

Культурные вызовы практической этике в сфере искусственного интеллекта: проблемы и решения

С. С. Кладько

Аннотация— Динамичное распространение технологий искусственного интеллекта (ИИ) усиливает необходимость в этических рамках, которые не только теоретически обоснованы, но и практически реализуемы. Практическая этика ИИ становится ключевой областью, заполняя разрыв между абстрактными этическими принципами и стратегиями, применимыми на практике. Данная статья исследует концепцию практической этики ИИ, акцентируя внимание на её роли в процессах принятия решений в реальных условиях и применении в различных отраслях. В отличие от традиционных этических теорий, практическая этика ИИ сосредоточена на операционализации таких принципов, как справедливость, прозрачность и подотчетность, с учетом технологических и социальных контекстов.

Ключевым аспектом практической этики ИИ является её взаимодействие с культурным разнообразием, фактором, который часто игнорируется в основных обсуждениях. Культурные нормы и ценности существенно влияют на то, как системы ИИ разрабатываются, внедряются и воспринимаются в разных обществах. В статье рассматривается пересечение культурных вопросов и этики ИИ, выявляются как вызовы, так и возможности. Через анализ существующих этических рамок, таких как утилитаризм, деонтология и этика добродетели, и их актуальности для ИИ, обсуждение расширяется до включения культурно инклюзивных практик. Примеры из здравоохранения, образования и управления демонстрируют, как культурно ориентированные системы ИИ могут повысить уровень доверия и эффективности, одновременно минимизируя предвзятость. Также в исследовании предлагается набор практических рекомендаций для учёта культурного компонента в вариантах практической реализации этики в сфере ИИ, что направлено на формирование более целостного подхода к управлению ИИ.

Ключевые слова— искусственный интеллект, культура, кросс-культурный менеджмент, этика в сфере искусственного интеллекта.

I. ВВЕДЕНИЕ

Динамичный рост технологий искусственного интеллекта (ИИ), постоянно сталкиваясь с вопросами, касающимися конфиденциальности, справедливости, безопасности, и ответственности, усиливает и актуальность исследования этих этических аспектов, ставших неотъемлемой составляющей дискурса об ИИ

по мере того, как технологии все глубже внедряются в социальные, экономические и политические процессы. Действительно, ускоренное развитие ИИ ведёт к сложным вызовам, требующим критического и оперативного рассмотрения этических последствий в процессе его внедрения [1], [2]. В частности, с увеличением объёмов данных, используемых для обучения ИИ, существенно возрастают риски дискриминации и предвзятости алгоритмов, использующихся в правосудии, здравоохранении и трудоустройстве (что может негативно сказываться на уязвимых социальных группах и усиливать социальное неравенство)[3], а также риски случайного или намеренного раскрытия личной информации (что усиливает важность создания этических стандартов для защиты данных пользователей) [4]. Возрастающая сложность распределения ответственности между стейкхолдерами (под которыми в данной статье понимаются, прежде всего, разработчики и владельцы и пользователи ИИ-решений, а также прочие акторы, поименованные в Кодексе этики в сфере искусственного интеллекта РФ) требует разработки более чётких стандартов и политик ответственности [2] и необходимости установления мер безопасности для предотвращения случайного или злонамеренного использования этих технологий [5].

Предполагается, что всё это должно усилить ответственное отношение к этике в сфере ИИ и, как следствие, существенно снизить потенциальные негативные последствия применения решений на основе ИИ и, тем самым, повысить доверие общества к наиболее динамично развивающимся технологиям. Однако немалое препятствие в продвижении практической этики представляет сама этика, вернее сохраняющееся отношение к ней как к некоему теоретическому конструкту для академических исследований. Следует согласиться с тем, что в рассуждениях о преобладающем подходе (деонтологии, утилитаризме или этике добродетели) теряется осознание того, что этика в сфере ИИ направлена, прежде всего, на реальное взаимодействие связанных с ней стейкхолдеров, которые зачастую крайне недовольны навязыванием им абстрактных конструктов, способных оказать негативное влияние на их конкурентоспособность [6]; как следствие, существенный рост за последние несколько лет исследовательских работ по этике в сфере ИИ [7] мало способствует эффективности её внедрения.

Статья получена 20 ноября 2024.

Статья подготовлена по итогам выступления на конференции «Интернет и современное общество», IMS-2023.

С.С. Кладько, АНО НЕЙМАРК (e-mail: s.kladko@neimark-it.ru).

В этой связи представляется важным постараться перейти от теоретической этики в сфере ИИ к практической, направленной на непосредственное применение этических постулатов для решения конкретных вопросов и дилемм, возникающих непосредственно в процессе проектирования, разработки и использования ИИ-систем. Задачи практической этики в сфере ИИ заключаются в разработке ориентиров, помогающих стейкхолдерам принимать прикладные решения в условиях неопределенности и быстрого технологического прогресса и, тем самым, избегать возможного ущерба (как репутационного, так и финансового) и даже, возможно, добавлять реальной стоимости своим разработкам. Именно такой подход позволяет преодолеть разрыв между теоретическими представлениями об этике и реальными проблемами, возникающими при разработке и внедрении ИИ.

