МГИМО УНИВЕРСИТЕТ

AI CENTRE MGIMO
EMPOWER INNOVATIONS

Анна Абрамова
Анастасия Рыжкова
Юлия Церех

ОЦЕНКА ЭТИЧЕСКИХ АСПЕКТОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА НАЦИОНАЛЬНОМ И МЕЖДУНАРОДНОМ УРОВНЯХ. УРОВЕНЬ «ИССЛЕДОВАТЕЛЬСКИХ И НАУЧНЫХ ЦЕНТРОВ»

коллекция исследований ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ ДЛЯ РАЗВИТИЯ

приоритет2030
Лидерами становятся

МОСКОВСКИЙ ГОСУДАРСТВЕННЫЙ ИНСТИТУТ
МЕЖДУНАРОДНЫХ ОТНОШЕНИЙ
(УНИВЕРСИТЕТ) МИД РОССИИ

Центр ИИ МГИМО (У) МИД России

Анна Абрамова, Анастасия Рыжкова, Юлия Церех

ОЦЕНКА ЭТИЧЕСКИХ АСПЕКТОВ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА НА НАЦИОНАЛЬНОМ И МЕЖДУНАРОДНОМ УРОВНЯХ.

УРОВЕНЬ «ИССЛЕДОВАТЕЛЬСКИЕ И НАУЧНЫЕ ЦЕНТРЫ»

Исследование

Москва, 2022

ISBN 978-5-6047689-6-9

Авторы:
Анна Абрамова, к.э.н., директор Центра искусственного интеллекта МГИМО, руководитель кафедры Цифровой экономики и искусственного интеллекта группы компаний АДВ в МГИМО (У) МИД России
Анастасия Рыжкова, к.т.н., научный сотрудник Центра искусственного интеллекта МГИМО
Юлия Церех, младший научный сотрудник Центра искусственного интеллекта МГИМО

Аннотация
Этические аспекты искусственного интеллекта (ИИ) становится одним из основных элементов мягкого права в регулировании национального и международного рынка. В декабре 2021 г. ЮНЕСКО приняла Рекомендацию по этическим аспектам искусственного интеллекта, в которой излагаются подходы к международному мягкому регулированию, уделяющие особое внимание этике. Международная система оценки этических аспектов ИИ может стать основой для оценки этического воздействия в сочетании с Рекомендацией и Принципами ИИ ОЭСР, а также подходами к классификации ИИ, разработанной экспертами данной организации. Методология ставит человека во главу угла и включает все ключевые заинтересованные стороны на протяжении всего жизненного цикла системы ИИ. Данные для индекса могут быть взяты из существующих баз данных ЮНЕСКО, ОЭСР, ЮНКТАД, ВЭФ. Но специфика темы расширяется за счет уточнения и проработки данных, которые могут быть добавлены в таблицы национальной статистики на макроуровне, а также из исследований, которые охватывают микроуровень.

Оценка этических аспектов Искусственного интеллекта на национальном и международном уровнях. Уровень исследовательских и научных центров

JEL F01, F20, F42, F53, F55, F60, F63, F68
Ключевые слова: искусственный интеллект, этика, индекс, мягкое право

Обложка: canva.com

Moscow, 2022

Содержание

Используемые сокращения

ИИ	Искусственный интеллект
ИКТ	Информационно-коммуникационные технологии
ИС	Интеллектуальная собственность
ННФ	Национальный научный фонд
ОЭСР	Организация экономического сотрудничества и развития
ЮНЕСКО	Специализированное учреждение Организации Объединённых Наций по вопросам образования, науки и культуры
НИОКР	Научно-исследовательские и опытно-конструкторские работы
ВЭФ	Всемирный экономический форум

Введение

Этот документ является следующим шагом в продолжающемся исследовании Центра ИИ МГИМО по разработке международной системы оценки этических аспектов искусственного интеллекта. В отдельных статьях подробно обсуждаются все ключевые субиндексы, включая основные индикаторы и сложные вопросы.

Международная система оценки этических аспектов искусственного интеллекта — это комплексный подход, в котором участвуют все ключевые участники жизненного цикла ИИ — государство, бизнес, гражданское общество, исследовательские центры/мозговые центры. Более того, мы предлагаем оценку по трем субиндексам областей, которые способствуют устойчивому развитию ИИ на всех этапах и влияют на все группы участников – грамотность в области ИИ, инвестиции в НИОКР и развитие инфраструктуры ИКТ.

