

ИССЛЕДОВАНИЯ ЭТИЧЕСКИХ ПРОБЛЕМ ПРИМЕНЕНИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА В БИБЛИОТЕЧНОМ И ИНФОРМАЦИОННОМ СЕКТОРЕ

ETHICS CASE STUDIES OF ARTIFICIAL INTELLIGENCE FOR LIBRARY AND INFORMATION PROFESSIONALS*

Эндрю Кокс (Andrew Cox), Университет г. Шеффилда, Факультет информационных технологий, Великобритания

Реферат. Искусственный интеллект (AI, ИИ) не только открывает захватывающие возможности для расширения доступности знаний, но и порождает множество этических проблем, касающихся пристрастности, прозрачности, объяснимости и ответственности, конфиденциальности, защиты и безопасности, а также того, как ИИ влияет на человеческий выбор и свободу. В статье подробно рассматриваются данные вопросы и в качестве примера ИИ анализируется использование ChatGPT, выдвигающее на первый план многие из них. Несмотря на то что в кодексах профессиональной этики высокого уровня прописаны необходимые принципы, в них нет ясного объяснения, как реагировать на назревающие проблемы в области ИИ. Однако сценарии, где затрагиваются этические соображения, позволяют в краткой и наглядной форме представить ключевые дилеммы, чтобы привлечь внимание к данной проблематике и побудить к ее обсуждению. В работе приведены восемь сценариев, в которых поставлены актуальные для специалистов библиотечного и информационного дела дилеммы, связанные с ИИ. Сценарии находятся в открытом доступе и могут быть использованы по лицензии CC-BY-SA.

Ключевые слова: искусственный интеллект — применение в библиотеках, ChatGPT, конфиденциальность, этика, защита персональных данных, сценарии.

Введение

Искусственный интеллект (AI, ИИ) открывает захватывающие возможности для специалистов в области библиотечного и информационного дела и пользователей. ИИ сулит расширение доступности знаний благодаря новым способам автома-

тизированного описания и извлечения информации из коллекций [1; 2]. Он позволит обеспечить гибкость и персонализацию при предоставлении информации. Основанные на ИИ чат-боты и голосовые помощники способны предоставить доступ к информации в режиме диалога и поддержки. Кроме того, ИИ можно применять для анализа и прогнозирования поведения пользователей. Необходимость разьяснения сути ИИ — это новый аспект информационной грамотности.

Вместе с тем ИИ поднял в обществе целый шквал опасений по поводу этических аспектов, особенно связанных с предвзятостью, осмысленностью и конфиденциальностью [3; 4]. На фоне этих сомнений несколько веб-сайтов занялись отслеживанием случаев нарушения этики в области ИИ, например репозиторий «ИИ, алгоритмические и автоматизированные инциденты и прения» (AIAAIC)¹ и «База данных инцидентов в области ИИ» (AIID)². В первом из источников с момента его создания в 2019 г. по март 2024 г. было зарегистрировано около 1400 инцидентов. В ответ на это многие организации опубликовали рекомендации по этическому использованию ИИ [5]. Существуют и полезные отраслевые руководства о том, как следует подходить к разработке ИИ с соблюдением этических норм в различных областях, например в образовании [6; 7]. Подобные целевые руководства бесценны для информационных специалистов, работающих в конкретных отраслях. Настоящая работа написана с позиции того, что сотрудники библиотек занимают уникальную ценностную позицию в отношении ИИ, которая и сама по себе заслуживает всестороннего изучения. Призвание библиотекаря к расширению доступности знаний привносит уникальный взгляд на этику использования ИИ, а информационным

* <https://www.degruyter.com/document/doi/10.1515/9783111336435-012/html?lang=en>

специалистам необходимо больше ресурсов, чтобы помочь продумать этику применения ИИ в своей профессиональной деятельности.

Профессиональная этика

Одним из определяющих свойств профессиональной практики является стремление к соблюдению этических норм. Однако кодексы по этике, публикуемые профессиональными организациями, чаще носят рекомендательный характер [8], они скорее излагают общепризнанные ценности на абстрактном уровне, нежели подробно разъясняют, как вести себя в тех или иных ситуациях. Кроме того, их обновление происходит не столь регулярно, чтобы учесть новые трудности этического характера. Прогрессивные технологии, по-видимому, становятся одним из ключевых источников подобных новых этических дилемм, но спорным является вопрос о том, насколько часто они поднимают принципиально новые вопросы [9].

