

УДК 174

О. Ю. Демченко

Этика и технологии искусственного интеллекта: политический аспект

Аннотация:

Развитие технологий, в частности технологий искусственного интеллекта, ставит перед человечеством важный этический вопрос о взаимодействии «человек-машина». Значимым представляется осмысление и выработка единых этических принципов, которые должны лечь в основу общепризнанной политики, применяемой к технологиям в целом и к «умным» машинам в частности. В силу специфики своей деятельности и нарастающей необходимости внутреннего регулирования работы систем частные компании принимают на себя роль составителей и главных пользователей этических протоколов, таким образом становясь ключевыми стейкхолдерами в вопросе этического осмысления работы машин. В настоящее время назрела необходимость наладить комплексное взаимодействие институтов власти, общества и компаний по вопросам этического осмысления работы алгоритмов и их взаимодействия как с другими алгоритмами, так и с людьми. В статье приводятся результаты анализа различных документов корпоративного и международного уровня, на основе которых предполагается строить политику работы с технологиями искусственного интеллекта на уровне государств, а также представлены тенденции развития данного направления политической мысли.

Ключевые слова: этика технологий, искусственный интеллект, законодательное регулирование технологий, технологическая политика

Об авторе: Демченко Оксана Юрьевна, Государственный академический университет гуманитарных наук (ГАУГН), аспирант, независимый консультант по технологическому бизнесу; эл. почта: svetoxana@gmail.com

В начале 2020-х гг. мы можем наблюдать развитие направления искусственного интеллекта (далее – ИИ) как двигателя не только технологической, но и онтологической мысли. Онтология познания со всеми сложностями ее изучения продолжает ставить

вопросы о мышлении, его принципах и этических аспектах взаимодействия «человек-машина».

Инициаторами изучения темы этического осмысления работы машин стали такие представители научной мысли как Алан Тьюринг [8] и Джон Сёрл [1].

Десятилетия спустя после появления первых работ на эту тему перед человечеством все так же стоит вопрос моральной ответственности технологических алгоритмов перед обществом за результаты своих действий (или бездействий), направленных как на другие алгоритмы, так и на человека. При этом вопрос моральной ответственности все так же остается непроясненным и в юридическом смысле, что требует его дальнейшей проработки и междисциплинарных исследований. Однако общая острота обозначенной проблемы диктует необходимость скорейшего выявления единых этических правил разработки и применения человеком алгоритмов на основе ИИ.

Особенность данной публичной сферы – ее кажущаяся отделенность и независимость от общественно-значимых процессов, а также традиционное сведение вопросов создания регламентов работы систем к технической, второстепенной задаче. В то же время техническое регулирование и этика технологий тесно связаны с такими направлениями мысли как: общемировой прогресс, динамика развития обществ, политическая культура, экономическое развитие и стагнация и др.

Осмысление и выработка единых этических принципов, которые должны лечь в основу общепризнанной политики, применяемой к технологиям в целом и к «умным» машинам в частности, становятся важными направлениями внутренней и внешней политики государств. Высокие технологии, основанные на прикладном применении алгоритмов ИИ, вносят изменения в целые отрасли. Зачастую эти изменения носят революционный, «взрывной» характер, что отражено в использовании для их влияния английского слова «disrupt», которое можно перевести как «нарушение, подрыв, слом». Так, например, «медицина вообще и радиология в частности претерпевает значительные изменения, связанные с широким внедрением информатики, машинного обучения, биобанкинга и персонализированной медицины» [11, с. 3]. Нельзя отрицать и потенциал влияния алгоритмов на основе ИИ на общественное мнение и принятие политических решений: то, что таким серьезным образом может повлиять на общество и его наиболее значимые сферы, не может оставаться вне области государственного регулирования.

