета при определенных показателях сахара в крови или онкологического заболевания в ответ на расшифровку данных анализов на онкомаркеры могут возыметь неоднозначное воздействие на психику пациентов. Не менее пугающими могут оказаться для неподготовленных пациентов данные об отрицательной динамике по протекающим хроническим заболеваниям. Программы ИИ мониторят текущее состояние здоро-

вья и постоянно сравнивают их с предыдущими. Не вполне однозначные для интерпретации результаты, которые могут при этом формироваться, могут всерьез пугать пользователей программ. Специалисты отмечают также возникновение у пациентов ипохондрических состояний, появляющихся в ответ на данные мониторинговых программ, различные неврозы, тяжелые состояния психологической мнительности, равно как состояния зависимости от постоянной оценки здоровья с помощью соответствующих программ. Некоторые пациенты могут жаловаться на ощущение, что их показатели здоровья стали действительно снижаться в ответ на информацию, сформированную мониторинговой программой. Как правило, врачи-люди ограничивают долгосрочность прогнозов ближайшими 5–10 годами, однако программы ИИ в состоянии формировать высокоточные прогнозы на гораздо большие сроки. Возможный психотравмирующий эффект от ознакомления с ними представляет собой определенную этическую проблему.

Пользователи мониторинговых программ подписывают соглашения о том, что все данные, которые собирает программа, а также все результаты ее промежуточных или итоговых прогнозов должны становиться доступными самому пациенту. Информация о состоянии здоровья, включая риски, не может быть конфиденциальной для пациента и открытой только для лечащих врачей. Однако не всегда возможно обеспечить правильную оценку пациентом тех данных, с которыми он знакомится. Неверные интерпретации или ошибочные оценки, к которым может склоняться пациент, не будучи врачом-профессионалом, могут всерьез исказить его картину собственного здоровья. Однако в тех случаях, когда пациент вполне адекватно понимает данные прогнозов и возможных рисков, его психологическое состояние также может пострадать в связи с реальными угрозами, существующими для его здоровья. Если в руки пациента попадает отчет, в котором приводятся данные о 70%-й вероятности повторного инсульта в ближайшие 2–3 года или 60%-й риск микроинфаркта в ближайшие годы, то качество его жизни может значительно снизиться. В ожидании реализации данных прогнозов пациенты могут жаловаться на депрессии, возникновения фобий,

психосоматических расстройств и неврозоподобных состояний. Некоторые пациенты отмечают ухудшение субъективного самоощущения после того, как начали пользоваться мониторинговыми программами. В ряде случаев могут диагностироваться аутосуггестивные состояния или появление фантомных болей. Этически непроясненной в данном случае остается степень пользы высоко вероятностных прогнозов, если пациенты, информированные об исходах своих заболеваний, а также сроках их наступления, получают серьезный психологический ущерб, который может превышать опасность самих прогнозируемых заболеваний. Отдельной этической проблемой является вопрос о том, верно ли данную ситуацию оценивать в утилитаристских терминах превышения вреда или пользы либо пациент вправе, не мотивируя это, отказаться от подобных мониторингов, наложив вето на автоматизированную оценку своего здоровья и расчеты рисков [Müller, 2019].

4. На сегодняшний день в начальном, но подающем большие надежды, положении находится сфера развития ИИ, направленного на генетическое тестирование. Подобные программы осуществляют генетическое консультирование по результатам проведенной генетической экспертизы. Наибольшей популярностью пока пользуются программы, которые применяются в пренатальном исследовании плода и прогнозировании возможных рисков генетических мутаций. В отличие от имеющихся программ оценки, используемых людьми-генетиками, созданные программы ИИ лучше и точнее рекомендуют, на какие именно генетические мутации стоит проверить пациентку. Если раньше, помимо основного списка возможных генетических аномалий развития плода, пациентка желала провести тест по оставшимся видам потенциальных мутаций (свыше 500 тысяч), то ей следовало точно определить, какую именно аномалию необходимо тестировать. Как правило, врач-генетик объяснял пациентке, что ей следует точно знать, какие именно были у родственников будущего ребенка аномалии развития (врожденная тугоухость, заячья губа, врожденная глаукома и проч.). Развернутый тест на все возможные патологии развития ранее нельзя было провести в силу колоссальной технической трудоем-

кости [Mishra, 2017]. В настоящее время программы ИИ на основе некоторых введенных данных формируют собственные рекомендации по генетическому тестированию, а в обозримом будущем, вероятно, пациентам будет предоставляться возможность пройти тотально расширенный тест (вместо специализированно расширенного) на более чем 500 мутаций, с которым будет справляться ИИ-программа по генетическому тестированию.

Кейс. *Используемое последние 8 лет исследование Panorama позволяет уже на первом триместре беременности обнаруживать различные хромосомные патологии с точностью выше 99% (Current Trends in Early Prenatal Diagnosis, 2000). Информация полученная путем анализа ДНК плода обрабатывается и идентифицируется с помощью технологий, основанных на разделении ДНК матери и плода из образца плазмы. Технология разработана для неинвазивного пренатального исследования, ее использование было одобренно в США. Программа позволяет провести стандартное обследование с участием менее 10 возможных мутаций. Для тестирования на наличие других мутаций (из списка свыше 500 позиций) пациент или его личный генетик должны точечно указать, какая именно мутация вызывает наибольшее опасение. Если к исследованию Panorama будет подключена система ИИ, то плод можно будет проверять на весь спектр потенциальных мутаций без использования предварительных информативных вводных данных. Такая технология позволит во много раз повысить эффективность самого исследования и снизить риски того, что та или иная аномалия останется незамеченной.*

Этические проблемы генетических обследований представляют собой отдельную область современной медицины и здравоохранения [Khoshnood, Babak, De Vigan, 2005]. Вдвойне проблематичными выглядят вопросы оправданности генетических тестирований в сфере пренатальной диагностики. После масштабного внедрения данной технологии (фактически с начала 1970-х годов в большинстве западных стран и в конце 1980-х в России) в обществе не сти-

хают бурные дискуссии. Этическое напряжение вокруг данной темы связано с понятными причинами — возможным прерыванием беременности в случае, если матери становится известно о поражениях плода. Активное обсуждение привело к созданию целого ряда нормативных актов, призванных создать правовую базу для медико-генетической помощи. Однако этическая экспертиза данного вида помощи развита в существенно меньшей степени. Главный этический принцип генетического тестирования — это добровольность проведения скрининга и обязательное консультирование специалистом до и после проведения процедуры. Общение со специалистом позволяет пациентке сделать выбор, наиболее релевантный ее этическим ценностям. В частности, если по тем или иным религиозным соображениям, запрещающим аборты, пациентка отказывается от скрининга, никто не оспаривает ее права. При проведении генетического консультирования важно соблюсти общий этический принцип автономии личности, а также принцип медицинской этики о непричинении вреда. В частности, обоих супругов информируют о том, что цель данного теста — выявить возможные патологии плода. В свою очередь, данная информированность и проведенная уточненная диагностика может послужить основанием для решения о прерывании беременности. Соответственно, будущие родители (окончательное решение принимает всегда мать) вправе не проводить тестовый скрининг и отказаться от диагностики в случае, если они не хотят брать на себя обязательство принимать такое решение. При внедрении программ ИИ регулирование этической стороны медицинской помощи данного типа, судя по всему, усложнится. Обширный спектр возможных параметров тестирования на генетические аномалии усложнит процедуру взвешивания всех за и против принимаемого решения. Нужно будет определять, для каждого ли отдельного случая матери нужно принимать решение в соответствии со своими ценностями или речь всегда будет идти о любом скрининге в целом. Последний вариант представляет собой значительную трудность, поскольку различные аномалии и их последствия для последующего развития ребенка далеко не одинаковы. Одни патологии значительно тяжелее

других и, наоборот, далеко не для всех случаев возможность прерывания беременности обсуждается в качестве релевантной меры. Кроме того, в ряде случаев будущие родители хотели бы знать о возможных патологиях, чтобы лучше подготовиться к ним после появления ребенка на свет. В ряде случаев сами медики заинтересованы в проведении подобных экспертиз, чтобы иметь возможность оказать адекватную оперативную помощь при рождении больного ребенка. Как следствие, расширение возможностей программ ИИ по генетическому тестированию должно повлечь значительное расширение, детализацию и спецификацию этически релевантных процедур ознакомления родителей с их правами и возможностями по каждому из тестируемых параметров. По-видимому, проведение самой процедуры консультирования ИИ будет затруднено в силу нежелательности формализации данного процесса. Родителям предстоит принимать важные решения, и лучше, если это будет проходить в неформальной и непринужденной обстановке, где они могут быть уверены, что с ними общается человек как носитель определенных этических взглядов и моральных ценностей. Также важно создать защиту от автоматизированных тестирований на генетические мутации, «зашитые» в различные другие медицинские процедуры, доверенные ИИ. Поскольку родители каждый раз заново и самостоятельно принимают решение о том, готовы ли они быть информированными о наличии у будущего ребенка возможных патологий и отклонений, необходимо, чтобы доступ к подобной информации всегда предварительно с ними согласовывали.