Результаты исследования, посвященного осведомленности об этике использования радиочастотной идентификации (RFID) в библиотеках, показывают, что при использовании технологий информационные специалисты порой чрезмерно полагаются на производителей в плане обеспечения этичности их применения [10]. Специалистам для понимания последствий использования ИИ в их профессиональной деятельности необходимы и другие материалы, помимо общих кодексов по этике. Международная федерация библиотечных ассоциаций и учреждений (ИФЛА) опубликовала обстоятельный обзор о последствиях применения ИИ для профессиональной этики [11]. В докладе Т. Падиллы (Thomas Padilla) для организации OCLC представлено практическое руководство по развитию ИИ [12]. При этом в арсенале специалистов по-прежнему остаются лакуны в плане средств, позволяющих изучать этические аспекты новых технологий.

Сценарии, иллюстрирующие этические проблемы, могут сыграть важную роль в изложении нравственных дилемм в доступной форме и стать стимулом для осмысления и обсуждения необходимых ответных мер и действий. Сценарии рассказывают истории и служат прекрасным методом освещения непростых тем. Они позволяют передать суть повседневных задач и способствуют эффективному обучению взрослого населения. Существуют наборы сценариев по вопросам этики применительно к работе с информацией в целом (например, [13–15]). Однако ИИ настолько важен, что для него требуется отдельный набор сценариев. Цель данной статьи заключается в том, чтобы объяснить, какие соображения лежат в основе развивающегося набора сценариев по этике ИИ для библиотечных и информационных специали-

тов [16]. Последняя версия сценариев³ доступна по лицензии CC-BY-SA, что дает возможность использовать их в учебном процессе или в конкретных организационных контекстах [17].

Этические аспекты использования искусственного интеллекта

Понимание этики использования ИИ обусловлено его дефиницией. При этом дать определение ИИ довольно сложно, поскольку это комплексная и постоянно эволюционирующая идея. По сути, под ИИ подразумевается выполнение компьютерами действий, аналогичных мыслительной деятельности и принятию решений, осуществляемых человеком. Однако достижение этого потенциала было целью информатики еще с 1950-х годов. С течением времени понимание сущности интеллекта трансформировалось, менялись и технологии, которые, казалось бы, могли предложить некую разновидность интеллекта. Ориентация лишь на сегодняшние технологии отнюдь не значит, что ИИ будет проще понять или правильно определить. Например, в документе Глобального института McKinsey от 2018 г., посвященном необходимости изменения профессиональных навыков в связи с новыми технологиями, был проведен опрос организаций, знакомых как минимум с одной из технологий в области автоматизации, искусственного интеллекта или передовых цифровых технологий и их применения в бизнесе из следующего списка: большие данные и углубленная аналитика, алгоритмы машинного обучения/искусственного интеллекта, автономные транспортные средства, распознавание образов, роботизированная автоматизация процессов, виртуальные помощники, автоматизация офисных процессов, портативные электронные устройства, Интернет вещей (IoT), персонализированное ценообразование и промоакции, 3D-печать, блокчейн и распределенный реестр [18].

В то же время в статье Я.-М. Ловендаля (Jan-Martin Lowendahl) и К.К. Уильямс (Kelly Calhoun Williams) говорится, что возможности ИИ включают в себя роботизированную автоматизацию процессов, компьютерное зрение, машинное обучение, понимание текста на естественном языке, виртуальных помощников или диалоговые интерфейсы, физическую робототехнику, генерацию естественного языка, автономные транспортные средства [19]. Каждая из технологий несет в себе потенциальные этические проблемы. Дополнительная сложность заключается в том, что, поскольку в основе работы многих технологий лежат данные, вопросы этики ИИ затрагивают и многие из предыдущих дебатов об этике больших данных.

ИИ — это нечто значительно большее, чем совокупность технологий, какими бы мощными они ни были. В отличие от многих других техно-

логий, ИИ занял прочное место в общественном сознании благодаря своей популярности в научно-фантастических телепередачах, фильмах и книгах. В средствах массовой информации и других источниках ИИ зачастую подается как антиутопия, а не утопическая возможность. Восприятие ИИ не обязательно должно быть изолировано от реакций, порожденных подобными вымышленными представлениями. Если же говорить более конкретно, то ИИ также сопровождается активным движением, имеющим коммерческую ценность. Взгляд общества на ИИ сформирован компаниями, цель которых — продвигать свои продукты как нечто, способное преобразовать деятельность организаций. Отчасти речь может идти об ажиотаже по поводу функциональности, которая, к слову, не слишком отличается новизной. Можно также считать, что это связано с сильными в культурном плане представлениями о технологическом детерминизме и солюционизме [20].