В рамках данной статьи хотелось бы уделить особое внимание фактору, который, являясь критически важным, нередко упускается из виду в разработке и применении этических норм в ИИ: культура как совокупность убеждений, норм, ценностей и поведенческих моделей, присущих тем, кто разрабатывает, владеет или пользуется решениями на основе ИИ. «Невидимость» культуры зачастую создаёт впечатление того, что её можно частично или полностью проигнорировать в практической деятельности, тем более такой технологичной как ИИ, однако именно культурные особенности напрямую влияют на субъективную трактовку данных при работе с корпусами, принятие решений на основе культурного контекста, а также определяет приоритеты в иерархии этических принципов в различных целевых аудиториях. Более того, при всей незаметности культуры, уже сегодня отмечается её возрастающее влияние на возникновение конфликтных ситуаций в этике в сфере ИИ [8]. Однако, несмотря на постоянно растущий корпус практических кейсов, подтверждающих свою важность, культурный фактор, к большому сожалению, продолжает оставаться для этики в сфере ИИ если и не игнорируемым полностью, то, в лучшем случае, на второстепенных ролях.

Настоящая статья направлена на рассмотрение концепции практической этики в сфере ИИ в рамках ее операционализации для облегчения взаимодействия между стейкхолдерами, а также на исследование и формулирование прикладных подходов к этике в сфере ИИ, уделяя особое внимание культурному фактору.

Соответственно, главными вопросами исследования являются следующие:

1. Какие ключевые барьеры мешают эффективному применению этических принципов ИИ на практике, и как их можно преодолеть?
2. Какую роль играют культурные различия в восприятии и практическом внедрении этических стандартов ИИ?

II. РАЗРАБОТАННОСТЬ ПРОБЛЕМАТИКИ ИССЛЕДОВАНИЯ

Комплексная природа практической этики в сфере ИИ

как подхода, ориентированного на прикладное применение этико-философских подходов и принципов, воплощает в себе дуализм общих теоретических взглядов на то, что представляет из себя практическая этика вообще, а также этических принципов, действующих именно в сфере ИИ. В частности, в философии практическая этика как направление традиционно рассматривается с позиций утилитаризма, деонтологии и этики добродетели. Утилитаризм, основанный на идее максимизации пользы, в контексте ИИ (например, в области машинного обучения) фокусируется на оценке его воздействия на общественное благосостояние [9] и, теоретически, требует от разработчиков ИИ учитывать не только экономическую, но и социальную выгоду. Деонтологический подход основан на идее соблюдения моральных обязательств и принципов, независимо от возможных последствий, требуя в рамках ИИ соблюдения определенных этических принципов, таких как конфиденциальность, честность и уважение к личности, даже если это не всегда приносит наибольшую выгоду (Floridi et al [10] строят свои принципы для ИИ именно на деонтологии). Наконец, этика добродетели фокусируется на развитии морального характера личности, подчёркивая в контексте проектирования ИИ-решений систем, которые способствуют развитию добродетелей, таких как честность, забота и сострадание [11]. Применение того или иного подхода к решению многочисленных этических дилемм в рамках разработки и применения ИИ-решений имеют свои особенности, а ожидания от их практической реализации могут иметь характер конфликта (более подробная информация об это в части III).

Истоки этики ИИ восходят к общим принципам прикладной этики в отношении технологий, включая идеи, сформулированные Wiener'ом [12], подчеркивавшим важность осознания этических последствий автоматизации и кибернетики. В 1970-е годы Weizenbaum придал новую актуальность данной проблеме, предостерегая от некритического использования машин и моральных рисках для решения человеческих проблем [13]. К концу XX века исследователи стали не просто ещё более акцентировано поднимать вопрос этической ответственности разработчиков перед пользователями и обществом в целом, но и задумываться о том, как запрограммировать машины так, чтобы они придерживались норм морали [14].

Следующим важным шагом в переходе от общих моральных принципов к конкретным прикладным решениям стали попытки формализации принципов этики ИИ [15], в итоге постепенно ставшими основными объектами для практической этики ИИ, что особенно ярко проявилось в 2010х годах, когда IT-индустрия постепенно начала претворять эти принципы в индустриальные этические стандарты, инструкции и кодексы [16],[17]. Параллельно с этим исследователи и подключившиеся к ним IT-компании начинают уделять повышенное внимание алгоритмическим предвзятостям

как одними из основных направлений приложения наработок практической цифровой этики [18][19],[20], что можно считать в определённой степени работающим вариантом перехода этики в сфере ИИ от теории к практике. В то же время проявилась и обратная тенденция (от конкретного к общему), выразившаяся в выработке системных подходов идеи учёта экосистемы цифрового взаимодействия в целом [21] или разработка системы принципов, способных минимизировать потенциальные риски катастрофического масштаба, исходящие из ИИ [22]. Однако несмотря на постоянно растущее число предложений и метрик, практическая этика в сфере ИИ остаётся пока далека от внимания стейкхолдеров как инструмент, способный быть действительно важным для создаваемых ими решений и продуктов.