5. В настоящее время значительное внимание уделяется разработке таких программ медицинской помощи пациентам, которые могли бы отчасти разгрузить врачей, а именно предоставить всю информацию о лечении напрямую [Petervan, Rysewyk, Pontier, 2018]. Подобные программы создают методики лечений по прямому запросу пациента — на основе их жалоб и симптомов. При развитии данных технологий пациент освобождается от посещения врача и в то же время не занимается самолечением: фактически его лечением занимается умный помощник или «умный ИИ-доктор». Подобная технология имеет много плюсов. Среди очевидных — из-

бавление врачей от большой и подчас рутинной работы. В случае простых состояний, таких как головная боль, повышенное давление или высокая температура, «умный доктор» вполне может сформировать план лечения и поход к человеку-врачу не понадобится. Менее очевидный, но вполне реальный плюс — решение проблемы психологических страхов, связанных у некоторых людей или детей с посещением медицинских учреждений или врачей. При использовании программ-помощников некоторые пациенты могут полностью избавиться от так называемого «страха белых халатов».

Кейс. *Британская компания Your.MD запустила в 2015 году программу [Arkin, Borenstein, 2016], которая позволяет пациенту сформулировать любой симптом на простом языке и узнать от программы вероятный диагноз и рекомендации по лечению. Например, пациент может сказать: «Меня тошнит» или «У меня горячий лоб», программа выделит симптом, чтобы потом найти ему соответствие в огромной картотеке, созданной Your.MD и насчитывающей более 1,5 млн симптомов. Далее программа найдет наилучшее соответствие симптому и предложит план лечения. Еще одна, Medtronic, разработала приложение, способное предсказывать вероятность кризов за несколько суток до события и рекомендовать принять сильнодействующее лекарство либо увеличить терапевтическую дозу препарата до значительных доз. Еще одно медицинское приложение OME использует данные генетики. Приложение предоставляет рекомендации на основании генетических данных пользователя и оценивает риски заболеваемости той или иной патологии в зависимости от болезней родственников.*

Внедрение данной технологии при всех своих плюсах сопровождается целым комплексом этических проблем. Далеко не все врачи приветствуют широкое внедрение программ, которые предоставляют пациентам данные об их состоянии здоровья. Беспокойство вызывает не только очевидные риски неверного прогноза, но, скорее, неадекватные решения, которые могут принимать сами паци-

енты на основе полученных данных. Так, известно, что решение об удалении яичников, которое было принято известной актрисой Анджелиной Джоли, представлялось разным медикам спорным. В ходе консилиумов не удавалось выработать единого мнения. На основании генетической экспертизы риск возникновения рака яичников оценивался в 50%. Программа ИИ, к который прибегли для повышения достоверности оценки, предоставила примерно те же результаты. После этого сама актриса настояла на проведении операции, притом, что единого мнения по этому случаю по-прежнему не было. Врачи опасаются, что люди могут принять неверное и неоправданное решение в отношении своего здоровья при использовании различных диагностических приложений. В силу недостаточной профессиональной подготовки, различных психологических особенностей, таких как мнительность или ипохондрия, пациенты могут склоняться к радикальным мерам в тех случаях, когда с адекватной медицинской точки зрения они выглядят неоправданными. Возможны и другие случаи, когда пациенты в силу своих особенностей не придадут достаточного внимания тем предписаниям, которые выдаст программа. Они могут саботировать решения, предлагаемые ИИ-помощником, но проявить большую ответственность при посещении человека-врача, который может стать дополнительной психологической поддержкой или попытаться убедить пациента лично. Еще одной проблемой использования ИИ-врачей в виде различных программ и приложений является трудность установления ответственности в случае назначения неверного лечения или различных врачебных ошибок для тех случаев, когда врачом выступает компьютерная программа. Следует ли возлагать ответственность на производителя программы (программистов) или того, кто реализует данный продукт? В случае с человеком-врачом ответственность за неверное лечение или причинение вреда пациенту несет сам врач, однако в случае с использованием компьютеризированного помощника эта мера неочевидна. Наконец, проблематичными становятся отношения между пациентом и врачом. В силу названных трудностей врачи могут избегать активного внедрения ИИ-помощников, поскольку реакция

пациентов на полученные рекомендации может быть непредсказуемой. Если пациент станет настаивать на радикальных и неоправданных методах лечения на основе полученных прогнозов и врач окажет ему содействие, то в какой степени будет распределяться ответственность между всеми участниками процесса, не вполне ясно. Врачи также могут склоняться к тому, чтобы скрывать какие-то детали диагноза или плана лечения в случае, если пациент все-таки решится обратиться к врачу за разъяснениями компьютеризированных предписаний или подтверждения диагноза, или плана лечения, сформированного машиной. При непосредственном общении врач сам принимает решение, какой информацией делится с пациентом. Однако если ИИ-помощник будет запрограммирован на больший объем доступной для пациента информации, это может создать двусмысленную, неопределенную и даже потенциально конфликтную ситуацию между врачом и пациентом. Возможно, при более масштабном внедрении компьютерных помощников в область медицинской помощи придется более детально регламентировать, в какой степени пациенты могут «перепроверять» решения ИИ-врачей или просить дополнительных разъяснений у людей-врачей, и как будет распределяться ответственность в подобных случаях.

3.2. Перечень рекомендаций

Важность всесторонней экспертной поддержки при внедрении ИИ-технологий трудно переоценить. Необходимо отметить, что это одна из тех областей, для регулирования которой желательно подключение специалистов самых разных областей. Помимо участия самих медицинских работников, важно участие психологов, философов, специалистов по медицинской этике, а также таких более частных областей, как философия сознания, нейрофеноменология, нейроэкзистенциализм, энактивизм и др. Данная область имеет высокий потенциал для релевантности собственно философской и психологической экспертиз. Роль гуманитариев, а также специалистов по социальным наукам крайне важна

для проведения квалифицированных экспертиз соответствующих ИИ-программ. Некоторые эксперты считают, что использование ИИ-технологий в медицине должно быть регламентировано на государственном уровне.

Решение ряда вопросов, связанных с внедрением и распространением ИИ-программ в области здравоохранения, стоит поручить общественному контролю, а также надзору соответствующих государственных органов. В частности, в ведении этих структур должны находиться вопросы контроля полноценной и всесторонней информированности граждан о возможностях и последствиях применения той или иной медицинской ИИ-программы. Врачи, которые предоставляют пациентам ИИ-прогнозы, должны убедиться, что они хорошо осведомлены об их плюсах и минусах. Эксперты должны консультировать пациентов по поводу ИИ так же, как это делают опытные специалисты по генетическому тестированию. Для того чтобы применение ИИ-технологий в области медицины и здравоохранения могло способствовать сохранению здоровья, увеличению продолжительности жизни, улучшению ее качества, а также повышению благосостояния общества в целом, не только сами медики, но и независимые исследователи, политики и различные общественные институты должны учитывать все возможные этические риски внедрения ИИ-программ и действовать с осторожностью. Для этого им должны помогать соответствующие группы экспертов, проектные команды специалистов, работающих на базе междисциплинарных подходов, они же должны оказывать посильную помощь для предотвращения возможных злоупотреблений при применении ИИ-технологий. Наиболее реалистичным в краткосрочной перспективе было бы решение данной задачи с помощью назначения каждым медицинским учреждением отдельных специалистов по медицинской этике, а также этике внедрения ИИ в области медицины.