Подобные идеи преподносят технологические изменения как неотвратимые и как способные беспрепятственно разрешить комплексные социальные проблемы. Они также воплощают в себе сексистские и расистские установки. Вместо того чтобы обусловленные технологиями изменения принимать как данность, обществу следует с учетом этических норм думать о выборе, который могут сделать индивиды, социальные группы и общество. Следует поставить вопрос: позволяют ли технологии решать проблемы, которые требуют решения, принимая во внимание потребности человека. Нужно противостоять ажиотажу по поводу ИИ и власти технологических гигантов, определяющих повестку дня. Кроме того, существуют значительные социальные последствия, связанные с процессами датафикации и слежки за данными, а также проблемы обеспечения устойчивости, вызванные высокими потребностями ИИ в электроэнергии. Эти многочисленные спорные точки зрения свидетельствуют о более серьезных препятствиях, которые существуют в области ИИ.

Необходимо также признать, что существуют этические проблемы в разработке ИИ, управлении им, оценке, а не только в его технических аспектах [21]. Критики полагают, что в настоящее время этические аспекты являются второстепенными в процессе создания ИИ, а заявления крупных технологических компаний об этике зачастую являются не более чем моральным политесом. Складывается ощущение, что в процесс разработки и использования ИИ этическое мышление должно быть внедрено гораздо глубже. Поскольку многие сферы применения ИИ оказывают широкое социальное воздействие, необходимо, чтобы все заинтересованные стороны присутствовали или были представлены на всех стадиях процесса разработки и внедрения. Проектирование ИИ не должно

замыкаться на экспертах. Вероятно, наилучшим способом защиты общества является правовое регулирование, а не упование на то, что компании будут придерживаться этических норм. Таким образом, вопрос, который необходимо задать в отношении ИИ, заключается не только в том, является ли его применение этическим и нейтральным с точки зрения социального воздействия, а в том, способствует ли оно активному продвижению социальной справедливости. Подобные соображения свидетельствуют о сложности определения этических норм ИИ в любом контексте.

Значительная часть дискуссий относительно этики ИИ сосредоточена в основном на следующих взаимосвязанных вопросах: пристрастность, прозрачность, объяснимость и отслеживаемость, конфиденциальность, защита и безопасность, а также влияние на человеческий выбор и свободу [5; 22]. СМИ уделяют особое внимание случаям, когда ИИ демонстрирует свою предвзятость, например, когда система распознавания лиц не распознает людей с темным цветом кожи. В значительной степени проблемы обусловлены предвзятостью исторических данных, которые использовались для обучения алгоритма. В других случаях предвзятость, по-видимому, вызвана кадровым составом индустрии ИИ, который в основном включает преимущественно молодых представителей белой расы, чьи узкие взгляды отражаются в создаваемом ими ИИ. Другой областью, вызывающей особую озабоченность, является прозрачность, объяснимость и ответственность. Если инструменты ИИ учатся принимать решения на основе данных, а не путем, который определил человек-кодировщик, сразу же возникает вопрос о том, как объяснить процесс и его результаты. Принцип работы ИИ не всегда понятен даже разработчику алгоритма, а значит, объяснить его может быть весьма сложно. Затем, если процесс принятия решений ИИ непрозрачен, возникает вопрос об ответственности. Кто несет ответственность, если ИИ совершает ошибку? И как можно добиться полностью информированного согласия?

Конфиденциальность представляет собой еще одну серьезную проблему при применении ИИ, связанную с персонализацией и адаптацией, которые основываются на объединении персональных данных из различных источников. Хранение подобных персональных данных создает угрозу для безопасности и сохранности. ИИ ставит фундаментальные вопросы о человеческой самостоятельности. В чем заключается роль человека, когда решения принимают компьютеры? Одним из примеров того, как ИИ может рассматриваться в качестве угрозы человеческой самостоятельности, является концепция подталкивания (англ. Nudge theory), когда дизайн системы направлен на то, чтобы повлиять на определенную группу людей. Важным аспектом автоматизации является ее влия-

яние на работу, в том числе на профессиональную деятельность. Если ИИ станет более эффективным, он сможет заменить специалистов во многих областях. ИИ могут использовать для контроля над работой человека и превратить приносящую удовлетворение работу в рутину. Аналогичным образом он способен расширить круг задач специалистов, позволяя им сосредоточиться на более значимых и комплексных аспектах своей работы. Влияние на трудовые ресурсы является важным фактором нравственного воздействия ИИ.

