

УДК 17:004.8

А. И. Панов

АНАЛИЗ РЕШЕНИЯ ПРОБЛЕМ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА С ЭТИЧЕСКОЙ ТОЧКИ ЗРЕНИЯ В СОВРЕМЕННОМ МИРЕ

Аннотация. В последние несколько лет искусственный интеллект (ИИ) продолжает демонстрировать свое положительное влияние на общество, хотя иногда и с сомнительными с этической точки зрения последствиями. Создание и поддержание общественного доверия к ИИ было определено как ключ к успешным и устойчивым инновациям. В статье обсуждаются проблемы, связанные с внедрением этических принципов ИИ, и представлен комплексный взгляд, который охватывает высокоуровневые этические вопросы ИИ.

Ключевые слова: интеллект, принцип, механизм, система, уровень, ценность, разработка, вопрос, проблема, проектирование, этика, алгоритм.

Когда речь заходит об ИИ и этике, есть два взаимосвязанных аспекта этой темы. Один из них посвящен тому, как ответственно проектировать, разрабатывать и проверять технологии и системы искусственного интеллекта, чтобы люди могли адекватно ответить этическим и юридическим проблемам, особенно касающимся человеческих ценностей. Другой – это использование самого ИИ в качестве средства для достижения определенных целей. В данной статье будет уделено внимание первому вопросу. В последние несколько лет искусственный интеллект продолжает демонстрировать свое положительное влияние на общество, хотя иногда и с сомнительными с этической точки зрения последствиями. Безответственное отношение к ИИ может в теории оказывать разрушительное воздействие на человечество не только в сфере безопасности данных, неприкосновенности частной жизни, но и на уровне трудовых прав. Создание и поддержание общественного доверия к ИИ было определено как ключ к успешным и устойчивым инновациям. Таким образом, вопрос об этическом или ответственном ИИ привлек внимание высокого уровня. Частными компаниями, исследовательскими институтами и общественными организациями было выпущено около ста принципов и руководств по этическому ИИ, и сложился определенный консенсус относительно этих принципов. С другой стороны, принципы и рекомендации далеки от обеспечения надежности систем искусственного интеллекта.

Классификация существующих вопросов по этике ИИ.

Значительное количество исследований было направлено на решение этических проблем ИИ. Глобально на сегодняшний день выделяются три главные проблемы.

1. Реализация рамок этических принципов высокого уровня. Рамки определяют важные этические и правовые принципы, которых должны придерживаться ответственные технологии и системы искусственного интеллекта. Иногда дополнительно разделяют эти принципы высокого уровня на руково-

дящие принципы на командном, организационном и отраслевом уровнях. Эти принципы высокого уровня трудно реализовать на практике по многим причинам, главным образом по технологическим.

2. Разработка этических алгоритмов. Значительный объем исследований был посвящен этическим алгоритмам, в которых формулировка некоторых этических свойств поддается анализу с математической и логической точек зрения. К ним относятся такие свойства, как конфиденциальность и справедливость для определенных типов систем искусственного интеллекта. Это охватывает механизмы, которые имеют дело с предварительной обработкой данных (для устранения предвзятости или индивидуалистических характеристик), сам процесс обучения (для учета этических ограничений), изученные модели (для дальнейшего соответствия этическим ограничениям) и прогнозные результаты (для исправления остаточной предвзятости или выявления индивидуалистической информации). Однако эти механизмы ориентированы на алгоритмы с ограниченной теоретической эвристикой, ограничены небольшим числом поддающихся количественной оценке свойств и небольшим подмножеством этических принципов и человеческих ценностей. В большинстве случаев эти алгоритмы, учитывающие этические нормы, слишком сложны, чтобы их можно было объяснить менее опытным заинтересованным сторонам, и не связаны с более широким процессом принятия решений. Они также не связаны с процессами разработки программного обеспечения, особенно с методами проектирования систем, разработкой требований или процессами проектирования, ориентированными на пользователя.

3. Человеческие ценности в разработке программного обеспечения и их практическая реализация. В последнее время появляются новые исследования данного аспекта, в том числе:

- расширение методов проектирования, основанных на ценности (например, проектирование с учетом ценности VSD (value-sensitive design));
- расширение исследований человеческого фактора в области производительности и удобства использования с учетом человеческих ценностей, но все же ограниченное их небольшим подмножеством;
- методы разработки программного обеспечения для внедрения человеческих ценностей и этических соображений на протяжении всего цикла разработки;
- шаблоны архитектуры и дизайна, которые могут обеспечить (с сильными математическими гарантиями) определенные этические или связанные с человеческими ценностями атрибуты качества, такие как конфиденциальность.

Внедрение этики искусственного интеллекта в практику.

Этика ИИ может быть внедрена различными способами. Существует много подходов и механизмов, выходящих за рамки высокоуровневых принципов и низкоуровневых алгоритмов.

1. *Требования.* В дополнение к обычным функциональным и нефункциональным требованиям этические принципы могут быть определены как подмножество требований к системам искусственного интеллекта. Для решения проблемы автономии ИИ может потребоваться механизм снижения рисков.