Опубликованная в феврале 2022 г. Общая структура международной системы оценки этических аспектов искусственного интеллекта (Абрамова, Рыжкова и Церех, 2022) охватывает все основные элементы индекса, представляющего группы ключевых показателей. Этот исследовательский документ сосредоточен на подробном освещении вклада субиндекса «Исследовательские центры/мозговые центры» в восприятие и развитие этических аспектов ИИ, а также освещает проблемные аспекты для всех типов исследовательских организаций.

Структура исследования следующая: первый раздел является введением, второй посвящен методологии Международной системы оценки этических аспектов искусственного интеллекта, сфокусированным на возможном вкладе исследовательских центров в отношении этики ИИ, а последний раздел посвящен обсуждению, охватывающему наиболее сложные вопросы в области этических аспектов ИИ. практическая реализация субиндекса в рамках различных групп исследовательских центров.

Авторы благодарны руководителям и координаторам проекта «Национальный приоритет 2030» за возможность проведения исследования.

Ландшафт исследовательских центров: растущее разнообразие

Исследовательские центры являются одним из основных столпов в развитии ИИ. Растущие инвестиции в ИИ подстегивают спрос на технологические достижения, понимание потенциала технологии для повышения эффективности, обсуждение междисциплинарных вопросов, включая этические аспекты ИИ. Растущий исследовательский интерес к этическим аспектам ИИ был зафиксирован за счет роста количества публикаций, резко возросших после 2015 года, в 2019 году их число достигло 70 в год.[1]

Исследовательские центры могли бы внести свой вклад в развитие вышеупомянутых аспектов в рамках проводимой ими деятельности. Растущий ландшафт исследовательской деятельности в области ИИ можно классифицировать на основе следующих ключевых характеристик — акцент на ИИ, уровень исследований (государственный, частный), отрасли применения, локализации, участия в различных сотрудничествах.

Одним из первых индикаторов, о котором стоит упомянуть, можно назвать «фокус на ИИ». Большое количество специалистов начинали свои исследования с центров с широким спектром деятельности от анализа экономического развития цифровых технологий. Позже в центрах такого типа появились лаборатории или инициативы, ориентированные на ИИ. Одним из ярких примеров является Институт Брукингса, некоммерческая общественная организация, проводящая исследования в области развития на национальном и международном уровнях. Одной из последних инициатив Института является создание Форума по сотрудничеству в области ИИ, в котором этика

[1] NSF partnerships expand National AI Research Institutes to 40 states.July 29, 2021. https://hai.stanford.edu/news/state-ai-10-charts

является одним из вопросов для обсуждения (Керри С., МЕЛЬЦЕР Дж., РЕНДА А., 2022).

Последние несколько лет ознаменовались новой волной роста государственных инвестиций в ИИ в странах-лидерах. Государственная долгосрочная поддержка фундаментальных исследований в области ИИ в первые десятилетия создала почву для дальнейших краткосрочных инвестиций со стороны частного сектора (ОЭСР, 2021). За период 2001-2019 гг. государственное финансирование США увеличилось в 17 раз.

В 2020 году Национальный научный фонд (NSF) учредил семь национальных научно-исследовательских институтов ИИ. В 2021 году список пополнился новыми одиннадцатью центрами, занимающимися содействием развитию ИИ в сотрудничестве с ведущими частными компаниями и федеральными агентствами. По данным ННФ, совокупные инвестиции достигают 220 млн долларов.[2]

Кроме того, частный бизнес создает исследовательские подразделения ИИ, ориентированные в основном на технологические достижения, а не на исследования этики ИИ. По данным Salesforce, этические аспекты ИИ были одной из главных проблем потребителей в 2018 году. Руководители ИИ из бизнеса ответили принятием принципов ИИ и введением административных единиц или групп по этическим аспектам ИИ. В этом отношении субиндекс «Бизнес» лучше отражает вклад частного сектора в развитие этических аспектов ИИ.

В отношении отраслевого распределения может применяться подход ОЭСР для анализа финансирования НИОКР, связанных с ИИ (ОЭСР 2021). Ключевыми секторами могут быть «общие методы ИИ, предпосылки и влияние ИИ (например, образование и обучение и социальное влияние), области ИИ (такие как компьютерное зрение и обработка естественного языка), медицинские приложения ИИ и немедицинские области применения

[2] NSF partnerships expand National AI Research Institutes to 40 states. July 29, 2021.https://beta.nsf.gov/news/nsf-partnerships-expand-national-ai-research

ИИ (таких как бизнес и социальные науки). Для исследовательских центров из каждого из упомянутых выше секторов могут применяться этические метрики ИИ.

Метрики локализации могут охватывать национальный, региональный или международный уровни. В этом отношении может применяться методология ВОИС в отношении баз данных ИС.