Среди существующих документов и концепций, отражающих этические нормы работы с технологиями, можно встретить несколько универсальных принципов, получивших свое распространение в юридических документах, научных и

публицистических работах по теме, а также локальных (в том числе внутрикорпоративных) этических формулярах:

- приватность;
- подотчетность;
- безопасность и защита;
- прозрачность и объяснимость;
- справедливость и отсутствие дискриминации;
- человеческий контроль над технологией;
- профессиональная ответственность;
- соблюдение и развитие человеческих ценностей.

Значимыми акторами сферы этического регулирования технологий на базе ИИ выступают пользователи электронных систем, их разработчики, компании-производители, отраслевые ассоциации, государства, международные объединения и союзы, люди, не являющиеся активными пользователями систем. Все они в той или иной степени или активно развивают окружающую их цифровую среду, или испытывают ее влияние, или находятся под ее давлением.

По итогам аналитической оценки текущего нормативного, этического и иного регулирования сферы информационных технологий на основе ИИ, могут быть выделены следующие общественно-значимые аспекты, приобретающие политическое и общемировое значение в контексте развития технологий и обществ:

1. «Локомотивы и балласт».

Данный аспект иллюстрирует высокую скорость принятия мер по совершенствованию нормативного регулирования и высокоразвитую техническую оснащенность одних стран в сравнении с другими.

Чтобы проанализировать данный аспект, перечислим страны, уже располагающие собственной национальной стратегией по развитию ИИ, которая вписана в общую канву государственной нормативной системы: Россия, США, Китай, Европейский Союз (общая стратегия блока и отдельные стратегии стран-участниц: Германия, Франция, Италия, Испания, Финляндия, Швеция, Дания, Литва), Канада, Япония, Южная Корея, Сингапур, ОАЭ, Тайвань, Турция, Великобритания, Мексика, Индия, Кения, Австралия (бюджетные документы по ИИ).

Из формулировок стратегий можно сделать следующие выводы. Крупнейшие экономики, такие как США и Китай, делают ставку на собственном лидерстве в общемировом процессе развития области ИИ, а также на национальной безопасности,

эффективной защите данных и укреплении суверенитета страны. США подробнее раскрывает вопрос развития кадрового потенциала, тогда как Китай фиксирует значимость наличия на рынке программного обеспечения с открытым исходным кодом. Для Европейского союза наиболее значимым аспектом стала нормативная составляющая: регулирование работы систем, выработка единых правил и принципов. Стратегия России апеллирует к общечеловеческим принципам морали и права: защита свобод человека, безопасность, прозрачность, технологический суверенитет, целостность инновационного цикла, разумная бережливость и др.

По итогу анализа особенностей названных стран и их стратегических документов, можно признать, что за исключением нескольких государств, для которых основной целью написания стратегических документов в сфере развития ИИ является привлечение внимания мировой общественности (например, Мексика, Кения), все остальные могут быть охарактеризованы как страны первого и второго мира, находящиеся в статусе развитых и активно развивающихся экономик. Из этого можно сделать вывод, что страны, имеющие нормативное и техническое преимущество, получают возможность многократно развить свои текущие наработки в области разработки различных алгоритмов, беспрепятственно применять их в своей практике, экспортировать решения на внешние рынки, получив преимущество в доступе к ресурсам, кадрам и потенциалу, связанным с разработкой и применением соответствующих технологий. Государства, расположенные на другом конце этой условной шкалы и выступающие импортерами нормативных инициатив и технологических разработок стран-лидеров, находятся в статусе пассивных пользователей без возможности активно формировать свою собственную технологическую и законодательную повестку.

Одно из наиболее важных факторов преимуществ стран-лидеров перед странами-потребителями – кадровый потенциал. На его развитие оказывают влияние все факторы технологического превосходства: от наличия более совершенной системы образования до распространенности высокотехнологичных производств и доступа к электронным ресурсам. В этом смысле для сферы высоких технологий справедлив ранее преимущественно экономический тезис «богатые богатеют, бедные беднеют»: уже сегодня можно говорить о труднопреодолимом технологическом отставании стран с менее развитыми образовательными институтами и технологическими возможностями, а на горизонте 2027-2030 гг. это отставание фактически превратится в барьер.