III. ДУАЛИЗМ ПРОТИВОРЕЧИЙ ПРАКТИЧЕСКОЙ ЭТИКИ В СФЕРЕ ИИ

В контексте ИИ практическая этика направлена на то, чтобы адаптировать и внедрять этические стандарты и нормы в процессы разработки, внедрения и использования технологий ИИ, охватывая такие важные вопросы, как конфиденциальность, справедливость, подотчетность, прозрачность и объяснимость алгоритмов. Казалось бы, уже одно это делает её особенно актуальной в условиях стремительного развития ИИ и возрастающих социальных, экономических и культурных последствий его внедрения. Однако описанная выше комплексная природа практической этики в сфере ИИ также остаётся подверженной вызовам и нерешённым вопросам как общей практической этики, так и этики в сфере ИИ в целом. Это касается, прежде всего, применимости постулатов различных моральных теорий, которые, получив поле для практической реализации, зачастую вступают в противоречие друг с другом. В этом плане традиционно большая часть критик достаётся утилитаризму, постулирующему максимизацию пользы, даже если это требует уступок в плане индивидуальных прав или конфиденциальности (например, поддержка использования технологии распознавания лиц в общественных местах, если это значительно улучшает безопасность, несмотря на нарушение конфиденциальности). Всё это противоречит идеям деонтологии (максимизация пользы не оправдывает игнорирование правил или прав) и этики добродетели (моральные качества создателей или пользователей первичны). Впрочем, здесь можно наблюдать и своеобразный «конфликт» между этикой добродетели и деонтологией в случае, если добродетельные действия не соответствуют установленным правилам (например, приоритет сострадания в здравоохранении на основе ИИ может потребовать нарушения строгих норм конфиденциальности ради улучшения состояния пациентов). При этом, при всем кажущейся теоретизации подобных философских рассуждений, противоречия в данных подходах оказывают влияние на процессы прикладного применения этики в сфере ИИ. Следует признать, что утилитаристский подход является

наиболее понятным (и, в принципе, наиболее удобным) вариантом этики для многих стейкхолдеров при разработке или эксплуатации соответствующих технологических решений, так как он во многом созвучен бизнес-задачам, которые они ставят перед собой. В то же время, большинство документов, декларирующих приверженность бизнеса этике в сфере ИИ (кодексы, инструкции и т.д.) написаны явно с точки зрения деонтологии, а наши ожидания высоких моральных убеждений у тех же разработчиков уже ближе к этике добродетели. При этом любая попытка как-то примирить тот же самый утилитаризм и деонтологию в рамках этики в сфере ИИ ведёт к возникновению ещё больших противоречий [23].

Процесс перехода от теории к практике этики в сфере ИИ, в свою очередь, оказывается под влиянием вызовов от стейкхолдеров для соответствующих ИИ-решений. К ним, в частности, можно отнести следующие:

1. Сложность интерпретации этических принципов. Этические нормы часто формулируются в абстрактной форме, что делает их сложными для практического применения. Представленные высокоуровневые принципы требуют конкретизации, чтобы их можно было адаптировать к прикладным ситуациям, но процесс такой конкретизации может варьироваться и приводить к различиям в трактовке [6], [24].

2. Отсутствие единых стандартов и нормативных документов, что приводит к разночтениям и непоследовательности в применении этических норм, а также к риску создания решений с недостаточной этической проработкой. Например, в разных юрисдикциях требования к конфиденциальности данных могут отличаться, что приводит к сложности интеграции единого подхода [1],[6].

3. Сложность измерения и верификации этических критериев, что усложняет их верификацию и последующую оценку. Например, определить уровень «справедливости» алгоритма крайне сложно, так как требуется учитывать множество факторов, влияющих на его работу [24].

4. Ограниченность в ресурсах и затраты на этическую оценку. Этические проверки и аудит требуют значительных ресурсов, в том числе временных и финансовых. Компании, особенно стартапы или малые и средние предприятия, могут не располагать достаточными ресурсами для проведения глубоких этических проверок, что часто приводит к компромиссам в области этики ради сокращения расходов и сроков разработки [25].

5. Проблемы с прозрачностью и объяснимостью. Большинство современных алгоритмов, таких как нейронные сети, сложны для интерпретации, что затрудняет их прозрачность и объяснимость. Для этичного применения ИИ важно, чтобы алгоритмы были понятны и объяснимы для конечных пользователей и регуляторов. Однако многие модели остаются «чёрными ящиками», что вызывает вопросы о их справедливости и надёжности и снижает соответствующее доверие к ним со стороны пользователей [26].

6. Конфликт интересов и коммерческая тайна.

Организации, разрабатывающие ИИ, часто находятся под давлением инвесторов и руководства, требующих ускоренного внедрения и монетизации решений. Это приводит к конфликту интересов между необходимостью соблюдения этических норм и задачами бизнеса, что нередко выливается в недостаточное внимание к этическим аспектам. Более того, коммерческая тайна может препятствовать прозрачности и объяснимости алгоритмов, что затрудняет проведение этических проверок.

Ко всему прочему, излишнее увлечение стандартизацией результатов в итоге в огромное количество декларируемых этических принципов [27], а интерес IT-индустрии к этому вопросу—в сторону некоего камуфляжа [25], осознанного создания стейкхолдерами впечатления у различных аудиторий о якобы следовании этике в своих разработках и применении соответствующих решений. В результате, этика в сфере ИИ, непреднамеренно для себя, стала удобным щитом для компаний, не желающих более строгого юридического регулирования своей деятельности [28] и прикрывающихся красивыми по звучанию, но во многом пустыми по реальному наполнению терминами «справедливость», «подотчётность», «прозрачность» и т.д. Сегодня (на конец 2024) этика в сфере ИИ получает новую критику в своей нереалистичности [29] и, что самое страшное, в своей ненужности [24]. Появление и активное развитие генеративного искусственного интеллекта вынудило правительства многих стран к более активному внедрению в сферу ИИ норм жесткого законодательства, которое, несмотря на протесты со стороны ведущих компаний, медленно, но устанавливает правовые рамки, тесня, тем самым, так долго декларируемую идею о том, что компании, осознавая этические рамки, будут осуществлять на основе последних саморегулирование. Подобная ситуация после нескольких десятилетий наработки базы и методик этики в сфере ИИ позволяет предположить в ней если не кризис, то вполне полноценное предкризисное состояние.