В дополнение к основным этическим вызовам ИИ, как уже говорилось, существуют вопросы, связанные непосредственно с профессиональными ценностями в области информации. Так, ИФЛА подробно рассмотрела некоторые из основных проблем обеспечения свободы информации и самовыражения, возникающих в результате использования ИИ [11]. Несмотря на то что им уделяется меньше внимания в широких дискуссиях об этике ИИ, становится очевидным, что специалисты библиотечного и информационного профиля могут внести свой особый вклад в рассмотрение этики ИИ. Персонализация и адаптация к индивидуальным потребностям, которые зачастую являются главным пунктом в заявлениях о возможностях ИИ, также порождают эффект «пузыря фильтров», который фактически ограничивает свободу доступа к информации. В частности, применение ИИ для модерирования форумов может привести к нарушению свободы слова, особенно если он предназначен для излишне осмотрительного блокирования материалов, которые могут оказаться спорными [23]. И если использование ИИ незаметно блокирует доступ к информации, то его применение для создания дипфейков имеет гораздо более серьезные последствия с точки зрения доверия к информации. ИФЛА обратила внимание на то, что ИИ способен нарушить неприкосновенность частной жизни и породить у людей страх перед слежкой за тем, что они ищут и читают, а это может привести к будущим отрицательным последствиям. Таким образом, растущее использование ИИ способствует росту опасений относительно рисков слежки и надзора за данными [23].

Проблемы, связанные с ChatGPT

С момента своего запуска в ноябре 2022 г. ChatGPT компании Open AI⁴ значительно способствовал росту информированности о разработках в области ИИ, но также усилил беспокойство по поводу этических норм в сфере ИИ. Вероятно, причина столь поразительного и впечатляющего эффекта ChatGPT заключается в его способности оперативно выполнять широкий спектр сложных задач, включая написание текста в различных стилях, а также написание кода. ChatGPT пред-

ставляет собой скорее разновидность общего ИИ, нежели узкого ИИ, с которым мы сталкивались ранее. При грамотном использовании ChatGPT и подобные ему технологии несут в себе множество преимуществ для работы с информацией, например позволяют составлять краткие конспекты сложных текстов и помогают облегчить процесс работы с письменной речью. Однако они также служат иллюстрацией множества этических проблем, присущих ИИ в целом, и обсуждение этого предоставляет возможность их конкретизировать.

Проблемы предвзятости, осмысленности и приватности, которые, по всей видимости, являются неотъемлемой частью ИИ, очевидны и в технологии ChatGPT [24]. Поскольку GPT частично обучается на данных из открытых источников, он отражает социальные предубеждения, которые могут воспроизводить гендерные стереотипы о трудовой деятельности, например подразумевать, что врач — это мужчина, и выдает утверждения, которые представляются необъективными по отношению к конкретной религии, например индуизму [25]. Другим источником обучающих данных для ChatGPT является платформа Reddit⁵, пользователям которой присущи свои собственные предубеждения. Если же использовать тренеров людей для борьбы с предвзятостью обучающих данных, можно получить уже новые формы предвзятости [26]. ChatGPT может также показаться малопонятным, поскольку компания отказывается открыто говорить о том, на каких данных он основан и как работает. Что касается защиты конфиденциальности, то он был временно запрещен в Италии в связи с недостаточной степенью защиты личных данных и отсутствием проверки возраста. Опасения по поводу конфиденциальности все еще сохраняются, что вынудило ряд компаний предостеречь сотрудников от использования этого сервиса.

Отметим, что с точки зрения достоверности информации ChatGPT имеет серьезные проблемы. Если говорить о бесплатной версии, то корпус данных, на которых она основана, не обновлялся с сентября 2021 г., что снижает ее надежность и может ввести неопытных пользователей в заблуждение. Более того, одна из наиболее часто обсуждаемых особенностей ChatGPT заключается в том, что он искажает информацию вплоть до выдумывания ссылок для подтверждения собственных утверждений. ChatGPT авторитетно и убедительно излагает многие темы, однако информация, которую он выдает, часто не соответствует действительности. Он не указывает ссылки на свои источники, и их трудно отследить. Повторный поиск дает другой результат, что ограничивает возможность проверки информации. Наличие популярного инструмента, не отвечающего этим основным критериям достоверности информации, может рассматриваться как подрыв доверия к информации в целом. В случае с ИИ, генерирующим

изображения, повышается риск порождения правдоподобных фальшивок и подделок.