Методология

Обзор возможных сценариев

Субиндекс по оценке этических аспектов ИИ на уровне аналитических центров требует привлечения широкого круга участников (университетов, школ, ученых) или организация сбора специальной статистики на уровне государства.

В основу методики оценки авторы взяли экспертный подход. Такой подход наиболее актуален для вопросов проведения масштабного периодического многофакторного исследования.

Во-первых, авторы рассмотрели и сравнили пять наиболее распространенных и проверенных методов оценки сложных и динамических систем, таких как:

- мозговой штурм;
− анализ слабых и сильных сторон;
− метод построения карты;
− метод Дельфи;
− экспертная оценка.

Каждый из рассмотренных методов имеет свои особенности и ограничения в применении.

1) Мозговой штурм

Этот метод довольно хорош для оценки на уровне аналитических центров, но не подходит для рамочного подхода.

Как правило, мозговой штурм проводится внутри проектной команды с возможностью привлечения к работе стороннего эксперта. Эксперт может обладать широкими или, наоборот, узкоспециальными знаниями, что, по мнению руководителя проектной группы, важно при реализации проекта.

Алгоритм метода достаточно прост и состоит из нескольких шагов:

1. Участники составляют максимально подробный список параметров, актуальных для проекта

2. Параметры с наименьшей вероятностью реализации удаляются из лонг-листа большинством участников.

Достоинства метода: скорость получения результата, простота реализации метода.

Недостатки метода: качество анализа напрямую зависит от опыта и кругозора лиц, участвующих в мозговом штурме.

Возможность применения метода оценки этических аспектов использования технологий ИИ:

− требует привлечения опытной проектной команды для внедрения аналогичных продуктов,

− высокая стоимость

− сложность привлечения профильных специалистов.

2) Анализ слабых и сильных сторон

Этот метод эффективен, но имеет слишком сложную практическую реализацию даже при использовании новых цифровых технологий. Принимая во внимание широкий спектр акторов в рассматриваемой тематике этот метод мало подходит для периодической оценки в рамках многофакторного анализа.

Метод аналогичен методу анализа предположений, однако команда проекта составляет список потенциальных параметров, выявляя и впоследствии анализируя их слабые/сильные стороны.

Преимущества: детальное рассмотрение параметров Индекса.

Недостатки:

− длительная реализация метода;

− чрезмерная детализация метода;

− качество анализа напрямую зависит от опыта и кругозора привлекаемых специалистов.

Возможность применения метода оценки этических аспектов использования технологий ИИ: команда проекта при недостаточном опыте может упустить существенные параметры и аспекты.

3) Метод построения диаграмм

Для подхода на уровне исследовательских центров этот метод слишком дорог и имеет сложности с поиском соответствующих специалистов.

Метод осуществляется внутри проектной команды с возможностью приглашения внешнего эксперта. Анализ проходит в три этапа:

− составление причинно-следственных связей,

− создание блок-схемы реализуемых процессов,

− составление диаграмм воздействия.

Преимущества: качественный учет потенциальных рисков проектов.

Недостатки: реализация метода построения диаграмм требует от проектной команды навыков работы с данным методом и значительных временных затрат.

Возможность применения метода оценки этических аспектов использования технологий ИИ: применение этого навыка требует специальных компетенций и опыта.

4) Метод Дельфи

Этот метод эффективен в случае использования цифровых технологий, но для общепринятых рамок оценки этических аспектов ИИ имеет слишком длительный временной диапазон. Для оценки деятельности исследовательских центров метод достаточно хорош к применению.

Метод Дельфи предполагает проведение большого анонимного опроса внешних и внутренних экспертов, обобщение собранных данных, выдачу заполненных анкет другой экспертной группе с последующим очным обсуждением результатов, а затем повторное проведение анонимного опроса с подведением итогов. окончательные результаты и составление списка потенциальных рисков.

Достоинства: Качественная проработка.

Недостатки: метод требует длительной реализации и финансовых ресурсов для реализации.

Возможность применения метода оценки этических аспектов использования технологий ИИ: метод требует больших затрат времени и средств.

5) Экспертная оценка

Метод экспертных оценок аналогичен методу Дельфи, однако предполагает открытый опрос экспертов, имеющих опыт как в области деятельности исследовательских центров, так и в области работы искусственного интеллекта.

Преимущества: качественная проработка выявления потенциальных рисков.

Недостатки: требуется создание базы специалистов, готовых участвовать в большом суре.

Возможность применения метода оценки этических аспектов использования технологий ИИ: метод требует много времени.