IV. КУЛЬТУРНЫЕ ВЫЗОВЫ ПРАКТИЧЕСКОЙ ЭТИКЕ В СФЕРЕ ИИ

Фактор культуры, понимаемый для IT-индустрии не только как совокупность нормативно-ценностных убеждений, присущих государству или обществу, но и как особенности субкультуры профессиональных сообществ, таких как разработчики, исследователи, регуляторы и пользователи технологий, при всей своей кажущейся незаметности, создаёт немало дополнительных препятствий, способствующих замедлению внедрения принципов этики в сфере ИИ в практику. Вряд ли это может считаться неожиданным, исходя из того, что, вне зависимости от всех попыток быть максимально этичными в трактовке данных, интерпретация последних напрямую зависит от социокультурных установок или стереотипов, носителем которых является тот или иной стейкхолдер, объективно неспособный абстрагироваться от них от

них в процессе работы с данными [30]. Культура формирует чрезвычайно стойкую индивидуальную систему убеждений, которые невозможно отделить от профессиональной деятельности стейкхолдеров, определяя их отношение к этике и влияя, тем самым, на интерпретацию этических принципов. Например, разработчики в странах с сильной коллективистской культурой могут проявлять большее внимание к вопросам социальной справедливости и коллективного блага, тогда как в странах с преобладанием индивидуализма большее значение могут придавать вопросам конфиденциальности и индивидуальных прав.

Различная комбинация этических приоритетов в разных культурах может влиять на проектирование ИИ-систем (в некоторых культурах основной акцент может делаться на прозрачности и ответственности, тогда как в других — на эффективности и инновациях, что приводит к различным подходам в разработке ИИ). Вышеупомянутая принадлежность к соответствующей профессиональной культуре также влияет на то, как именно будет реализован на практике тот или иной этический принцип (культура стартапов в Кремниевой долине часто характеризуется на преодоление любых барьеров на пути проекта, что не всегда согласуется с этическими подходами, принятыми в других секторах, таких как медицина или государственное управление). Наконец, культурные различия также играют важную роль в восприятии философских подходов к этике ИИ (утилитаризм может находить большую поддержку в западных культурах, в то время как восточные общества могут придавать больше значения коллективным обязанностям и деонтологическим принципам) [1].

Проблема культурной интерпретации напрямую касается и системы этических принципов, которые, казалось бы, называются одинаково в кодексах и инструкциях компаний в разных частях света: (подотчётность, справедливость, прозрачность, конфиденциальность и т.д.) но зачастую имеют разное смысловое наполнение в разных культурах, означая «разное в разных местах» [31]. И действительно, «прозрачность» в одних сообществах предполагает детальные технические пояснения, а в других это не вызовет доверия, «справедливость» может отличаться в коллективистских и индивидуалистских культурах, а «privacy» для ИИ-решений (в частности, разработки для социального скоринга) даже в пограничных культурах (например, Китая и Индии) может иметь абсолютно разную ценность [32]. Игнорирование подобных, казалось бы, очевидных фактов компаниями и исследователями при попытках не критического буквального следования этическим установкам, присущим западной культуре, всё чаще и чаще получает от других стейкхолдеров соответствующее сопротивление, основанное на их собственных культурных ценностях [33].

V. ВОЗМОЖНЫЕ РЕШЕНИЯ ДЛЯ КУЛЬТУРНЫХ ВЫЗОВОВ

В условиях стремительного внедрения систем ИИ возрастает необходимость учитывать в рамках практической этики в сфере ИИ разнообразие

культурных контекстов с целью обеспечения решениям на основе ИИ дополнительных преимуществ и соответствия ожиданиям других стейкхолдеров. В этом отношении стандартными, но сохраняющими свою значимость выглядят рекомендации по регулярному аудиту наборов данных с целью минимизации культурных предвзятостей, пилотному тестированию систем ИИ на культурно репрезентативных группах пользователей, а также созданию механизмов обратной связи в отношении культурного восприятия проектных и внедренческих решений. Конечно, не теряет свою актуальность и традиционный призыв к междисциплинарному сотрудничеству (хотя текущая «затеоретизированность» этики в сфере ИИ поставила достаточно обоснованные препятствия в восприятии идей академического сообщества у практиков ИИ-индустрии [23]).

Однако более существенным в плане работы с культурными вызовами могла бы стать разработка культурно релевантных метрик для продуктов ИИ, что является не только этической необходимостью, но и соблюдением практических ожиданий того, что системы ИИ «уважают» и отражают разнообразные культурные ценности, способствуя повышению доверия к решениям на основе ИИ.

Как представляется, важную роль в этом процессе могло бы сыграть перенимание опыта кросс-культурного менеджмента, некоторые положения которого (в частности, применение культурных измерений) уже находят свою практическую реализацию в этическом тестировании ИИ-продуктов [34]. Важным условием при этом является избежание излишнего усложнения методик этического тестирования (аудитов) (не следует забывать о том, что одним из основных препятствий для эффективного внедрения практической этики в сфере ИИ является опасение реального бизнеса в том, что подобные методики лишь создадут дополнительное необоснованное препятствие оперативному внедрению соответствующего ИИ-решения на рынок). В этом случае представляется допустимым проводить экспресс-анализ на основании 2-3 культурных измерений, информацию о которых можно быстро найти в интернете (более подробно такой подход изложен в [8]).

VI. ОГРАНИЧЕНИЯ

Представленные в данной статье рекомендации имеют ряд ограничений, задающих направления для дальнейших исследований практической этики в области ИИ и её взаимосвязи с культурным контекстом.