ChatGPT способен нанести ущерб авторскому праву, поскольку он использует текст и данные из Интернета без предварительного разрешения. Вполне возможно, что извлечение данных из общедоступных источников само по себе является актом эксплуатации. Компания OpenAI заявляет, что получала доступ к контенту на условиях добросовестного использования, однако если она воспроизвела нечто близкое к ранее опубликованному тексту, на нее могут быть поданы те же иски о нарушении авторских прав, которые сейчас рассматриваются в отношении других продуктов генеративного ИИ [27]. Существуют также и вопросы обеспечения справедливости. С появлением новых, возможно, более эффективных сервисов с платной подпиской цифровое неравенство между теми, кто в состоянии позволить себе такие услуги, и теми, кто не может этого сделать, будет увеличиваться.

Прочие сложности касаются не столько функций ChatGPT или способов его использования, сколько того, как он был разработан. Например, в разоблачительной статье издания Time было выявлено, что просмотром и очисткой проблемного контента занимались кенийские сотрудники, которые получали крайне низкую заработную плату и минимальную поддержку [28]. И хотя компания OpenAI изначально создавалась как идеалистическая некоммерческая организация и ее называют «открытой», прозрачность компании и ее продуктов вызывает сомнения [29]. Кроме того, возникают разногласия по поводу экологической устойчивости GPT-моделей [25; 30]. При этом генеративный ИИ весьма требователен с точки зрения энергопотребления.

На более широком уровне ChatGPT, как представляется, может оказать существенное воздействие на социальную сферу, например, поставить под угрозу источники дохода людей, работающих в таких отраслях, как издательское дело, журналистика и маркетинг, а также в сфере разработки программного обеспечения. Он уже оказывает значительное влияние на подходы учебных заведений к оцениванию и методам обучения письменной речи, и существует немало опасений, что он может негативно повлиять на навыки письменной речи и критического мышления студентов [31]. Создается впечатление, что при запуске данного сервиса подобные последствия не принимались во внимание. Резкое внедрение меняющего парадигму инструмента и при этом полное отсутствие беспокойства о более широких социальных последствиях свидетельствует о могуществе крупных технологических компаний [32]. В ежегодном аналитическом отчете института AINow подчеркивается необходимость обуздать ничем не регулирующую власть технологических гигантов [3]. Опасения, по-видимому, присутствуют и в самой отрасли,

в результате чего лидеры в области ИИ, включая Илона Маска, в 2023 г. призвали к шестимесячному перерыву в развитии ИИ [33]. Они обосновали паузу тем, что ИИ оказывает воздействие, которое никто, даже его создатели, не может понять, предсказать или с уверенностью контролировать [34].

Перед информационными специалистами стоит немало проблем. Если бы последствия применения ИИ для библиотек ограничивались его непосредственным использованием при оказании информационных услуг, то они смогли бы справиться с этической стороной вопроса. Однако в реальности библиотека обязана работать в условиях более широкой информационной среды и использовать информационные продукты, созданные в совершенно иных контекстах. Библиотека также должна помогать пользователям разобраться в том, как ИИ воздействует на их жизнь. Не исключено, что нормативные положения позволят сдерживать наиболее негативные проявления ИИ, но гораздо вероятнее, что библиотечные специалисты станут участвовать в разъяснении границ применения этой технологии и давать пользователям рекомендации относительно получения более качественных ответов с помощью оптимальных запросов.

Сценарии

Исходя из изложенного выше, мы можем сделать вывод, что в сценариях, касающихся этики использования ИИ, необходимо изучить ряд аспектов в рамках библиотечной деятельности.