Формула подсчета итогового результата

Авторы основывали оценку групп показателей на индексе значимости, который рассчитывается по формуле:

$$r_{ij}^{k} = \alpha_{ij}\beta_{ij}^{k}, \qquad (1)$$

где

r_{ij}^{k} - значимость i-го показателя, оцененного j-м респондентом, с точки зрения влияния на k-фактор,

$i=(1...N),$ где N – количество параметров, рассматриваемых в исследовании,

$j=(1...n),$ где n - количество полученных ответов,

$k=(1...5),$ где 1…5 – номера групп влияния соответственно (соответственно стоимость, время выполнения ИТ-проекта, качество продукта, среда, безопасность),

α_{ij} - вес значимости показателя i, оцененный j-м респондентом,

β_{ij}^{k} - величина «эффекта» влияния показателя на рассматриваемого заинтересованного лица и/или преследуемые им цели.

Для оценки среднего значения показателей рассчитывается Индекс значимости показателя по формуле:

$$R_i^k = \frac{\sum_{j=1}^{n} r_{ij}^k}{n} = \frac{1}{n}\sum_{j=1}^{n} \alpha_{ij}\beta_{ij}^k \quad (2)$$

Предлагаемые параметры для расчета

РЕГИСТРАЦИЯ ИНТЕЛЛЕКТУАЛЬНОЙ СОБСТВЕННОСТИ:

− количество внедрений продуктов на базе технологий ИИ в исследовательских центрах и аналитических центрах,

− динамика реализации кейсов, связанных с этическими аспектами ИИ (текущая/предыдущая)

− объем проекта

− стоимость проекта

− количество патентов

− динамика количества новых патентов

− количество привлеченных участников

− уровень проекта: местный, национальный, транснациональный

− количество проектов ИИ без реализации этических аспектов ИИ, созданных исследовательскими центрами и аналитическими центрами.

− количество проектов ИИ без реализации аспектов этики ИИ с привлечением исследовательских центров и аналитических центров.

− соблюдение этических требований ЮНЕСКО

ЭТИЧЕСКИЕ ПРОБЛЕМЫ В НИОКР:

− количество случаев по этическим аспектам ИИ, случайно замеченных сотрудниками исследовательских центров и аналитических центров,

− количество обращений в компетентные органы для решения проблем

− сектор /-ы, где произошел случай, связанный с этическими аспектами ИИ

− предполагаемые негативные сценарии развития

− количество реальных дел, вновь защищающих от негативных последствий этических аспектов ИИ

ПУБЛИКАЦИИ ПО ЭТИЧЕСКИМ АСПЕКТАМ ИИ:

− количество рецензируемых научных журналов с публикациями по этическим аспектам ИИ

− динамика количества рецензируемых научных журналов с публикациями по этическим аспектам ИИ

− количество публикаций об этических аспектах ИИ в рецензируемых научных журналах

− динамика количества публикаций об этических аспектах ИИ в рецензируемых научных журналах

− формат научного журнала

− Охват аудитории журнала

− соблюдение этических требований ЮНЕСКО

МЕРОПРИЯТИЯ, ОСВЕЩАЮЩИЕ ЭТИЧЕСКИМ СПЕКТАМ ИИ:

− группа показателей, охватывающая национальные и международные конференции и форумы по этическим аспектам ИИ, которые могут предоставить всем заинтересованным сторонам возможность обменяться мнениями по вопросам для дальнейших исследований.

− количество популярных мероприятий в области этических аспектов ИИ:

- местный
- национальный
- по всему миру

− динамика количества событий

- местный
- национальный
- по всему миру

− количество участников

− плата за участие

− доступность мероприятия для широкого круга лиц

− профессиональный уровень участников

− соблюдение этических требований ЮНЕСКО

ЭТИЧЕСКИЕ ВОПРОСЫ ИССЛЕДОВАТЕЛЕЙ, РАБОТАЮЩИХ С ИИ:

− количество исследователей, занимающихся этическими проблемами ИИ

− динамика исследователей, занимающихся этическими проблемами ИИ (текущая/предыдущая)

− объем проекта

− стоимость проекта

− количество привлеченных участников

− уровень: местный, национальный, транснациональный

− соответствие проекта этическим требованиям ЮНЕСКО

Заключение

Растет число исследовательских центров, занимающихся ИИ, которые финансируются из государственных и частных источников. Национальная политика и приоритеты в области ИИ ставят этические аспекты ИИ в среднесрочную перспективу развития сектора ИИ. Мультидисциплинарные исследования в настоящее время привлекают все больше внимания как компьютерщиков, так и социологов. Одними из лучших индикаторов, отражающих текущий этап развития, являются публикации и события. Но необходимы дальнейшие улучшения в сборе данных о регистрации ИС для реструктуризации обсуждения этических исследований ИИ, особенно в области патентования и защиты авторских прав.