Прежде всего, одним из значительных ограничений является начальная стадия интеграции культурных аспектов в практическую этику ИИ. Многие существующие этические рамки, хотя и теоретически обоснованы, недостаточно адаптированы к локальным культурным нормам и ценностям из-за их ориентации на универсальные принципы. Будущие исследования должны быть направлены на уточнение этих рамок путем проведения межкультурных эмпирических исследований, которые отразят разнообразные

этические перспективы и практики. Это позволит создать более тонкое («выверенное») понимание этики в сфере ИИ.

Еще одной проблемой является ограниченная доступность данных, представляющих культурное разнообразие. Системы ИИ в значительной степени зависят от данных для обучения и принятия решений, однако часто используемые наборы данных не учитывают специфики различных культурных контекстов. Этот дефицит данных может привести к предвзятости и искажениям, подрывающим этическое применение ИИ. Именно для решения этих вопросов необходимо более активное участие во внедрении практической этики в сфере ИИ специалистов по культурной антропологии, социологии и кросс-культурному менеджменту.

Техническая сложность внедрения культурных метрик в системы ИИ также представляет собой ограничение. Многие модели ИИ не обладают способностью динамически интерпретировать и учитывать культурно специфические ценности и социальные нормы. Для улучшения понимания и уважения культурных нюансов со стороны ИИ необходимы достижения в области объяснимого ИИ, обработки естественного языка и алгоритмов машинного обучения. Эти технологические инновации должны сопровождаться междисциплинарными подходами, включая участие представителей различных культурных групп в процессах разработки и оценки.

Наконец, данная статья в основном ориентирована на концептуальный анализ с ограниченным обсуждением практических примеров. Будущим исследованиям в сфере влияния культурных факторов на практическую этику в сфере ИИ следует быть более сфокусированными на практическом (посредством сравнительных исследований реальных кейсов в различных регионах внедрения продуктов на основе ИИ) внедрении этики в сфере ИИ с учетом культурной адаптации.

Решение указанных ограничений позволит ученым и практикам быть уверенными в том, что этика в сфере ИИ будет развиваться в направлении, одновременно учитывающем глобальную значимость этических принципов и культурное уважение к стейкхолдерам.

VII. ЗАКЛЮЧЕНИЕ

Исследование практической этики в области ИИ, особенно в контексте культурного разнообразия, подчеркивает необходимость разработки гибкого и адаптируемого подхода к этическому регулированию в сфере ИИ. Сегодня практическая этика постепенно выходит за рамки теоретических рассуждений, фокусируясь на внедрении этических принципов в реальных условиях, что дополнительно позволяет решениям на основе ИИ функционировать наиболее ответственно в разных культурных контекстах. Предлагаемое исследование акцентирует внимание на ключевой роли культурного фактора в формировании этики в сфере ИИ, подчеркивая, что этические рамки не могут быть эффективно обозначены без учета

культурных ценностей, норм и ожиданий различных целевых аудиторий.

Культурное разнообразие создает как возможности, так и вызовы для внедрения таких этических принципов, как справедливость, ответственность и прозрачность в системы ИИ. Несмотря на то, что глобальные этические стандарты предоставляют основу для формирования политики, их применение часто сталкивается с трудностями в культурных контекстах из-за различий в интерпретации этих принципов. Как показано в статье, гибридный подход, сочетающий глобальные стандарты с учетом культурных особенностей, представляет собой перспективный путь для преодоления этих противоречий путём разработки культурно-ориентированных метрик и методологий. Кроме того, в обсуждении подчеркивается важность междисциплинарного сотрудничества для продвижения практической этики ИИ. Интегрируя идеи из кросс-культурного менеджмента, социологии, IT-индустрии и культурной антропологии, исследователи и практики могут создавать такие системы ИИ, которые не только соответствуют техническим и функциональным целям, но и уважают культурные ценности, способствуя этическому единству.

В заключение, по мере того как ИИ продолжает интенсивно проникать во все сферы человеческой жизни, решение вопросов культурных аспектов в этическом регулировании становится не просто опциональным, но необходимым. На основании этого, будущим исследованиям следует углубиться в эмпирические примеры, кросс-культурную валидацию этических метрик и создание динамичных рамок, способных адаптироваться к сложной реальности культурных взаимодействий. Только через такие усилия этика ИИ может перейти от теоретической идеи к практической реальности (или другими словами, перейти от «что» к «как» [35]), которая будет соответствовать ожиданиям разных культур и, тем самым, стать действительно важной для помощи эффективному развитию ИИ-решений на благо стейкхолдеров и общества в целом.