С точки зрения человечества как вида и цивилизации это, может, и не носит катастрофического характера, и не наносит физический урон странам третьего мира,

однако и не способствует их развитию, преодолению социально-экономических проблем и выходу на новый уровень благосостояния. В некоторой степени текущее положение вещей созвучно раннеколониальному периоду в истории Европы, когда одни народы имели абсолютное преимущество перед другими на основании непреодолимого технологического барьера в виде более развитого военного дела, наличия огнестрельного оружия, продвинутого судостроения, средств навигации и т.п.

2. «Корпоративный киберпанк».

Накопленный опыт применения технологий на основе ИИ позволяет говорить, что этические правила, закладываемые в технологии, протоколы работы алгоритмов и электронных систем носят социально-значимый характер. Они влияют на повседневную жизнь людей, чьи будни связаны с использованием технологий, и здесь интересы хозяйствующих субъектов и государства пересекаются, вынуждая национальные правительства вырабатывать собственный независимый взгляд на технологии и прогресс. При этом государственной машине требуется время на соотнесение практики применения любых правил и инициатив с реальностью их применения, тогда как скорость развития технологий, напротив, диктует ускоренный темп выработки этих самых правил. В связи с этим, компании, напрямую заинтересованные в выработке правил, становятся двигателями процессов их создания, формализации и внедрения, так как отсутствие таковых не дает возможности выводить товары на рынок, а следовательно, получать прибыли от реализации.

В каком-то смысле изменения на государственном уровне «отстают» от темпов коммерциализации технологий и их внедрения во внутрикорпоративные процессы. Корпорации получают обусловленную объективными рыночными причинами возможность предлагать, обосновывать и лоббировать собственные трактовки этических правил и закреплять их как в сознании людей, так и в появляющемся впоследствии законодательстве. В данном случае срабатывает «эффект первого», что приводит к закреплению первоначальных идей, а не более поздних регламентов.

Менять устоявшиеся и укоренившиеся в сознании людей установки непросто даже для государства, поэтому высока вероятность того, что текущие версии и формулировки этических правил работы электронных систем на основе ИИ будут перенесены и в национальные законодательства. Примером такого укоренения может служить знаменитый «принцип QWERTY», когда не самая эффективная с точки зрения скорости набора знаков раскладка клавиатуры для ЭВМ была зафиксирована на значительном числе электронных устройств ввода и получила наиболее широкое

распространение несмотря на то, что существуют и более эффективные с точки зрения эргономики раскладки [5].

Тот же эффект может произойти и с выработкой правил, которые в конечном счете получают закрепление в законодательстве и окажутся негибкими к дальнейшим изменениям. Компании становятся главными интересантами разработки и внедрения правил, а значит, прикладывают максимум усилий для сохранения статус-кво в качестве лидеров мнений. Речь идет о таких глобальных мировых корпорациях-гигантах как: Amazon, Google, Facebook, Apple, Microsoft.

В России данное направление влияния корпоративного сектора на экономику в целом представлено такими компаниями как: Сбербанк, Яндекс, Mail.ru. Также можно обозначить, что влияние имеют и компании, не имеющие четкой национальной принадлежности, такие как Telegram.

В деле развития технологических компаний вопрос национальной принадлежности перестает носить столь значимый характер; «транснациональность» как основа глобального влияния таких компаний диктует особенности взаимодействия государств с ними. Так, на протяжении 10 лет с 2005 г. Правительство США вело дело против компании Google, направленное на внедрение ограничений и регулирование процесса создания всемирной библиотеки Google Books [7]. В это же время компания Amazon вынуждена пересмотреть собственные инновации на складах, так как скорость работы автоматизированного направления логистики значительно превышает предельную возможную скорость работы человека, что при пересечении автоматизированного и ручного подразделения влечет за собой рост усталости и травм персонала [6]. Сфера защиты данных регулярно страдает от недостаточного регулирования, например, в СМИ регулярно освещаются утечки данных крупнейшего банка России Сбербанка [2] или главной мировой социальной сети Facebook [10].