Еще одним решением было бы формирование этических комитетов при крупных медицинских учреждениях (в частности, научно-исследовательских институтах) по аналогии с этическими комитетами в ведущих западных университетах и исследова-

тельских центрах. Помимо медиков в них могут входить отраслевые эксперты, специалисты по этике, представители курирующих отрасль органов власти, представители гражданского общества и НКО, защищающие права пациентов. В частности, следует продумать этическую сторону защиты конфиденциальности данных. Нарушение врачебной тайны является одним из самых неприемлемых последствий автоматизации сбора и обработки данных при работе с пациентами. Искусственному интеллекту необходим доступ к гигантским массивам данных, чтобы выдавать релевантные прогнозы. Однако все собранные медицинские данные (личные карточки пациентов, лабораторные данные, снимки, протоколы УЗИ, данные анамнеза и заключения врачей) являются личными данными и подпадают под определение конфиденциальной информации, и работа с ними должна регулироваться законом о защите персональных данных. Решение о том, в каких объемах и по каким алгоритмам передача данных была бы возможна, целесообразно возложить на этические комитеты, работающие при медицинских учреждениях. Наряду с этим достаточно общим нормативным требованием возникает и целый ряд достаточно тонких этических последствий разглашения личных данных о состоянии здоровья граждан. В случае если оцифрованные данные (электронные карты) окажутся в открытом доступе, то они могут быть использованы различными организациями в собственных интересах. Например, если данные карт станут доступны страховым компаниям, то весьма вероятно изменение условий обслуживания граждан, в зависимости от их текущего коэффициента здоровья. Справедливость подобных мер — это дискуссионный вопрос. Однако наиболее несправедливым представляется коррекция условий обслуживания граждан на основе долгосрочных прогнозов по их здоровью, сформированных высокотехнологичным ИИ с развитой функций прогнозирования. В частности, если страховые компании прибегнут к завышению стоимости полисов для пациентов, у которых ИИ определяет в перспективе высокие риски онкологических заболеваний, инфарктов или инсультов, требующих дорогостоящего медикаментозного лечения и реабилитации, то подобная мера пред-

ставляется несправедливой. Также следует принимать защитные меры для предотвращения ситуаций, в которых потенциальные работодатели могут отказывать соискателям или изменять условия найма сотрудников на основе различных ИИ-прогнозов, в том числе генетического прогнозирования. Разъяснения подобных вопросов, лежащих в ведении этической экспертизы, должны проводиться соответствующими специалистами. Очевидно, что развитие ИИ-технологий в сфере здравоохранения требует важных модификаций в правовой базе, однако большая их часть должна проводиться на основе этических экспертиз. Ближайшими исполнителями данных требований могли бы стать специальные этические комитеты, работающие при медицинских учреждениях. В идеале специалисты данных комитетов должны иметь подготовку по двум квалификациям — в сфере медицины и врачебного дела, а также в области этики ИИ. Одним из вариантов решения данной, прежде всего государственной, задачи была бы переподготовка медицинских кадров по профилю этики ИИ в рамках повышения квалификации или дополнительного образования при разработке соответствующих образовательных стандартов. Из числа данных специалистов можно было бы формировать потенциальный контингент этических комитетов, функционирующих при каждом медицинском учреждении.

4. ЖКХ. Системы «Умный дом»

Наряду с самыми разными сферами внедрения ИИ наиболее активное и широкое применение компьютерных технологий происходит в области повседневной жизни. Одной из наиболее востребованных технологий современного внедрения ИИ в бытовую сферу является технология «Умный дом». Данная технология получает все большее распространение как в частных домах и домовладениях, так и в качестве централизованной услуги, предлагаемой современными управляющими компаниями и комплексами жилищно-коммунального хозяйства. Концепция «Умного дома» была сформулирована Институтом интеллектуального здания в Вашингтоне в 1970-е годы. «“Умный дом” — здание, обеспечивающее продуктивное и эффективное использование рабочего пространства благодаря оптимизации его четырех основных элементов: структуры, систем, служб и управления, а также взаимоотношений между ними» [Анненков, Рыбкин, 2013, с. 3].

Эта концепция заключает в себе ряд требований по обслуживанию и эксплуатации помещения. Как правило, система «Умного дома» охватывает оборудование, управляющее отоплением, освещением, связью и электронными устройствами. Они управляются дистанционно и обладают высокоавтоматизированной настройкой, работа которой основана, в том числе, на автоматическом сборе и анализе данных. Дистанционное управление системой «Умного дома» может осуществляться с помощью портативных устройств личного пользования (любых гаджетов: смартфонов, планшетов, компьютеров) или при помощи централизованных систем, подключенных к дому управляющими компаниями.

Проект позволяет комплексно решать проблемы экономии электроэнергии, пожарной безопасности, управления освещением, кондиционирования, охраны, контроля физического доступа, управления оборудованием вне дома. Такие дома — источники больших потоков данных для проекта «умного города». Они открывают возможности оптимизации потребления ресурсов в масштабе как городского района, так и города в целом.

В основу технологии «Умного дома» положены критерии совершенствования комфорта, безопасности, оптимизации потребления и экологичности.

Автоматизация коммунального обслуживания значительно увеличивает защищенность собственников и их имущества. Так, установка круглосуточного наблюдения за своей квартирой или домом, которая автоматически управляется с помощью приложения на смартфоне, позволит оперативно сообщить о возникновении пожара или вторжении третьих лиц в режиме спецсигнала, поступающего сразу на телефон. Также можно подключить запись с камер видеонаблюдения и их автоматическую передачу в облачное хранилище данных, чтобы в случае вторжения квартирных воров или наступления других страховых случаев, воспользоваться возможностями реконструкции событий. Кроме того, при подключении различных охранных опций технологии «Умный дом» у владельцев появляется возможность удаленного управления системами сигнализаций. Например, система сможет самостоятельно включить сирену или систему пожаротушения, если произошло задымление, или заблокировать выход и послать сигнал в службы безопасности в случае вторжения третьих лиц.

Немаловажным достоинством возможностей «Умного дома» является контроль за такими системами, как электричество, водоснабжение и отопление. Автоматизированные датчики оповестят, если обнаружится утечка воды, возгорание, задымление в помещении, произойдут поломки в трубах или будет нарушена целостность проводки. Система сработает по двум направлениям — во-первых, она включит необходимые системы устранения аварий, а во-вторых, оповестит собственника, соответствующие аварийные службы и управляющие компании. Большое преимущество такого подхода, что ущерб от наступления аварийных ситуаций при оперативной реакции может быть минимизирован. Кроме того, технологии «Умного дома» позволяют улучшить систему безопасности дома и обеспечить мобильность и гибкость контроля. Управлять домом можно дистанционно, даже если речь идет о больших расстояниях (например, из других стран). Наконец, технология «Умный дом»

может оказывать поддержку самим жильцам. Например, встроенные камеры и усилители звука позволят знать обо всем, что происходит в большом доме: о местонахождении ребенка и информировать об этом родителей, если ребенок попытается открыть окно, оно заблокируется, если засунет пальцы в розетку, система отключит электричество, если пожилому человеку станет плохо и он упадет на пол вызовет бригаду скорой помощи и т.д.

Второе немаловажное преимущество технологии «Умный дом» — гарантия высокого уровня комфортабельности жилища. В таком доме многие системы поддержания помещения в порядке поручаются роботу: уборка, мытье посуды, глажка белья, загрузка белья в стиральную машину — все это может выполняться в автоматизированном режиме. Поддержание благоприятного климата в помещении также возлагается на различные электронные системы. Климат-контроль, интенсивность освещения, температурный режим, уровень влажности, проветривание и очистка воздуха от вредных примесей управляются с помощью различных систем, поддерживаются через мобильные приложения или установленные в доме приборы. Различные системы обслуживания максимально удобны в эксплуатации, поскольку управляются дистанционно, например, по голосовой команде можно подозвать робота-пылесоса или через команду в приложении. Также можно распорядиться о приготовлении ванны или включении чайника. Благодаря таким разработкам для помещения можно задать наиболее комфортные и, что наиболее важно, индивидуальные параметры, в том числе для каждого из проживающих, не зависеть напрямую от централизованных систем коммунальных услуг и в целом максимизировать комфорт от пребывания в доме.

Помимо прочего, технология «Умный дом» позволяет экономить потребление ресурсов за счет оптимизации расходования [Анненков, Рыбкин, 2013]. Поскольку при выполнении различных операций расходуется всегда определенное количество воды или электроэнергии, то избыточных и неучтенных расходов можно будет успешно избегать. Так, при автоматической подаче воды в стиральную или посудомоечную машину, или при поливе участка

расходы воды, как правило, снижаются. Также автоматизированное использование отопления в зависимости от повышения или понижения температуры в квартире должно способствовать устранению необоснованного расходования ресурсов. Точно так же при установке датчиков уличного света можно автоматизировать регулирование света в квартире. При дополнении этой функции датчиками движения можно добиться того, что свет будет включаться и выключаться автоматически в зависимости от присутствия или отсутствия в помещении человека. Также с помощью датчиков тепла или движения «умный дом» позаботится о выключении света в тех комнатах, где никого нет, если пользователь сам забыл это сделать. Подача тепла будет зависеть от смены температуры за окном, а яркость света — от времени суток.