БИБЛИОГРАФИЯ

- [1] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines." *Nature Machine Intelligence*, 1(9), 2019, pp.389–399.
- [2] C. Cath, "Governing artificial intelligence: ethical, legal and technical opportunities and challenges," *Phil. Trans. R. Soc. A*.37620180080, 2018, <http://doi.org/10.1098/rsta.2018.0080>.
- [3] S. Noble, "Algorithms of Oppression: How Search Engines Reinforce Racism." *Science* 374, 542–542, 2021, DOI:10.1126/science.abm5861
- [4] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, "The ethics of algorithms: Mapping the debate." *Big Data & Society*, 3(2), 2016, <https://doi.org/10.1177/2053951716679679>.
- [5] M. Brundage, B. Garfinkel, S. Avin, J. Clark, and H. Toner, "The malicious use of artificial intelligence: Forecasting, prevention, and mitigation," 2018, <https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf>.
- [6] DataRobot's State of AI Bias Report Reveals 81% of Technology Leaders Want Government Regulation of AI Bias (Report). DataRobot, 2022. URL: <https://www.datarobot.com/newsroom/press/datarobots-state-of-ai-bias-report-reveals-81-of-technology-leaders-want-government-regulation-of-ai-bias/> (дата обращения 30.09.2024).
- [7] Artificial Intelligence Index Report. Stanford Institute for Human-Centered Artificial Intelligence, April 2024 <https://aiindex.stanford.edu/report/> (дата обращения 11.10.2024).
- [8] S. Kladko, "Cultural-Ethical Evaluation in the Launch of AI Education Technologies," in *Agents and Multi-agent Systems: Technologies and Applications 2023*. Vol 35, 2023, pp. 265–275. https://doi.org/10.1007/978-981-99-3068-5_24.
- [9] R. Binns, "Fairness in machine learning: Lessons from political philosophy," in *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency*, 81:1–11, 2018 <https://doi.org/10.48550/arXiv.1712.03586>.
- [10] L. Floridi, J. Cowls, M. Beltrametti, et al., "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," *Minds & Machines*, Vol. 28, 2018, pp. 689–707. <https://doi.org/10.1007/s11023-018-9482-5>.
- [11] A. Sison, and I. Ferrero, "How different is neo-Aristotelian virtue from positive organizational virtuousness?" in *Special Issue: Virtues and Virtuousness: Will the Twain Ever Meet?* Volume24, IssueS2, August 2015, pp. S78–S98. <https://doi.org/10.1111/beer.12099>
- [12] N. Wiener, "Cybernetics: Or Control and Communication in the Animal and the Machine," MIT Press, 1948.
- [13] J. Weizenbaum, "Computer Power and Human Reason: From Judgment to Calculation," W. H. Freeman and Company, 1976.
- [14] J. H. Moor, "What is Computer Ethics?" *Metaphilosophy*, 16, 1985, pp. 266–275. <https://doi.org/10.1111/j.1467-9973.1985.tb00173.x>
- [15] L. Floridi, "What is the Philosophy of Information?" January 24, 2002, available at SSRN: <https://ssrn.com/abstract=3862333> or <http://dx.doi.org/10.2139/ssrn.3862333>
- [16] R. Chatila, J.C. Havens, "The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems," in *Aldinhas Ferreira, M., Silva Sequeira, J., Singh Virk, G., Tokhi, M., E. Kadar, E. (eds) Robotics and Well-Being. Intelligent Systems, Control and Automation: Science and Engineering*, vol 95. Springer, 2019, pp.11–16 https://doi.org/10.1007/978-3-030-12524-0_2
- [17] Ethically Aligned Design A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems, 2016. Available: http://www.standards.ieee.org/develop/indconn/ec/autonomous_systems.html (дата обращения: 07.10.2024)
- [18] J.M. Alvarez, A.B. Colmenarejo, A. Elobaid, et al., "Policy advice and best practices on bias and fairness in AI," *Ethics Inf Technol* 26, 31, 2024, <https://doi.org/10.1007/s10676-024-09746-w>
- [19] Z. Obermeyer, B. Powers, Ch. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science* 366, 2019, pp. 447–453, DOI:10.1126/science.aax2342
- [20] Google AI Principles (2018). <https://ai.google/responsibility/principles/> (дата обращения: 10.11.2024)
- [21] L. Floridi, "Establishing the rules for building trustworthy AI," *Nat Mach Intell* 1, 2019, pp. 261–262, <https://doi.org/10.1038/s42256-019-0055-y>
- [22] N. Bostrom, "Superintelligence: Paths, Dangers, Strategies," Oxford University Press, 2017
- [23] I. Spiegel, "Criticizing Ethics According to Artificial Intelligence," 2024. <https://arxiv.org/html/2408.04609v1>
- [24] L. Munn, "The uselessness of AI ethics," in *AI Ethics*, Vol. 3, 2023, pp. 869–877. <https://doi.org/10.1007/s43681-022-00209-w>.
- [25] L. Floridi, "Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical," *Philos. Technol.* 32, 2019, pp. 185–193, <https://doi.org/10.1007/s13347-019-00354-x>
- [26] W.J. von Eschenbach, "Transparency and the Black Box Problem: Why We Do Not Trust AI," *Philos. Technol.* 34, 2021, pp. 1607–1622, <https://doi.org/10.1007/s13347-021-00477-0>
- [27] N.K. Corrêa, C. Galvão, J.W. Santos, et al., "Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance," in *Patterns (N Y)*, Vol. 4(10):100857, 2023, doi: 10.1016/j.patter.2023.100857.
- [28] B. Mittelstadt, "Principles alone cannot guarantee ethical AI," *Nature Machine Intelligence*, Vol. 1, 2019 pp. 501–507. <https://doi.org/10.1038/s42256-019-0114-4>.
- [29] D. Lauer, "You cannot have AI ethics without ethics," in *AI Ethics*, Vol. 1, 2021, pp. 21–25, <https://doi.org/10.1007/s43681-020-00013-4>.
- [30] A. Caliskan, J. Bryson, A. Narayanan, "Semantics derived automatically from language corpora contain human-like biases," *Science*, Vol. 356 (6334), 2017, pp.183–186.
- [31] A. Hagerty, I. Rubinov, "Global AI Ethics: A Review of the Social Impacts and Ethical Implications of Artificial Intelligence," 2019, <https://arxiv.org/abs/1907.07892> (дата обращения: 17.10.2024).