2. *Проектирование.* Могут быть предложены шаблоны и механизмы проектирования для учета этических принципов. В распределенных системах обучения также существуют проблемы с доверием между участвующими узлами. Например, в федеративных системах обучения координатор обучения может стать единственной точкой отказа. Таким образом, при разработке моделей и механизмов следует тщательно учитывать все заинтересованные стороны.

3. *Операции.* Для постоянной проверки соблюдения этических принципов необходимы механизмы непрерывной валидации и мониторинга. Вовлеченные заинтересованные стороны должны определить пороговое значение для значительного воздействия, которое запускает механизмы проверки.

4. *Управление.*

Сертификация этической зрелости. Схема сертификации зрелости этики ИИ может быть разработана для оценки этической зрелости организации в отношении управления проектами искусственного интеллекта. На основе анализа проектов ИИ организации могут быть выданы сертификаты этической зрелости различных уровней. Уровень сертификата может быть повышен или понижен.

Проверка этики. Для соблюдения этических принципов могут быть проведены внутренние и внешние обзоры этического воздействия. Ожидается, что к проведению обзоров будут присоединяться представители заинтересованных сторон.

Проектная группа. При создании команды разработчиков следует учитывать разнообразие членов команды (например, происхождение, культуру и дисциплины специализации).

Таким образом, были рассмотрены основные проблемные вопросы ИИ с этической точки зрения в современном мире, а также проанализированы существующие решения и механизмы, некоторые из которых будут с высокой вероятностью введены в эксплуатацию в будущем. ИИ – это технология, являющаяся флагманом науки в последние годы, интерес и внимание к которой неустанно растут, именно поэтому предельно важно заострять внимание не только на технологической, но и на этической и моральной сторонах вопроса развития технологии искусственного интеллекта.

СПИСОК ИСПОЛЬЗОВАННОЙ ЛИТЕРАТУРЫ

1. Этические вопросы искусственного интеллекта [Электронный ресурс] // Хабр. – Режим доступа: habr.com/ru/company/kaspersky/blog/421791/. – Дата доступа: 30.05.2022.

2. **Карпов, В. Э.** К вопросу об этике и системах искусственного интеллекта / В. Э. Карпов, П. М. Готовцев, Г. В. Ройзензон // Философия и общество. – 2018. – № 2 (87). – С. 84–105.

3. Этика искусственного интеллекта [Электронный ресурс] // Википедия. – Режим доступа: ru.wikipedia.org/wiki/Этика_искусственного_интеллекта. – Дата доступа: 03.06.2022.

4. **Разин, А. В.** Этика искусственного интеллекта / А. В. Разин // Философия и общество. – 2019. – № 1 (90). – С. 57–73.

5. Этичное применение искусственного интеллекта [Электронный ресурс] / П. М. Готовцев [и др.] // Этика и «цифра»: проблемы цифровых технологий: аналит. докл. – Режим доступа: ethics.cdto.ranepa.ru/3_3. – Дата доступа: 06.06.2022.

УДК 94(47).084.9

Д. Н. Постников

ИСТОРИЯ РАСПАДА СССР СКВОЗЬ ПРИЗМУ АНАЛИТИКИ ТРАВМЫ

Аннотация. Категория травмы рассмотрена в статье в качестве специфического объекта междисциплинарного исследования. Автор обосновывает возможность применения аналитики травмы в процессе изучения истории распада СССР. Сделан вывод о том, что нарратив о прекращении существования Советского Союза наполнен личными и общественными репрезентациями в настоящем.

Ключевые слова: историческая травма, аналитика травмы, распад СССР, постсоветское пространство.

Социальные потрясения и тот опыт катастроф, которые пережило человечество в XX в., безусловно, требует осмысления. Воспоминания о прошлых травмирующих событиях конструируются в различных социальных и культурных контекстах. Дискуссии о травматическом опыте как предмете социального анализа ведутся не одно десятилетие [5, 7].

Во вступительной статье к сборнику «Травма: пункты» отмечено, что изучение травмы гуманитарными и общественными науками может быть выстроено по трем ключевым линиям:

1) понимание травмы как опыта утраты («травма-как-утрата»), когда «утрата становится причиной переоценки уже несуществующего «целого», а «сама картина «целостности» возникает лишь впоследствии – в процессе осмысления утраты» [7, с. 9];

2) исследование травмы как символической матрицы, определяющей общую систему повествовательных координат («травма-как-сюжет»);

3) рассмотрение травмы в качестве консолидирующего события [7, с. 8–9].

В процессе рассмотрения истории распада СССР в контексте обращения к травматическому опыту важным является ответ на вопрос о том, насколько аналитика травмы (сформировавшаяся по большому счету в американском варианте) эффективна при изучении советского прошлого и событий, развернувшихся на постсоветском пространстве в 1990-е гг. И тогда становится очевидным то, что исследование истории распада СССР «нуждается в самостоятель-