Технологии «Умный дом» успешно внедряются в России в последние 10–20 лет. Пока рано говорить о централизованном оснащении домов ИИ-технологиями, однако широкое развитие и внедрение систем «Умный дом» входит в планы усовершенствования коммунальных услуг ЖКХ в России в ближайшее будущее [Дементьев, 2016].

В России ГИС ЖКХ — это единая федеральная централизованная информационная система, запущенная в опытном режиме в 2016 году, функционирующая на основе программных и технических средств и информационных технологий. Они обеспечивают сбор, обработку, хранение, предоставление, размещение и использование информации о жилищном фонде, стоимости и перечне услуг по управлению общим имуществом, о работах по содержанию и ремонту общего имущества в многоквартирных домах. В ней аккумулируется информация о предоставлении коммунальных услуг и поставке необходимых ресурсов, размере платы за жилое помещение и коммунальные услуги, задолженности по указанной плате, объектах коммунальной и инженерной инфраструктур, а также иной информации, связанной с жилищно-коммунальным хозяйством.

Согласно порталу ГИС ЖКХ, устройствами учета потребления электрической энергии оборудованы более 43 млн узлов учета,

устройствами учета потребления газа более 14,5 млн. Таким образом, объем только этого сегмента ЖКХ, по самой скромной оценке, составляет около 250–300 млрд рублей. Пока совсем незначительно охвачены области умного освещения и альтернативных источников энергии на городских и дворовых территориях, практически не контролируется потребление энергоресурсов в местах общего пользования (подъезды, подвалы), хотя анализ энергобаланса и возможность определять проблемы, приводящие к перерасходу, очевидно, экономически вполне обоснованы. Девелоперы и строители до последнего времени занимались продажей квартир в домах, построенных по технологиям XX века. То есть при их строительстве не учитывалась возможность создания «умного» здания. Установка интеллектуальных систем учета (ИСУ) на базе беспроводной технологии приводит к удорожанию квадратного метра недвижимости не более чем на 80 рублей, при этом обеспечивая инфраструктуру для «Умного дома». По прогнозу J'son & Partners Consulting, в 2018–2022 годах в РФ ежегодно будет подключаться к сети 3–5 млн новых устройств. К 2022 году их число достигнет более 40 млн единиц, а объем рынка составит 86 млрд рублей.

Само внедрение данной технологии требует некоторых усилий, в частности, детального исследования рынка, привлечения заинтересованных инвесторов, изучения потребностей самих пользователей и адаптации под российские системы коммунальных служб. Большая часть этих задач — технического и экономического характера, однако ряд проблем по внедрению и эксплуатации ИИ-технологий в области ЖКХ связан и с этическими вопросами. Система «Умный дом» при многочисленных своих удобствах, тем не менее, не лишена определенных трудностей социального и этического характера.

4.1. ИИ в области ЖКХ

1. Одной из важнейших составляющих технологии «Умный дом» является автоматизированный сбор данных об эксплуатации помещения. Для того чтобы оптимально регулировать функционал по-

мещения с помощью автоматизированного контроля, ИИ должен наладить качественную обратную связь с помещением. В частности, ему требуется аккумулирование и анализ тех данных, которые поступают из самого эксплуатируемого помещения. Если в вечерние часы вторников и четвергов электричество расходуется меньше, а в другие дни больше — ИИ-система запишет этот результат. В следующие вечера вторников и четвергов система автоматизированной подачи электроэнергии может учесть данные прошлых расходований. Чтобы ИИ-система эксплуатации помещения была действительно гибкой, ей требуется фиксировать любые мелочи пользования помещением и корректировать услуги в зависимости от этих данных. Системе нужно предлагать тот объем услуг, который вырисовывается как наиболее реалистичный с точки зрения аналитического прогноза, сделанного на основе собранных данных. В этом смысле программа ИИ, встроенная в систему «Умный дом», работает по аналогии с интерактивными тестовыми системами, которые подстраиваются под ответы тестируемого, или как рекламные компании, которые собирают данные о пользователях из соцсетей и других интернет-ресурсов, чтобы предложить им потенциально необходимые услуги. Во всех этих случаях на входе срабатывает сбор и анализ данных об особенностях потребления, а на выходе предложение об услуге, наиболее релевантной с точки зрения обнаруженных закономерностей.

Кейс. *После внедрения автоматизированной системы сбора данных по расходу электроэнергии системами Мосэнерго конечным потребителям стали предлагаться усредненные расчеты затрат электроэнергии исходя из прошлых данных. При работе с собственниками нежилых помещений, сдаваемых в аренду, данная программа предлагает не только расчет средних показателей, но и предположения о смене арендодателей, а также типе их производственной деятельности. Так, при значительном изменении объемов трат электроэнергии система может выдать предположение, что назначение помещения было изменено с представительского офиса под ночной клуб или круглосуточный магазин.*

Использование системой «Умный дом» технологии сбора и анализа данных по эксплуатации помещения порождает ряд этических недоразумений. Как и прочие системы ИИ, технологии «Умного дома» достаточно внимательны и ресурсоемки, чтобы накопить огромный объем данных о привычках владельцев домов, их частных пристрастиях, нормах потребления, затратах и расходах, организации быта и прочих особенностях эксплуатации помещений. Определенное опасение у хозяев домов могут вызывать возможная утечка информации или ее попадание в руки третьих лиц. Дальнейшее использование данной информации может скрывать в себе возникающие угрозы безопасности жизни и благополучию владельцев домов. Фактически доступ к системам управления домом означает доступ к информации об образе жизни семьи. По данным о том, как и что работает в помещении, можно почти полностью воссоздать уклад семьи, привычки хозяев и всех домочадцев, реконструировать потребности, склонности, определенные традиции и даже ценности, которых они придерживаются. Если данная информация попадет в руки злоумышленником или мошенников, различные злоупотребления окажутся вполне возможными. Также если подобная информация станет достоянием коммерческих структур, то это может привести к активному предложению специализированной рекламы, в получении которой собственники дома не заинтересованы.

Также определенные этические затруднения связаны с тем, что, если информация собирается, классифицируется и анализируется, то мы можем получить сведения и привычках, тратах или контактах человека в объемах, многократно превышающих персональное знание человека о самом себе. Далеко не всем собственникам может понравиться, что системы эксплуатации в итоге знают о них больше, чем они сами о себе. Предоставление подобной информации третьи лицам может привести к нежелательным, опасным и вредным для жителей «умных» домов последствиям. Фактически речь идет о нарушении неприкосновенности частной жизни граждан. Собственники могут быть очень недовольны тем фактом, что данные, которые аккумулируются системой, анализи-

руются по не вполне понятным самим владельцам критериям и параметрам и, таким образом, могут приводить к неожиданным выводам. Фактически система может формировать такие сведения о человеке, которые не очевидны для него самого. Подобная информация вдвойне деликатна, поскольку не подлежит прямому контролю со стороны человека. Эта информация превышает его знание о самом себе, и с этической точки зрения ее единственным возможным адресатом мог быть сам субъект. Во всех других случаях распространения данной информации речь будет идти о направленной слежке с возможными злоупотреблениями в последующем. Этическая сложность данной проблемы в том, что создается образ человека, который может не соответствовать его внутреннему образу и собственному представлению о самом себе. Подобный «объективный дубликат» человека может стать источником различного рода этически некорректного использования. Самое безобидное следствие формирования так называемых потребительских двойников — их возможное использование для неадекватной контекстной рекламы, различных коммерческих предложений и проч. Однако он также может стать материалом для возможных мошеннических действий в случае утечки или нарушения конфиденциальности. Владельцы домов могут стать жертвами шантажа или вымогательства в результате перепродажи данных о частной жизни домовладельцев. Мошенники могут специально охотиться за такими данными, зная о том, что современные коммунальные системы на основе технологий «Умного дома» аккумулируют подобные сведения о людях, получение которых в недавнем прошлом было весьма трудоемким и дорогостоящим процессом.

2. Одной из важнейших возможностей технологии «Умный дом» является повышенное внимание к усовершенствованию безопасности жилища. Во многих домах уже давно используются системы круглосуточной видеозаписи. Снятые с видеокамер записи хранятся разное по продолжительности время и в случае необходимости предоставляются заинтересованным лицам. При внедрении технологий «Умный дом» организацию видеозаписей предполагается значительно расширить. Видеозаписи могут включаться, реагируя на

датчики движения, они могут быть установлены в квартирах и домах вблизи сложного оборудования, чтобы зафиксировать момент поломки. Например, режим круглосуточной видеозаписи может быть включен в подвальном помещении, что позволит точнее установить причину разрыва труб или возникновения протечек.