- [32] P. Wong, “Cultural Differences as Excuses? Human Rights and Cultural Values in Global Ethics and Governance of AI,” *Philos. Technol.* 33, 2020, pp. 705–715, <https://doi.org/10.1007/s13347-020-00413-8>
- [33] E. Goffi, L. Colin, S. Belouali, “Ethical assessment of AI cannot ignore cultural pluralism: a call for broader perspective on AI ethics,” in *The International Journal of Human Rights*, Vol. 1(2), 2021, pp. 151–175.
- [34] F. Trompenaars, “Artificial intelligence: control it or go with the flow,” October, 22, 2018, <https://www.peoplematters.in/article/technology/artificial-intelligence-control-it-or-go-with-the-flow-19612> (дата обращения: 02.11.2024).
- [35] J. Morley, L. Floridi, L. Kinsey, et al, “From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices,” in *Science and Engineering Ethics*, Vol. 26, 2020 pp. 2141–2168. <https://doi.org/10.1007/s11948-019-00165-5>

Кладько Сергей Сергеевич, АНО НЕЙМАРК (e-mail: s.kladko@neimark-it.ru).

Cultural Challenges to Practical AI Ethics: Problems and Solutions

S. S. Kladko

Abstract— The rapid proliferation of artificial intelligence (AI) technologies has amplified the need for ethical frameworks that are not only theoretically robust but also practically implementable. Practical ethics of AI emerges as a pivotal field, bridging the gap between abstract ethical principles and actionable strategies. This article delves into the concept of practical ethics for AI, emphasizing its role in real-world decision-making processes and its applicability across diverse industries. Unlike traditional ethical theories, practical AI ethics focuses on operationalizing principles like fairness, transparency, and accountability within the constraints of technological and societal contexts.

A critical aspect of practical AI ethics is its engagement with cultural diversity, a factor often overlooked in mainstream discourse. Cultural norms and values significantly influence how AI systems are designed, deployed, and perceived across different societies. This article examines how cultural issues intersect with AI ethics, highlighting both challenges and opportunities. Through an analysis of existing frameworks, such as utilitarianism, deontology, and virtue ethics, and their relevance to AI, the discussion extends to the incorporation of culturally inclusive practices. Case studies from healthcare, education, and governance illustrate how culturally attuned AI systems can enhance trust and effectiveness while mitigating biases. Finally, the article proposes a set of actionable guidelines and metrics for embedding cultural sensitivity into AI ethical audits, advocating for interdisciplinary collaboration to ensure AI systems are ethically sound and culturally resonant. By integrating cultural metrics into practical ethics, the article seeks to advance a more holistic and globally relevant approach to AI governance.

Keywords— artificial intelligence, AI Ethics, culture, cross-cultural management.