1. Понимание баланса между потенциальными преимуществами ИИ и связанными с ним этическими проблемами.
2. Спектр областей применения ИИ, актуальных для информационных специалистов, от непосредственного использования в рамках информационных услуг до применения в управлении услугами, а также в случаях, когда информационные специалисты могут оказывать поддержку организациям в развертывании ИИ.
3. Спектр различных контекстов информационной работы, например, здравоохранение, библиотеки, бизнес и правовой сектор.
4. Вопросы, связанные с данными, отражающие зависимость ИИ от больших данных.
5. Указание на целый ряд дилемм этического характера, которые освещаются в литературе, особенно на такие, как пристрастность, прозрачность, объяснимость и ответственность, конфиденциальность, защита и безопасность, а также то, как ИИ влияет на человеческий выбор и свободу.
6. Упоминание о проблемах, возникающих в связи с воздействием на трудовые и профессиональные функции.
7. Упор на проблемы этического порядка, которые особенно важны для информационных спе-

циалистов, например связанные с доступом к информации и свободой выражения мнений.

8. Указание на более глубокие вопросы, относящиеся к участию всех заинтересованных сторон в разработке и использовании ИИ и к его влиянию на социальную справедливость и устойчивость.

Сценарии были опубликованы на условиях лицензии CC-BY-SA, что позволяет специалистам в области библиотечного и информационного дела адаптировать их к местным потребностям, например путем внесения изменений с учетом отраслевой специфики и собственного организационного контекста.

Заключение

В условиях растущего интереса к ИИ как непосредственно в деятельности библиотечных и информационных специалистов, так и в прочих организациях, где они работают, назрела острая необходимость в проведении дискуссий об этике использования этих технологий на практике. В настоящей статье выявлены различные аспекты и проблемы. Сценарии, где затрагиваются этические соображения, позволяют в наглядной форме представить ключевые дилеммы, способствуя обсуждению и повышению уровня понимания ИИ. Предоставив сценарии в свободном доступе, авторы надеются, что они могут быть адаптированы для использования в различных организациях и также в целях преподавания.

Уместно закончить статью об этических нормах вопросами, а не ответами. Документ со сценариями завершается набором общих вопросов, к которым читатель будет подготовлен путем анализа восьми сценариев. Они отражают глубину профессиональных вызовов, которые ставит перед нами ИИ.

- Насколько существующие кодексы профессиональной этики помогают ориентироваться в подобных сценариях?

- Создает ли ИИ принципиально новые проблемы в области этики и ценностей профессии, и если да, то в каком плане?

- Как правило, в заявлениях о профессиональной этике основное внимание уделяется ответственности отдельного специалиста, но с какими этическими вызовами сталкивается организация?

- Требуется ли нам государственное регулирование ИИ вместо того, чтобы полагаться на соблюдение организациями и отдельными людьми этических норм?

Примечания

¹ <https://www.aiaaic.org/home>

² <https://incidentdatabase.ai/>

³ https://orda.shef.ac.uk/articles/dataset/Ethics_Scenarios_Of_Artificial_Intelligence_For_Information_and_Knowledge_Management_And_Library_Professionals_AFTER_FEEDBACK/17081138/1