Кейс. *При прохождении регистрации в современных аэропортах, а также при прохождении паспортного контроля в последние несколько лет используется автоматизированная система регистрации, а также снятия отпечатков пальцев. Многие из этих систем хранят в памяти историю перелетов граждан, а также данные их дактилоскопии. При этом указанные данные и в первую очередь данные дактилоскопии становятся автоматически собственностью иностранных государств.*

Как было отмечено выше, важной этической проблемой применения технологий «Умный дом» является неприкосновенность частной жизни. Беспокойствам и опасениям граждан способствует ряд мер, активно внедряемых в различных общественных местах. В целом не все граждане относятся позитивно к тенденции повышенного внимания к безопасности в ущерб неприкосновенности частной жизни. Многие методы профилактики преступлений, такие как, например, расширенные законы о прослушивании телефонных разговоров, комплекс камер наблюдения не только в общественных местах, но иногда и в приватных зонах для круглосуточного мониторинга, расширение биометрических возможностей и установка датчиков, интеграция личных данных граждан в специальные базы данных длительного хранения и т.д. Известно, что подавляющая часть действий, совершаемых гражданами через мобильные приложения или социальные сети, аккумулируется и сохраняется. В случае если эта информация перестанет быть конфиденциальной, то граждане окажутся перед фактом, что отныне о них известно все или почти все. Развитие технологий «Умный дом» вносит свою лепту в этот неоднозначный процесс. Фактически организация поддержки жилища в режиме «Умного дома» предпола-

гает круглосуточный и достаточно приватный видеомониторинг жизни домовладельцев [Arkin, Borenstein, 2016], при этом на записи непосредственно могут попадать самые интимные стороны жизни граждан. Если речь идет о применении технологий «Умный дом» в пределах больших жилищных комплексов, домовладений, многоквартирных домов, таунхаусов или общежитий, то, помимо обычных камер слежения, в соответствующей литературе упоминаются полицейские роботы, которые могут вести тайное наблюдение на расстоянии, например, обнаруживать скрытые наркотики или оружие, а также незаметно идентифицировать лица. Такие видеорегистраторы могут быть подключены к базам данных и проводить автоматическую проверку личных данных лиц, если в их отношении возникли какие-то подозрения [Sharkey, 2019]. Функция идентификации лица с привлечением биометрических данных может быть подключена как консьерж-служба здания. Управляющие компании, подключившие к обсуживаемым домам функцию «Умный дом», могут оказывать помощь различным службам безопасности и в том числе полиции, поскольку могут снабжать их необходимыми данными в случае необходимости.

Фактически коммунальная служба обслуживания и эксплуатации дома становится хранителем личных данных жильцов. В качестве дополнительной меры безопасности на входе в здание могут быть установлены распознаватели лиц, так что в здание смогут войти только постоянные жильцы или те, кого они приглашают. Для домашней безопасности роботы могут быть оснащены устройствами наблюдения, которые могут контролироваться третьими сторонами или теми, кто получит для этого доступ. Несмотря на технологическую автоматизацию охраны и безопасности жильцов и их собственности, далеко не все граждане с оптимизмом относятся к перспективе находиться под круглосуточным контролем видеокамер. Повсеместное рассредоточение приборов слежения и видеозаписи может восприниматься как тотальная слежка и контроль за частной жизнью. Жильцам может быть некомфортно от мысли о том, что данные об их частном быте могут попасть в руки третьих лиц, что возможно при современном уровне развития ха-

керских технологий. Определенную этическую трудность представляет собой опасение граждан о том, что данные о частной жизни могут стать предметом самых разных злоупотреблений. Риски мошенничества, шантажа, репутационных потерь, как и ухудшение отношений с домочадцами составляют суть тех опасений, которые побуждают людей критически относиться к некоторым из технологий «Умного дома». Многие способы последующего использования записываемых данных недостаточно продуманы и регламентированы. Именно это обстоятельство дает повод для этических споров. При постоянном сборе личных данных или видеосъемке люди могут не чувствовать себя в безопасности. Чувство дискомфорта, смущения, необходимость изменять поведение, тревожные состояния, невозможность расслабиться, будучи в одиночестве, все эти трудности вполне релевантны при обсуждении вопроса масштабов внедрения технологий «Умного дома» в ЖКХ. Вся трудность состоит в разграничении приватного пространства, которое должно оставаться неприкосновенным при любых обстоятельствах, и пространства коллективного пользования, которое может быть доступно для внешнего наблюдения. Если установка камер допускается на придомовых территориях, в подъездах, лифтах, на лестничных площадках, то можно ли распространить их размещение на помещения рядом с квартирами? Не ясно также может ли кто-то из третьих лиц получить доступ к данным видеонаблюдений, проводимых в квартирах, и при каких обстоятельствах. Служат ли, например, возгорание или затопление, возникшее в одной из квартир, основанием для просмотра данных видеозаписи в квартире, пострадавшими, от которых стали соседи по площадке? Что делать, если в квартире одного из соседей произошло ограбление, — допустимо ли в этом случае потребовать данные видеозаписей в приватных зонах жилища (в квартире другого соседа)? Этично ли будет запросить к просмотру материал квартирных видеозаписей в случае смерти собственника и невозможности запросить разрешение у его наследников? Также не вполне ясен статус возможных режимов видеонаблюдения. Например, не урегулирован вопрос о том, в какой момент наблюдение круглосуточных камер слежения или ро-

ботов с использованием технологий слежения может подпадать под категорию «виртуального обыска», для которого обычно требуется судебный ордер. Робот-шпион, установленный в доме для съемки оборудования и информирования о возникающих авариях, фактически осуществляет слежку. Кроме того, в его поле видения может попасть информация, которая ему не предназначалась. В какой степени жильцы должны быть информированы о видеозаписях и должны ли они давать письменное согласие о съемке в местах проживания — эти вопросы пока не получили должного прояснения.

3. Одним из важнейших принципов работы технологии «Умный дом» является максимизация комфорта проживания, улучшение качества жизни и возрастание чувства защищенности от пребывания в жилище. Автоматизация многих функций должна не только высвободить время домовладельцев, но и улучшить контроль за жилищем, практически сделав его всесторонним. При подключении технологий «Умный дом» собственники домов смогут автономизироваться от коммунальных служб, обрести большую независимость от управляющих компаний и самостоятельно решать многие задачи коммунального обслуживания.

***Кейс**. На конференции специалистов по компьютерной безопасности DefCon в 2016 году группа исследователей провела зрелищную инсталляцию на тему того, как можно легко взломать умный дом [Vincent, 2016]. Они привели пример с блокировкой термостата таким образом, что владелец не может больше изменять температуру. Другие бытовые приборы, имеющие подключение к Wi-Fi, тоже находятся под угрозой взлома хакерами и прочими недоброжелателями. Ясно, что и вся система безопасности может быть заблокирована через тот же канал, через который осуществляется управление — а именно дистанционно.*

Несмотря на многие преимущества технологии «Умный дом» в части повышения самостоятельности и автономии домовладельцев, а также их независимости от различных внешних коммунальных служб, реакция на активное внедрение может быть прямо противо-

положной. Одной из этических дилемм внедрения ИИ-технологий в области ЖКХ является феномен чувства беспомощности, которое характеризуется чувством утраты контроля над происходящим и опасений в возникновении возможных чрезвычайных ситуаций. Многие люди, чья активная деятельность связана с использованием различных автоматизированных систем, неизменно отмечают большой прилив энтузиазма, связанный с эффективностью и быстротой решения многих задач. Однако определенная часть пользователей отмечает некоторую угнетенность в связи с непониманием того, как в действительности работают «умные технологии», и поэтому они могут чувствовать себя более уязвимыми и незащищенными.

Внедрение автономизированных технологий на повседневном уровне, а именно в сфере каждодневной эксплуатации собственных жилищ, может усилить чувство потери контроля над приватным пространством, в котором человек находится большую часть времени. Этически неоднозначным является тот факт, что технологии, которые призваны способствовать повышению чувства защищенности, уверенности и независимости, могут приводить к обратным результатам. У пользователей ИИ-технологий может появляться ощущение многократно возросшей зависимости от различных технических устройств и приспособлений [Danaher, 2010]. Люди отмечают возрастающий страх от осознания того, что поломка одного-единственного устройства или выход из строя программного приложения может привести к катастрофическим последствиям.

По заказу UK Post Office было проведено исследование феномена, получившее название номофобии (страха потери доступа к сотовому телефону), которое показало, что около 53% людей в Великобритании, пользующихся мобильными телефонами, испытывают тревогу: если теряют его, если он отключается, если заканчиваются деньги на счету или пропадает покрытие сотовой сети. В этом же исследовании было показано, что по степени интенсивности панического страха до 45% людей ставят степень паники при потере сотового телефона на второе место после ощущения паники от угрозы оказаться в центре террористической атаки [D'Agata, 2019].