REFERENCES

- [1] A. Jobin, M. Ienca, and E. Vayena, "The global landscape of AI ethics guidelines." *Nature Machine Intelligence*, 1(9), pp.389–399, 2019
- [2] C. Cath, "Governing artificial intelligence: ethical, legal and technical opportunities and challenges," *Phil. Trans. R. Soc. A*.37620180080, 2018, <http://doi.org/10.1098/rsta.2018.0080>.
- [3] S. Noble, "Algorithms of Oppression: How Search Engines Reinforce Racism," *Science*, no. 374, pp. 542-542, 2021, DOI:10.1126/science.abm5861
- [4] B. D. Mittelstadt, P. Allo, M. Taddeo, S. Wachter, and L. Floridi, "The ethics of algorithms: Mapping the debate," *Big Data & Society*, 3(2), 2016, <https://doi.org/10.1177/2053951716679679>.
- [5] M. Brundage, B. Garfinkel, S. Avin, J. Clark, and H. Toner, "The malicious use of artificial intelligence: Forecasting, prevention, and mitigation," 2018, <https://arxiv.org/ftp/arxiv/papers/1802/1802.07228.pdf>.
- [6] *DataRobot's State of AI Bias Report Reveals 81% of Technology Leaders Want Government Regulation of AI Bias* (Report). DataRobot, 2022. URL: <https://www.datarobot.com/newsroom/press/datarobots-state-of-ai-bias-report-reveals-81-of-technology-leaders-want-government-regulation-of-ai-bias/> (accessed date: 30.09.2024).
- [7] *Artificial Intelligence Index Report. Stanford Institute for Human-Centered Artificial Intelligence*, April 2024 <https://aiindex.stanford.edu/report/> (accessed date: 11.10.2024).
- [8] S. Kladko, "Cultural-Ethical Evaluation in the Launch of AI Education Technologies," *Agents and Multi-agent Systems: Technologies and Applications*. Vol 35, pp. 265-275, 2023, https://doi.org/10.1007/978-981-99-3068-5_24.
- [9] R. Binns, "Fairness in machine learning: Lessons from political philosophy," in *Proceedings of the 2018 Conference on Fairness, Accountability, and Transparency*, vol. 81, pp. 1–11, 2018, <https://doi.org/10.48550/arXiv.1712.03586>.
- [10] L. Floridi, J. Cowls, M. Beltrametti, et al., "AI4People—An Ethical Framework for a Good AI Society: Opportunities, Risks, Principles, and Recommendations," *Minds & Machines*, Vol. 28, pp. 689–707, 2018, <https://doi.org/10.1007/s11023-018-9482-5>.
- [11] A. Sison, and I. Ferrero, "How different is neo-Aristotelian virtue from positive organizational virtuousness?," *Special Issue: Virtues and Virtuousness: Will the Twain Ever Meet?*, Volume24, Issue S2, August 2015, pp. S78-S98, <https://doi.org/10.1111/beer.12099>.
- [12] N. Wiener, "Cybernetics: Or Control and Communication in the Animal and the Machine," MIT Press, 1948.
- [13] J. Weizenbaum, "Computer Power and Human Reason: From Judgment to Calculation," W. H. Freeman and Company, 1976.
- [14] J. H. Moor, "What is Computer Ethics?" *Metaphilosophy*, no. 16, pp. 266-275, 1985, <https://doi.org/10.1111/j.1467-9973.1985.tb00173.x>.
- [15] L. Floridi, "What is the Philosophy of Information?" January 24, 2002, <http://dx.doi.org/10.2139/ssrn.3862333>.
- [16] R. Chatila, J.C. Havens, "The IEEE Global Initiative on Ethics of Autonomous and Intelligent Systems," in Aldinhas Ferreira, M., Silva Sequeira, J., Singh Virk, G., Tokhi, M., E. Kadar, E. (eds) *Robotics and Well-Being. Intelligent Systems, Control and Automation: Science and Engineering*, vol 95. Springer, pp.11-16, 2019, https://doi.org/10.1007/978-3-030-12524-0_2.
- [17] *Ethically Aligned Design A Vision For Prioritizing Wellbeing With Artificial Intelligence And Autonomous Systems*, 2016, http://www.standards.ieee.org/develop/indconn/ec/autonomous_systems.html (accessed date: 07.10.2024).
- [18] J.M. Alvarez, A.B. Colmenarejo, A. Elobaid, et al., "Policy advice and best practices on bias and fairness in AI," *Ethics Inf Technol*, no. 26, 31, 2024, <https://doi.org/10.1007/s10676-024-09746-w>.
- [19] Z. Obermeyer, B. Powers, Ch. Vogeli, and S. Mullainathan, "Dissecting racial bias in an algorithm used to manage the health of populations," *Science*, no. 366, pp. 447-453, 2019, DOI:10.1126/science.aax2342.
- [20] *Google AI Principles*, 2018, <https://ai.google/responsibility/principles/> (accessed date: 10.11.2024)
- [21] L. Floridi, "Establishing the rules for building trustworthy AI," *Nat Mach Intell*, no. 1, pp. 261–262, 2019, <https://doi.org/10.1038/s42256-019-0055-y>.
- [22] N. Bostrom, "Superintelligence: Paths, Dangers, Strategies," Oxford University Press, 2017.
- [23] I. Spiegel, "Criticizing Ethics According to Artificial Intelligence," 2024, <https://arxiv.org/html/2408.04609v1>.
- [24] L. Munn, "The uselessness of AI ethics," *AI Ethics*, vol. 3, pp. 869–877, 2023, <https://doi.org/10.1007/s43681-022-00209-w>.
- [25] L. Floridi, "Translating Principles into Practices of Digital Ethics: Five Risks of Being Unethical," *Philos. Technol.*, no. 32, pp. 185–193, 2019, <https://doi.org/10.1007/s13347-019-00354-x>.
- [26] W.J. von Eschenbach, "Transparency and the Black Box Problem: Why We Do Not Trust AI," *Philos. Technol.*, no. 34, pp. 1607–1622, 2021, <https://doi.org/10.1007/s13347-021-00477-0>.
- [27] N.K. Corrêa, C. Galvão, J.W. Santos, et al., "Worldwide AI ethics: A review of 200 guidelines and recommendations for AI governance," *Patterns*, vol. 4(10),100857, 2023, doi: 10.1016/j.patter.2023.100857.
- [28] B. Mittelstadt, "Principles alone cannot guarantee ethical AI," *Nature Machine Intelligence*, vol. 1, pp. 501–507, 2019, <https://doi.org/10.1038/s42256-019-0114-4>.
- [29] D. Lauer, "You cannot have AI ethics without ethics," *AI Ethics*, vol. 1, pp. 21–25, 2021, <https://doi.org/10.1007/s43681-020-00013-4>.
- [30] A. Caliskan, J. Bryson, and A. Narayanan, "Semantics derived automatically from language corpora contain human-like biases," *Science*, vol. 356(6334), pp.183-186, 2017.

- [31] A. Hagerty, and I. Rubinov, "Global AI Ethics: A Review of the Social Impacts and Ethical Implications of Artificial Intelligence," 2019, <https://arxiv.org/abs/1907.07892> (accessed date: 17.10.2024).
- [32] P. Wong, "Cultural Differences as Excuses? Human Rights and Cultural Values in Global Ethics and Governance of AI," *Philos. Technol.*, 33, pp. 705–715, 2020, <https://doi.org/10.1007/s13347-020-00413-8>.
- [33] E. Goffi, L. Colin, and S. Belouali, "Ethical assessment of AI cannot ignore cultural pluralism: a call for broader perspective on AI ethics," *The International Journal of Human Rights*, vol. 1(2), pp. 151–175, 2021.
- [34] F. Trompenaars, "Artificial intelligence: control it or go with the flow," October, 22, 2018, <https://www.peoplesmatters.in/article/technology/artificial-intelligence-control-it-or-go-with-the-flow-19612> (дата обращения: 02.11.2024).
- [35] J. Morley, L. Floridi, L. Kinsey, et al, "From What to How: An Initial Review of Publicly Available AI Ethics Tools, Methods and Research to Translate Principles into Practices," *Science and Engineering Ethics*, vol. 26, pp. 2141–2168, 2020, <https://doi.org/10.1007/s11948-019-00165-5>.

Kladko Sergei Sergeevich, ANO NEIMARK (e-mail: s.kladko@neimark-it.ru).