Текущие данные подчеркивают растущий интерес к этическим аспектам ИИ со стороны государственных органов и бизнеса. Но данные об исследованиях этические аспекты ИИ по-прежнему скудны и фрагментарны. Подиндекс Сбор показателей научно-исследовательских центров/мозговых центров с точки зрения международного сотрудничества может быть поддержан ЮНЕСКО и ВОИС путем введения дополнительных таблиц, ориентированных на этические оценки ИИ.

Библиография

Абрамова Анна и Рыжкова Анастасия и Церех Юлия Оценка этики ИИ на национальном и международном уровнях. Подход к структуре индекса и методологии (18 февраля 2022 г.). [Электронный ресурс]. Ссылка доступа: SSRN: https://ssrn.com/abstract=4096669 or http://dx.doi.org/10.2139/ssrn.4096669

AI Ethics: Another Step Closer to the Adoption of UNESCO's Recommendation. [Электронный ресурс]. Ссылка доступа: URL https://en.unesco.org/news/ai-ethics-another-step-closeradoption-unescos-recommendation-0

EUROPEAN COMISSION 2021. AI watch index. [Электронный ресурс]. Ссылка доступа: URL https://op.europa.eu/en/publication-detail/-/publication/15568192-a95f-11eb-9585-01aa75ed71a1/language-en/format-PDF/source-209026200

Kerry C., Meltzer J., Renda A. (2022). AI cooperation on the ground: AI research and development on a global scale. Report, November 4, 2022. [Электронный ресурс]. Ссылка доступа: URL https://www.brookings.edu/wp-content/uploads/2022/11/FCAI-October-2022.pdf

Multistakeholder group discusses ten building blocks towards creating inclusive AI policies. [Электронный ресурс]. Ссылка доступа: URL https://en.unesco.org/news/multistakeholder-group-discusses-ten-building-blocks-towards-creating-inclusive-ai-policies

Multistakeholder group discusses ten building blocks towards creating inclusive AI policies. [Электронный ресурс]. Ссылка доступа: URL https://en.unesco.org/news/multistakeholder-group-discusses-ten-building-blocks-towards-creating-inclusive-ai-policies

OECD 2021. Izumi Yamashita, Akiyoshi Murakami, Stephanie Cairns, Fernando Galindo-Rueda. Measuring the AI content of government-funded R&D projects: A proof of concept for the OECD Fundstat initiative. [Электронный ресурс]. Ссылка доступа: URL https://www.oecd-

ilibrary.org/science-and-technology/measuring-the-ai-content-of-government-funded-r-d-projects_7b43b038-en

OECD 2022. FRAMEWORK FOR THE CLASSIFICATION OF AI SYSTEMS [Электронный ресурс]. Ссылка доступа:URL https://www.oecd-ilibrary.org/docserver/cb6d9eca-en.pdf?expires=1646822229&id=id&accname=guest&checksum=0D6117C31817EA6B9FFDB7A65AACAFCF

UNESCO 2021. Recommendation on the ethics of artificial intelligence. [Электронный ресурс]. Ссылка доступа: URL https://unesdoc.unesco.org/ark:/48223/pf000038

WIPO 2019. WIPO Technology Trends 2019. Artificial Intelligence. [Электронный ресурс]. Ссылка доступа: URL https://www.wipo.int/edocs/pubdocs/en/wipo_pub_1055.pdf

Центр искусственного интеллекта МГИМО создан для расширения международного сотрудничества и взаимодействия со всеми субъектами цифровой экономики как на национальном, так и на международном уровне. Наше междисциплинарный научный подход сосредоточен на повестке дня международного сотрудничества, национальной политике в области ИИ и возможностях для бизнеса. Международная торговля и торговая политика (приоритет цифровой торговли), устойчивое развитие, этика ИИ — ключевые направления нашей деятельности.

На базе Университета МГИМО мы развиваем международную экспертную площадку по искусственному интеллекту с регулярными конференциями и круглыми столами, научными статьями и исследовательскими работами. Наша расширяющаяся сеть стратегических партнерств позволяет предоставлять консультационные и иные решения в области ИИ как для бизнеса, так и для государственных учреждений.

Центр основан в октябре 2021 года

Наши контакты

143007, Одинцово, Московская область,
Ново-Спортивная 3
https://aicentre.mgimo.ru
aicentre@inno.mgimo.ru
+7 903 623-95-15
https://t.me/aicentremgimo

AI CENTRE MGIMO
EMPOWER INNOVATIONS

приоритет2030^
Лидерами становятся