В большинстве случаев респонденты указывали, что данные страхи связаны со страхом утратить контроль над всеми банковскими счетами и прочими финансовыми инструментами, которые могут попасть в руки злоумышленников.

Ощущение мгновенной утраты всякого контроля весьма характерно и при возникновении проблем при эксплуатации систем «Умный дом». У домовладельцев может возникать ощущение, что в случае возможной поломки вся система одновременно выйдет из строя. Так, при возможном доступе злоумышленников к системе обслуживания дома они получают неограниченный доступ ко всем уровням жизнедеятельности, что несет риски, в том числе, для жизни и здоровья самих домовладельцев.

Как и всякая система, управляемая дистанционно, «Умный дом» подвержен взлому со стороны. Если запускать систему со стационарного компьютера или ноутбука, то высока вероятность внедрения вирусов в систему. Безопасность также зависит от того, к какой сети подключается пользователь для управления домом — общественной или частной. Многие из этих нюансов эксплуатации могут приводить к обратным результатам, а именно увеличивать психологическое чувство тревожности у собственников, ощущение, что их жилище недостаточно защищено или слишком уязвимо при воздействии различных технологических катастроф, при технических поломках и нестабильной работе программного обеспечения.

4.2. Перечень рекомендаций

Внедрение ИИ-технологий в сферу ЖКХ требует такой же этической поддержки, как и для областей здравоохранения, образования и судебной системы. Государству следует как можно тщательнее продумать возможности сохранности приватного пространства, а также определить критерии отделения частного от общественного. Этической экспертизе подлежит само определение приватного. В связи с виртуализацией многих областей жизнедеятельности современных индивидуумов требуется как можно яснее задать параметры отделения того пространства, где вступают в силу зако-

ны, оберегающие неприкосновенность частной жизни граждан. Эти данные следует отделять от тех, где работает приоритет защиты коллективных интересов, а также обеспечение коллективной безопасности. Внедрение технологий «Умный дом» напрямую касается этих вопросов.

Автоматизированная поддержка коммунальных услуг, эксплуатации помещений и безопасности пребывания в них максимизирует комфорт, однако порождает ряд определенных трудностей, в том числе, этического характера. Для их устранения требуется тесное сотрудничество с коммунальными службами, которые должны быть готовы всегда прийти на помощь в случае возможных перебоев с автоматизированными службами и услугами. В этой части был бы уместен контроль со стороны органов государственного управления, осуществляющих надзор за соблюдением этических норм. В первую очередь он должен касаться защиты личных данных граждан от утечки и передачи в третьи руки.

Одним из способов, реализующих данное требование, могло бы стать создание независимых центров, разрабатывающих технологии шифрования и защиты данных. По аналогии со службами безопасности, существующими в качестве подразделений банков и других финансовых учреждений, было бы целесообразно внедрить аналогичные службы во все организации, так или иначе осуществляющие сбор и использование личных данных граждан.

Организации жилищно-коммунального комплекса не могут стать исключением. Управляющие компании могут внедрять специальные подразделения, отвечающие за сохранность всех личных данных, поступающих в ведение УК в рамках предоставления услуг «Умный дом». Также требуется некоторая автономия и гибкость выбора на рынке коммунальных услуг. Любому собственнику жилья должна быть предоставлена возможность беспрепятственной смены управляющей компании, предоставляющей услуги «Умного дома», и замену его на другую компанию с аналогичным доступом к подобным услугам. Можно прогнозировать крайне негативные последствия для развития рынка технологий искусственного интеллекта в сфере ЖКХ в случае, если эти услуги будут монопо-

лизированы ограниченным количеством управляющих компаний. Также следует предоставить равный доступ всем жильцам одного дома к данным услугам. Унификация доступа необходима для равной заинтересованности собственников в соблюдении всех норм и требований обеспечения помещений соответствующими технологиями искусственного интеллекта.

Важной этической стороной технологий «Умного дома» является защита личных данных собственников, формируемых на базе доступа к их частному жилью. Для того чтобы автоматизированные системы коммунальных служб работали бесперебойно, им требуется аккумулировать и анализировать большие объемы данных, имеющих отношение к частной жизни граждан. Этот факт может вызывать напряжение у многих собственников жилья, так как формирование целых кластеров информации, содержащих данные о частном быте людей в объеме, подчас превышающем их собственную осведомленность, может таить в себе угрозу и риски быть использованными во вред и против интересов домовладельцев.

Следует разработать комплекс мер, направленных на профилактику нежелательных последствий от сбора данных. Поскольку в случае возможной утечки, хакерских атак, хищения данных мошенниками для последующих злоупотреблений данные могут быть использованы не в интересах граждан, следует установить критерии сегрегации сугубо приватных данных от нейтральных, а также процедуры отказа жильцов от предоставления приватной информации.

Аналогичным образом следует продумать также комплекс мер, направленных на снятие напряжения от осознания избыточной автоматизированности жилого пространства. Чувство утраты контроля над личным пространством, незащищенности от возможных поломок и непрозрачности технологий, управляющих повседневными практиками, могут вселять тревогу и дискомфорт. Возрастающее ощущение сверхценности технологий приводит к обратному результату, а именно возможной девальвации идеи тотальной автоматизации и цифровизации.

Работа с этой проблемой, исследование ее логики и причин лежит за пределами технической стороны внедрения искусственного

интеллекта в сферу организации быта и ведения хозяйства и, скорее, относится к этической области. Данные исследования должны проводиться специалистами соответствующего профиля, а именно квалифицированными экспертами по гуманитарным дисциплинам и социальным наукам.

Заключение

Развитие компьютерных и цифровых технологий в современном мире, а также активное развитие и внедрение искусственного интеллекта актуализируют новые проблемы человека и его существование в мире. Большая часть этих проблем непосредственно связана с этическими трудностями совмещения цифровых технологий и повседневной жизни человека. Перспективы внедрения различных компьютерных технологий в жизнь человека неоспоримы, однако они сопровождаются множеством серьезных и не всегда явных опасностей и рисков. В той мере, в которой государство прямо или косвенно участвует во внедрении и принятии управленческих решений в отношении использования искусственного интеллекта, оно заинтересовано в том, чтобы этические аспекты цифровизации принимались во внимание.

Индустрия искусственного интеллекта и робототехники в современном мире является одной из самых активно развивающихся отраслей. Вместе с тем при ее развитии, если и принимаются во внимание экономические законы и принципы развития рынка, они не вполне оснащены внутренней этической экспертизой, а также своего рода философской рефлексией, позволяющей понимать свою роль. Социально-философская рефлексия нередко медленно «догоняет» технологические инновации, что приводит к образованию «этического вакуума» [Moor, 1985]. Такой эффект означает, что ИИ-технологии развиваются по своим собственным законам, и это приводит к ряду личностных коллизий и противоречий, требующих решения с точки зрения философского анализа и этической экспертизы. Необходимо провести ревизию существующих этических лакун в области использования искусственного интеллекта и наметить пути их устранения.

В исследованиях применения искусственного интеллекта ключевым вопросом станет преодоление противоречий между современным «цифровым» образом человека, для которого использование технологий стоит наряду с его личными компетенциями, и человеком как субъектом, обладающим сознанием, волей, свобо-

дой выбора, а также этическими интуициями, которые трудно формализуются в технологиях искусственного интеллекта. Существует целый класс этических проблем, в том числе аксиология искусственного интеллекта, решение которых не может быть заменено никакими другими сугубо формальными технологиями. В значительной мере такая работа может быть адресована внешним экспертным союзам, профессиональным сообществам философов, психологов, социологов, различным независимым группам исследователей самого широкого профиля и т.д.

Не менее важен и общественный контроль за этой сферой. Однако необходимы государственный мониторинг и помощь государства в принятии и применении определенных принципов регулирования этических требований к развитию искусственного интеллекта. При использовании технологий ИИ в государственных масштабах и различных экономических секторах и отраслях нужно принимать во внимание гуманитарную и гуманистическую составляющие данного проекта.

В подготовленном тексте были обзорно (выборочно) представлены некоторые из возможных отраслей, где развитие технологий искусственного интеллекта обещает стать широко и повсеместно применяемым уже в ближайшие годы. Непосредственно данные области были выбраны также потому, что именно в них активно происходит внедрение технологий искусственного интеллекта уже сейчас.