⁴ <https://openai.com/>

⁵ <https://www.reddit.com/>

Список источников

1. Cordell R. Machine Learning and Libraries : A Report on the State of the Field. Washington, DC : LC Labs Library of Congress, 2020. URL: <https://labs.loc.gov/static/labs/work/reports/Cordell-LOC-ML-report.pdf> (дата обращения: 25.03.2025).
2. The Impact of AI, Machine Learning, Automation and Robotics on the Information Professions : A Report for CILIP. London : CILIP, 2021. URL: <https://www.cilip.org.uk/general/custom.asp?page=researchreport> (дата обращения: 25.03.2025).
3. AI Now 2018 report / M. Whittaker, K. Crawford, R. Dobbe, G. Fried, E. Kaziunas, V. Mathur, S. Myers West, R. Richardson, J. Schultz, O. Schwartz. AINow Institute, 2018. URL: <https://ainowinstitute.org/publications/ai-now-2018-report-2> (дата обращения: 25.03.2025).
4. AI Now 2019 report / K. Crawford, R. Dobbe, T. Dryer, G. Fried, B. Green, E. Kaziunas, A. Kak, V. Mathur, E. McElroy et al. AINow Institute, 2019. URL: <https://ainowinstitute.org/publications/ai-now-2019-report-2> (дата обращения: 25.03.2025).
5. Jobin A., Ienca M., Vayena E. Artificial Intelligence: The Global Landscape of AI Ethics Guidelines // Nature Machine Intelligence. 2019. Vol. 1, № 9. P. 389–399. DOI: 10.1038/s42256-019-0088-2.
6. Developing The Ethical Framework for AI in Education : Annex to The Ethical Framework for AI in Education. Institute for Ethical AI in Education, 2021. 15 p. URL: <https://www.buckingham.ac.uk/wp-content/uploads/2021/03/The-Institute-for-Ethical-AI-in-Education-Annex-Developing-the-Ethical-Framework-for-AI-in-Education-IEAIED-.pdf> (дата обращения: 25.03.2025).
7. A Pathway Towards Responsible, Ethical AI // Jisc : Repository. 2021. URL: <https://repository.jisc.ac.uk/8548/> (дата обращения: 25.03.2025).
8. Frankel M.S. Professional Codes: Why, How, and With What Impact? // Journal of Business Ethics. 1989. Vol. 8. P. 109–115. DOI: 10.1007/BF00382575.
9. Ferguson S., Thornley C., Gibb F. Beyond Codes of Ethics: How Library and Information Professionals Navigate Ethical Dilemmas in a Complex and Dynamic Information Environment // International Journal of Information Management. 2016. Vol. 36, № 4. P. 543–556. DOI: 10.1016/j.ijinfomgt.2016.02.012.
10. Ferguson S., Thornley C., Gibb F. How Do Libraries Manage the Ethical and Privacy Issues of RFID Implementation? A Qualitative Investigation into the Decision-Making Processes of Ten Libraries // Journal of Librarianship and Information Science. 2015. Vol. 47, № 2. P. 117–130. DOI: 10.1177/0961000613518572.
11. IFLA Statement on Libraries and Artificial Intelligence. International Federation of Library Associations and Institutions, 2020. 14 p. URL: https://www.ifla.org/wp-content/uploads/2019/05/assets/faife/ifla_

