

Секция «Научно-технологическая и инновационная политика в современной России и мире»

Этические нормы в технологиях искусственного интеллекта

Научный руководитель – Яковлева Александра Фёдоровна

Демченко Оксана Юрьевна

Аспирант

Государственный академический университет гуманитарных наук, Москва, Россия

E-mail: svetoxana@gmail.com

Аннотация: Текущее состояние развития технологий искусственного интеллекта ставит перед человечеством массу вызовов, в том числе этических. Этика взаимодействия «человек-машина» становится всё более сложной, также как и развиваемые человечеством технологии, что требует особого подхода к формированию доверия к технологиям, в том числе путём внутрикорпоративного и законодательного закрепления принципов этики работы алгоритмов. В разрезе общественной этики данный вопрос попадает в пространство политической мысли и стимулирует поиски подходов к разрешению накопившихся противоречий в том числе политическими инструментами. Сегодня значимую роль в разработке, составлении и применении этических принципов работы технологий и с технологиями искусственного интеллекта играют коммерческие компании, напрямую заинтересованные в вопросе. В докладе представлены и проанализированы корпоративные и международные примеры этических принципов, на основе которых предполагается строить политику работы с технологиями искусственного интеллекта на уровне государств, а также представлены тенденции развития данного направления политической мысли.

Последние годы мы можем наблюдать развитие направления искусственного интеллекта (Далее - ИИ) как двигателя не только технологической, но и онтологической мысли. Онтология познания со всеми сложностями её изучения по-прежнему ставит вопросы о мышлении, его принципах и этических аспектах взаимодействия «человек-машина».

Начиная с таких представителей научной мысли как Алан Тьюринг [Тьюринг, 1950] и Джон Сёрл [Сёрл, 1980], своё развитие получила тема этического осмысления работы машин.

Сегодня для нас как для человечества встает вопрос моральной ответственности технологических алгоритмов перед обществом за действия, направленные как на другие алгоритмы, так и на человека. Вопрос моральной ответственности по-прежнему остается непроясненным и в юридическом смысле, что требует глубокой проработки и исследований, однако общая острота обозначенной проблемы диктует необходимость выявления единых этических правил разработки и применения человеком алгоритмов на основе ИИ в самое скорое время.

Высокий уровень общественной значимости вопроса выводит его на политический уровень, и можно говорить, что технология и этика в изложенном вопросе приобретают общеполитическую значимость.

В ноябре 2019 года в России была принята «Национальная стратегия развития искусственного интеллекта на период до 2030 года», в которой обозначена необходимость разработки «этических правил взаимодействия человека с искусственным интеллектом» [Стратегия, 2019].

На данный момент в общемировом информационном поле не выражен общепризнанный лидер мнений, задающий повестку развития темы. Можно выделить государства

лидеры и компании-локомотивы [Fjeld, 2020], которые путем создания локальных документов стремятся обозначить свою позицию по вопросам этики работы технологий. Важно заметить, что в условиях отсутствия международных документов и четкого лидерства, по-прежнему остается пространство для появления новых политических акторов, а также для выявления отраслей, наиболее остро нуждающихся в соответствующей регламентации с необходимостью принятия локальных отраслевых документов.

Среди существующих документов, отражающих этические нормы работы с технологиями, можно встретить следующие общие принципы, получившие свое распространение в юридических документах, научных и публицистических работах по теме, а также локальных (в том числе внутрикорпоративных) этических формулярах:

- Приватность;
- Подотчётность;
- Безопасность и защита;
- Прозрачность и объяснимость;
- Справедливость и отсутствие дискриминации;
- Человеческий контроль над технологией;
- Профессиональная ответственность;
- Соблюдение и развитие человеческих ценностей.

На основании вышеприведенных принципов могут быть выработаны общие этические правила взаимодействия «человек-машина». Очевидно, встает вопрос выбора субъекта права, обладающего достаточным политическим весом и волей для разработки и ратификации документа общемирового значения. Предполагаем, что возможность ратификации подобного документа может быть доступна организациям, подобным Организации Объединенных Наций (ООН) или Всемирной организации здравоохранения (ВОЗ), если речь касается локальных отраслевых правил. Примером подобного документа может служить действующая Конвенция о дорожном движении от 08 ноября 1968 года [Конвенция, 1971], разработанная Экономическим и Социальным Советом ООН (ЭКОСОС).

Вопрос выработки универсальных этических принципов работы технологий и с технологиями является значимым в контексте стремительного развития алгоритмов на основе искусственного интеллекта.

В настоящий момент обозначенный в докладе вопрос остро нуждается в детальной проработке, и мировому сообществу в лице государств, общественности и бизнеса ещё предстоит объединить усилия в выработке единого взгляда на этические нормы работы алгоритмов искусственного интеллекта.

Источники и литература

- 1) OECD, Recommendation of the Council on Artificial Intelligence, OECD/LEGAL/0449, 2021 Edition
- 2) Fjeld Jessica, Nele Achten, Hannah Hilligoss, Adam Nagy, and Madhulika Srikumar. "Principled Artificial Intelligence: Mapping Consensus in Ethical and Rights-based Approaches to Principles for AI." Berkman Klein Center for Internet & Society, 2020
- 3) Азимов А. Хоровод. 1942
- 4) ГОСТ «Системы искусственного интеллекта. Способы обеспечения доверия. Общие положения» ГОСТ Р 59276-2020
- 5) Конвенция о дорожном движении. 08.11.1968 г

- 6) Национальная стратегия развития искусственного интеллекта на период до 2030 года ст. 49 п. ж., утвержденная указом Президента РФ от 10.10.2019 г. № 490
- 7) Сбер одним из первых в РФ утвердил принципы этики искусственного интеллекта. Электронный ресурс. Режим электронного доступа: [<https://press.sber.ru/publications/sber-odnim-iz-pervykh-v-rf-utverdil-printsipy-etiki-iskusstvennogo-intellekta>]. Дата последнего обращения 07.03.2021
- 8) Сёрл Д. Сознание, мозг и программы. 1980
- 9) Тьюринг А. Могут ли машины мыслить? 1950