Авторы книги надеются, что смогли внести посильный вклад в развитие общих ориентиров, направленных на поддержку этической стороны внедрения искусственного интеллекта в сферу государственного управления в России, а также обратить внимание читателя непосредственно на философско-этический уровень экспертной разработки. Основанием для радости является тот факт, что авторам книги удалось провести собственно междисциплинарное исследование и показать, как и в каком смысле специалисты из разных областей (государственного управления и философии) могут оказать друг другу посильную аналитическую помощь.

Итогом и заключительной идеей данной работы могло бы стать следующее — если искусственный интеллект и прочие технологии

призваны служить людям, оценивать и судить их должны именно люди, а не технологии. Победу в деле создания сложнейших технологий человечество уже одержало, однако остается выработать основательную этическую позицию по вопросам их применения в ходе обсуждения между исследователями и государственными управленцами.

Библиография

Альманах «Искусственный интеллект № 5: Предиктивная аналитика и системы поддержки принятия решений». ЦНТИ МФТИ, 2020. С. 169.

Анненков З.С., Рыбкин Е.Г. «Умный дом» — интеллектуальное здание // Материалы научно-технической конференции «Роль молодых ученых в обеспечении продовольственной и экологической безопасности России». М.: ФГБОУ ВПО МГУП, 2013. С. 3–6.

Дементьев А. «Умный» дом XXI века. Издательские решения, 2016. 142 с.

Добролюбова Е.И. Государственное управление по результатам в эпоху цифровой трансформации: обзор зарубежного опыта и перспективы для России // Вопросы государственного и муниципального управления. 2018. С. 70–93.

Кашанин А.В., Козырева А.Б., Курносова Н.А., Малов Д.В. Информационные технологии в правосудии: состояние и перспективы. Россия и мир. Аналитический доклад / под ред. А.В. Кашанина. НИУ ВШЭ, 2020. URL: https://ilr.hse.ru/news/379660230.html.

Клименко А.В., Минченко О.С. Государственное регулирование экономики: вопросы теории и лучшая практика // Вопросы государственного и муниципального управления. 2016. № 3. С. 7–30.

Кнутов А.В., Плаксин С.М. Индикаторы риска при осуществлении государственного контроля (надзора) // Законодательство. 2019. № 5. С. 36–45.

Нагорная М. Искусственный интеллект в судопроизводстве // Новая адвокатская газета, 2019. URL: https://www.advgazeta.ru/novosti/advokaty-i-yuristy-ob-iskusstvennom-intellekte-v-sudoproizvodstve/.

Национальная стратегия развития искусственного интеллекта на период до 2030 года. URL: http://publication.pravo.gov.ru/Document/View/0001201910110003.

Шевелёва Н.А. (науч. ред.). Право и информация: вопросы теории и практики: сб. матер. междунар. науч.-практ. конф. (Сборники

Президентской библиотеки. Сер. «Электронное законодательство». Вып. 7). СПб.: Президентская библиотека, 2018. 271 с. URL: https://www.prlib.ru/item/1168067/.

Aletras N., Tsarapatsanis D., Preoţiuc-Pietro D., Lampos V. Predicting judicial decisions of the European Court of Human Rights: A Natural Language Processing perspective // PeerJ Computer Science 2016. URL: https://peerj.com/articles/cs-93/.

Arkin R.C. The Robot Didn't Do It: A Position Paper for the Workshop on Anticipatory Ethics, Responsibility and Artificial Agents // Workshop on Anticipatory Ethics, Responsibility and Artificial Agents, 2013.

Arkin R., Borenstein J. Robots, Ethics, and Intimacy: The Need for Scientific Research // M.V. D'Alfonso, D. Berkich (eds). On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence. Springer Verlag, 2016. P. 299–309.

Asada M. Artificial Pain May Induce Empathy, Morality, and Ethics in the Conscious Mind of Robots // Philosophies. 2019. No. 38.

Bankins S., Formisa P. When AI Meets PC: Exploring the Implications of Workplace Social Robots and a Human-Robot Psychological Contract // European Journal of Work and Organizational Psychology. 2019. No. 29 (2). P. 215–229.

Barr A. Handbook of Artificial Intelligence / E. Feigenbaum. Los Altos, CA: William Kaufman, 1989. 442 p.

Bhorat Z. Do we still need human judges in the age of Artificial Intelligence? // open Democracy, 2017. URL: https://www.opendemocracy.net/en/transformation/do-we-still-need-human-judges-in-age-of-artificial-intelligence/.

Boyles R.J.M., Joaquin J.J. Why Friendly AIs Won't Be That Friendly: A Friendly Reply to Muehlhauser and Bostrom // AI and Society. 2019. P. 1–3.

Brożek B., Jakubiec M. On the Legal Responsibility of Autonomous Machines // Artificial Intelligence and Law. 2017. No. 25 (3). P. 293–304.

Burr C., Cristianini N., Ladyman J. An Analysis of the Interaction Between Intelligent Software Agents and Human Users // Minds and Machines. 2018. No. 28 (4). P. 735–774.

Capurro R. (ed.). Ethics and Robotics. IOS Press, 2009.

Coeckelbergh M. Responsibility and the Moral Phenomenology of Using Self-Driving Cars // Applied Artificial Intelligence. 2016. No. 30 (8). P. 748–757.

Coppin B. Artificial Intelligence Illuminated. Jones & Bartlett Learning, 2004.

Council of Europe, 1990. Recommendation R (90)13 of the Committee of Ministers to Member States on Prenatal Genetic Screening, Prenatal Genetic Diagnosis, and Associated Genetic Counselling // Texts of the Council of Europe on Biomedical Matters. Strasbourg, 1993. P. 53–55. CDBI/INF/(93)2.

Council of Europe. Convention for the Protection of Human Rights and Dignity of Human Being with regard to the application of biology and medicine: Convention on Human Rights and Biomedicine. Strasbourg, 1997. DIR/JUR 96,14.

D'Agata Ch. Nomophobia: Fear of being without your cell phone // CBS News. April 3, 2019.

Danaher J. Toward an Ethics of AI Assistants: An Initial Framework // Philosophy and Technology. 2010. No. 31 (4). P. 629–653.

Danaher J. The Rise of the Robots and the Crisis of Moral Patiency // AI and Society. 2019. No. 34 (1). P. 129–136.

Faggella D. AI and Machine Vision for Law Enforcement — Use-Cases and Policy Implications, 2019. URL: https://emerj.com/ethics-and-regulatory/ai-and-machine-vision-for-law-enforcement-use-cases-and-policy-implications/.

Feil-Seifer D., Matarić M.J. Dry Your Eyes: Examining the Roles of Robots for Childcare Applications // Interaction Studies: Social Behaviour and Communication in Biological and Artificial Systems. 2010. No. 11 (2). P. 208–213.

Floridi L., Taddeo M., Mittelstadt B., Wachter S., Cath C. Artificial Intelligence and the 'Good Society': The US, EU, and UK Approach // Science and Engineering Ethics. 2018. No. 24 (2). P. 505–528.

Fossa F. Legal Fictions and the Essence of Robots: Thoughts on Essentialism and Pragmatism in the Regulation of Robotics // M. Coeckelbergh, J. Loh, M. Funk, J. Seibt, M. Nørskov (eds). Envisioning Robots in Society — Power, Politics, and, Public Space. Amsterdam: IOS Press, 2018. P. 103–111.

Galliott J., Reed W. (eds). Ethics and the Future of Spying: Technology, National Security and Intelligence Collection. Routledge, 2016.

Giubilini A., Savulescu J. The Artificial Moral Advisor. The "Ideal Observer" Meets Artificial Intelligence // Philosophy and Technology. 2018. No. 31 (2). P. 169–188.

Goel A.K., David A. Joyner. Design of an Online Course on Knowledge-Based AI // AAAI. 2016. P. 4089–4094.

Haidt, J., Kesebir S. Morality // S. Fiske, D. Gilbert, G. Lindzey (eds.) Handbook of Social Psychology. 5th ed. Hobeken, NJ: Wiley, 2010. P. 797–832.

Hakli R., Mäkelä P. Moral Responsibility of Robots and Hybrid Agents // The Monist. 2019. No. 102 (2). P. 259–275.

Hibbard Bill. Super-intelligent machines // ACM SIGGRAPH Computer Graphics. 2001. Vol. 35. No. 1. P. 11–13.

Ho A. Deep Ethical Learning: Taking the Interplay of Human and Artificial Intelligence Seriously // Hastings Center Report. 2019. No. 49 (1). P. 36–39.

Holm S., Voo T.C. Brain-Machine Interfaces and Personal Responsibility for Action — Maybe Not As Complicated After All // Studies in Ethics, Law, and Technology. 2011. No. 4 (3).