- statement_on_libraries_and_artificial_intelligence.pdf (дата обращения: 25.03.2025).
12. *Padilla T.* Responsible Operations: Data Science, Machine Learning, and AI in Libraries. Dublin, OH: OCLC Online Computer Library Center, Inc. 2019.
 13. *Buchanan E.A., Henderson K.A.* Case Studies in Library and Information Science Ethics. London : McFarland, 2009. 175 p.
 14. *McMenemy D., Poulter A., Burton P.* A Handbook of Ethical Practice : A Practical Guide to Dealing with Ethical Issues in Information and Library Work. London : Chandos, 2014. 172 p.
 15. *Rösch H.* Edili: Ethical Dilemmas for Librarians and Other Information Workers: Case Studies // Technische Hochschule Köln. URL: <https://www.iws.th-koeln.de/edili/edili-start.php> (дата обращения: 25.03.2025).
 16. *Cox A.* The Ethics of AI for Information Professionals: Eight Scenarios // Journal of the Australian Library and Information Association. 2022. Vol. 71, № 3. P. 201–214. DOI: 10.1080/24750158.2022.2084885.
 17. *Cox A.* Ethics Scenarios of Artificial Intelligence for Information and Knowledge Management and Library Professionals // University of Sheffield. Dataset. 2021. DOI: 10.15131/shef.data.17081138.v1.
 18. *Bughin J., Hazan E., Lund S., Dahlström P., Wiesinger A., Subramaniam A.* Skill Shift: Automation and the Future of the Workforce. McKinsey Global Institute. Discussion Paper. 23 May 2018. New York : McKinsey & Co. URL: <https://www.mckinsey.com/featured-insights/future-of-work/skill-shift-automation-and-the-future-of-the-workforce> (дата обращения: 25.03.2025).
 19. *Lowendahl J.-M., Calhoun Williams K.* 5 Best Practices for Artificial Intelligence in Higher Education // Gartner group. 20 December 2018. URL: <https://www.gartner.com/en/documents/3895923> (дата обращения: 25.03.2025).
 20. *Mirza R., Seale M.* Who Killed the World? White Masculinity and the Technocratic Library of the Future // Topographies of Whiteness : Mapping Whiteness in Library and Information Science / ed. by G. Schlesselman-Tarango. Sacramento, CA : Library Juice Press, 2017. P. 171–197.
 21. *Greene D., Hoffmann A.L., Stark L.* Better, Nicer, Clearer, Fairer: A Critical Assessment of the Movement for Ethical Artificial Intelligence and Machine Learning // Proceedings of the 52nd Hawaii International Conference on System Sciences, 2019. URL: <https://scholarspace.manoa.hawaii.edu/server/api/core/bitstreams/849782a6-06bf-4ce8-9144-a93de4455d1c/content> (дата обращения: 25.03.2025).
 22. *Fjeld J., Achten N., Hilligoss H., Nagy A., Srikumar M.* Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI. Berkman Klein Center for Internet & Society Research Publication, 2020. 71 p. URL: https://www.researchgate.net/publication/339138141_Principled_Artificial_Intelligence_Mapping_Consensus_in_Ethical_and_Rights-Based_Approaches_to_Principles_for_AI (дата обращения: 25.03.2025).
 23. Privacy and Freedom of Expression in the Age of Artificial Intelligence Privacy International : Report // Privacy International. 11 May 2018. URL: <https://privacyinternational.org/report/1752/privacy-and-freedom-expression-age-artificial-intelligence> (дата обращения: 25.03.2025).
 24. ChatGPT // AIAAIC Repository of AI, Algorithmic and Automation Incidents and Controversies. Published December 2022. Last updated February 2024. URL: <https://www.aiaaic.org/aiaaic-repository/ai-algorithmic-and-automation-incidents/chatgpt-chatbot> (дата обращения: 25.03.2025).
 25. *Burruss M.* The (Un)ethical Story of GPT-3: OpenAI's Million Dollar Model // Matthew Burruss Blog. July 27 2020. URL: <https://matthewpburruss.com/post/the-unethical-story-of-gpt-3-openais-million-dollar-model/> (дата обращения: 25.03.2025).
 26. *Webb M.* Exploring the Potential for Bias in Chat GPT // Jisc : National Centre for AI. 26 January 2023. URL: <https://nationalcentreforai.jiscinvolve.org/wp/2023/01/26/exploring-the-potential-for-bias-in-chatgpt/> (дата обращения: 25.03.2025).
 27. *Dreben R.N., Julyan M.T.* Generative Artificial Intelligence and Copyright: Current Issues // Morgan, Lewis & Bockius LLP. 23 March 2023. URL: <https://www.morganlewis.com/pubs/2023/03/generative-artificial-intelligence-and-copyright-current-issues> (дата обращения: 25.03.2025).
 28. *Perrigo B.* Exclusive: OpenAI Used Kenyan Workers on Less Than \$2 Per Hour to Make ChatGPT Less Toxic // Time. 18 January 2023. URL: <https://time.com/6247678/openai-chatgpt-kenya-workers/> (дата обращения: 25.03.2025).
 29. *Hao K.* The Messy, Secretive Reality Behind OpenAI's Bid to Save the World // MIT Technology Review. 17 February 2020. URL: <https://www.technologyreview.com/2020/02/17/844721/ai-openai-moonshot-elon-musk-sam-altman-greg-brockman-messy-secretive-reality/> (дата обращения: 25.03.2025).
 30. *Groes Albin Ludvigsen K.* The Carbon Footprint of Chat GPT // Medium. 22 December 2022. Updated 20 March 2023. URL: <https://towardsdatascience.com/the-carbon-footprint-of-chatgpt-66932314627d> (дата обращения: 25.03.2025).
 31. *Cotton D.R.E., Cotton P.A., Shipway J.R.* Chatting and Cheating: Ensuring Academic Integrity in the Era of ChatGPT // Innovations in Education and Teaching International. 2024. Vol. 61, № 2. P. 228–239. URL: DOI: 10.1080/14703297.2023.2190148.
 32. *Telving T.* With ChatGPT, We All Act As Well-Behaved Children of Tech Capitalism // DataEthics. News. 23 January 2023. URL: <https://dataethics.eu/with-chatgpt-we-all-act-as-well-behaved-children-of-tech-capitalism/> (дата обращения: 25.03.2025).
 33. *Hern A.* Elon Musk Joins Call for Pause in Creation of Giant AI “Digital Minds” // The Guardian. 29 March 2023. URL: <https://www.theguardian.com/technology/2023/mar/29/elon-musk-joins-call-for-pause-in-creation-of-giant-ai-digital-minds> (дата обращения: 25.03.2025).
 34. Pause Giant AI Experiments: An Open Letter. Future of Life Institute. 22 March 2023. URL: <https://futureoflife.org/open-letter/pause-giant-ai-experiments/> (дата обращения: 25.03.2025).

Перевод **Марии Федотовой**,
Российская государственная библиотека