Khoshnood B., De Vigan C. Trends in prenatal diagnosis, pregnancy termination, and perinatal mortality of newborns with congenital heart disease in France, 1983–2000: A population-based evaluation // Pediatrics. 2005. No. 115. P. 950–101.

Lamanna C. Should Artificial Intelligence Augment Medical Decision Making? The Case for an Autonomy Algorithm // AMA Journal of Ethics. 2018. No. 9 (20).

Livingston S., Risse M. The Future Impact of Artificial Intelligence on Humans and Human Rights // Ethics and International Affairs. 2019. No. 33 (2). P. 141–158.

Matthias A. The Responsibility Gap: Ascribing Responsibility for the Actions of Learning Automata // Ethics and Information Technology. 2004. No. 6 (3). P. 175–183.

McDougall R.J. Computer Knows Best? The Need for Value-Flexibility in Medical AI // Journal of Medical Ethics. 2019. No. 45 (3). P. 156–160.

Michelle A., DeBaets M. Can a Robot Pursue the Good? Exploring Artificial Moral Agency // Journal of Evolution and Technology. 2014. No. 24 (3). P. 76–86.

Miller K., Wolf M., Grodzinsky F. Applying a Social-Relational Model to Explore the Curious Case of hitchBOT // M.V. D'Alfonso, D. Berkich (eds). On the Cognitive, Ethical, and Scientific Dimensions of Artificial Intelligence. Springer Verlag, 2019. P. 311–323.

Mishra A. Moral Status of Digital Agents: Acting Under Uncertainty // V. Müller (ed.). Philosophy and Theory of Artificial Intelligence. Springer, 2017.

Modell B. Current Trends in Early Prenatal Diagnosis. Geneva: WHO, 2000.

Monegain B. IBM Watson pinpoints rare form of leukemia after doctors misdiagnosed patient // Healthcare IT News, 2016. URL: https://www.healthcareitnews.com/news/ibm-watson-.

Moor James H. What is computer ethics? // Metaphilosophy. 1985. Vol. 16. No. 4. P. 266–275.

Muehlhauser L, Helm L. The singularity and machine ethics // Singularity Hypotheses. Berlin, Heidelberg: Springer, 2012. P. 101–126.

Müller V.C. Exploring the Coexistence of Human and Artificial Intelligence: Ethics of Responsibility and Literary Imagination // Journal of Ethics. 2019. No. 1 (124). P. 17–35.

Muntean I., Howard D. Artificial Moral Cognition: Moral Functionalism and Autonomous Moral Agency // Th. Powers (ed.). Philosophy and Computing:

Essays in Epistemology, Philosophy of Mind, Logic, and Ethics. Springer, 2017.

Myungsik K. Artificial Intelligence and Environmental Ethics: Moral, Legal Right of Artificial Intelligence // Environmental Philosophy. 2018. No. 25. P. 5–30.

Nilsson N.J. Artificial Intelligence: A New Synthesis. Elsevier, 1998.

Pei W.A.N.G. What Do You Mean by «AI»? // Artificial General Intelligence. 2008. P. 362–373.

Rahwan I. Society-in-the-Loop: Programming the Algorithmic Social Contract // Ethics and Information Technology. 2018. No. 20 (1). P. 5–14.

Raphael D. (ed.). Social Determinants of Health: Canadian Perspectives. Canadian Scholars' Press, 2009.

Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance) // Официальный журнал. 119. 04 May 2016. P. 1–88.

Richardson K. Rethinking the I-You Relation Through Dialogical Philosophy in the Ethics of AI and Robotics // AI and Society. 2019. No. 34 (1). P. 1–2.

Rouvroy A. Privacy, Data Protection, and the Unprecedented Challenges of Ambient Intelligence // Law and Ethics of Human Rights. 2010. No. 2 (1).

Russell S., Norvig P. Artificical Intelligence: A Modern Approach. Egnlewood Cliffs: Prentice-Hall, 1995.

Ryseweek S.P. van, Pontier M. (eds). Machine Medical Ethics. Springer, 2018.

Sharkey A. Autonomous Weapons Systems, Killer Robots and Human Dignity // Ethics and Information Technology. 2019. No. 21 (2). P. 75–87.

Strickland E. "AI Clinician" Makes Treatment Plans for Patients With Sepsis // IEEE Spectrum, 2018. URL: https://spectrum.ieee.org/the-human-os/biomedical/devices/ai-clinician-makes-treatment-plans-for-sepsis-patients.

Sun T.Q., Medagli R. Mapping the challenges of artificial intelligence in the public sector: Evidence from public healthcare // Government Information Quarterly. 2019. No. 36 (2). P. 368–383.

Tasioulas J. First Steps Towards an Ethics of Robots and Artificial Intelligence // Journal of Practical Ethics. 2017. No. 7 (1). P. 61–95.

Taylors A. Artificial Intelligence 'Judge' to Predict Outcome in European Court Trials? University Scientists Develop Software // University Herald. 2020. URL: https://www.universityherald.com/articles/45702/20161024/artificial-intelligence-judge-predict-outcome-european-court-trials.htm.

Tiefensee C., Marx J. Of Animals, Robots and Men // Historical Social Research. 2015. No. 40. P. 70–91.

Totschnig W. The Problem of Superintelligence: Political, Not Technological // AI and Society. 2019. No. 34 (4). P. 907–920.

Vallverdú J. Thinking Machines and the Philosophy of Computer Science: Concepts and Principles. IGI, 2010.

Vincent J. What counts as artificially intelligent? AI and deep learning, explained. Retrieved. The Verge, 2016. URL: https://www.theverge.com/2016/2/29/11133682/deep-learningaiexplained-machine-learning (accessed: 20.02.2019).

Warwick K. Artificial Intelligence: The Basics. 1st ed. London: Routledge, 2011.

Wiggers K. Health Tap's platform uses AI to dispense treatment advice // Venture Beat. 2018. URL: https://venturebeat.com/2018/08/21/health-taps-platform-uses-ai-to-make-treatment-recommendations/.

Woo-Kyu K. Consideration of Human Emotions About Artificial Intelligence — Focused on the Analysis of Newspaper Articles on Alpha Go vs Lee Sedol // Journal of Ethics. 2018. No. 1 (132) P. 181–201.

Yudkowsky E. Complex Value Systems in Friendly AI // J. Schmidhuber, K.R. Thórisson, M. Looks (eds). Artificial General Intelligence. AGI 2011. Lecture Notes in Computer Science. Vol. 6830. Berlin, Heidelberg: Springer, 2011.

Gasparyan Diana, Styrin Evgeny. Applied Problems of Artificial Intelligence Ethics Implementation in Russia: Industry Analysis and Justice System [Text] / D. Gasparyan, E. Styrin ; National Research University Higher School of Economics. — Moscow : HSE Publishing House, 2020. — 112 pp. — 500 copies. — ISBN 978-5-7598-2351-3 (pbk.). — ISBN 978-5-7598-2242-4 (e-book).

In this book we examine ethical problems which arise in the process of artificial intelligence (AI) technologies implementation in the industrial public administration in education, healthcare, public housing and utilities spheres as well as in the system of justice. In the beginning of the book we provide general analysis of applied problems, biases, dilemmas and difficulties related to the ethics of artificial intelligence technologies implementation in the selected areas, then we explain guiding principles and values of ethically oriented artificial intelligence and finally we analyze ethical problems of the AI implementation in the selected spheres. Each chapter of the book examines one of the areas and consists of two parts: in the first one we provide an overview of the ethical problems in the process of AI application and implementation; in the second chapter we propose a set of recommendations on possible resolution, elimination or minimization of the formulated ethical problems.

Научное издание

Гаспарян Диана Эдиковна, Стырин Евгений Михайлович

Прикладные проблемы внедрения этики искусственного интеллекта в России: отраслевой анализ и судебная система

Зав. редакцией *Е.А. Бережнова*
Редактор *Н.М. Дмуховская*
Художник *В.П. Коршунов*
Компьютерная верстка *Н.Е. Пузанова*
Корректор *Н.М. Дмуховская*

На обложке — фрагмент презентации робота Roboy, разработка Лаборатории искусственного интеллекта Цюрихского университета. 27 февраля 2013 г.
© Adrian Baer | Wikimedia Commons <Roboy_270213_1.jpg>

Национальный исследовательский университет
«Высшая школа экономики»
101000, Москва, ул. Мясницкая, 20
Тел.: 8 (495) 772-95-90 доб. 15285

Подписано в печать 21.12.2020. Формат 60×90/16
Гарнитура Newton. Усл. печ. л. 7,0. Уч.-изд. л. 5,2
Печать офсетная. Тираж 500 экз. Изд. № 2468. Заказ №