Особенную сферу сегодня представляет направление беспилотных автомобилей. Алгоритмы, построенные на основе имеющихся правил дорожного движения, с появлением «беспилотников» нуждаются в пересмотре. Возникает проблема разграничения ответственности водителя и ответственности разработчика алгоритма автовождения, что неминуемо порождает юридические коллизии в случае наступления нежелательного события (например, аварии по причине столкновения с другим участником движения, потери управления транспортным средством, сбоя в работе электроники и т.д.). То же касается и проблемы определения диагноза системой поддержки принятия врачебных решений, где ответственность за наступление

нежелательного события (например, неизлечение, инвалидность, смерть пациента и т.д.) так же может быть разделена между врачом, разработчиком и клиникой. В данном примере речь пока не идет о наделении технического устройства юридической правоспособностью; сегодня такое развитие событий не имеет смысла в силу относительного несовершенства работы систем. Человек (или люди, если речь идет о коллективе) по-прежнему остается единственным ответственным за работу систем.

3. «Золотой миллиард» технологий.

Качество кадров – один из основных факторов технологического превосходства одних стран над другими и один из самых сложных с точки зрения быстрого преодоления разрыва. Люди, с раннего детства имеющие доступ к технологиям и образовательным возможностям, неминуемо получают существенный перевес в возможностях карьерного роста, создании и развитии собственных технологических компаний, нахождении в сообществе таких же людей: с такими же возможностями, потенциалом и путями его реализации. Это формирует своего рода барьер и устойчивый пласт «среднего класса», обладающий сравнительно более высокими доходами и доступом к актуальным знаниям. Таким странам как Китай и Индия приходится труднее, потому как процент населения, относящегося к «среднему классу» в процентном соотношении к общему числу населения еще не так велик [3], хотя в абсолютных числах достигает цифры в 200 млн. человек.

Кроме того, важно учитывать, что само понятие «среднего класса» достаточно относительно: оно напрямую связано с общим уровнем благосостояния страны и ее места в мировой экономике, и нельзя рассматривать данную категорию в отрыве от остальных. Из этого следует, что более технологически развитая страна имеет больше возможностей возвращать, привлекать и удерживать ценные с технологической точки зрения кадры: таланты остаются в «системе» и развивают ее еще больше, таким образом усугубляя разрыв в кадровом потенциале между странами. Любой человек, стремящийся к личному благополучию, выбирает более технологичное, чистое и понятное будущее, где он может получить безопасность, медицинское обслуживание, образование, работу и интересный досуг. Люди, которые сознательно или ввиду жизненной ситуации отказываются от этого пути, остаются «вне системы», а значит не получают доступа к ее благам и не могут выбраться из порочного круга низкоквалифицированного труда в менее благоприятной среде ни индивидуально, ни на уровне общности. В каком-то смысле подобная ситуация схожа с концепциями «миров» в политической истории человечества: возможности на «лучшую жизнь» концентрировались в центрах Рах

Romana, Pax Britannica и Pax Americana. В меньшей, хотя также наглядной степени эти процессы воспроизведены в столицах и областных центрах государств с неравномерным уровнем социально-экономического развития. Подобная ситуация труднопреодолима естественным путем, что побуждает мировое сообщество уделять внимание программам глобального образования и стипендий для студентов из развивающихся стран, а сами эти страны развивать инициативы по возвращению подготовленных за рубежом кадров в страну, а в худшем случае – использовать административные запреты для принуждения квалифицированных кадров к работе в интересах национальной экономики.

Полагаем важным отдельно сказать о российской стратегии. В ноябре 2019 г. 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» [9]. Ключевыми технологиями данного направления развития признаются: компьютерное зрение, обработка естественного языка, распознавание и синтез речи, рекомендательные системы и интеллектуальные системы поддержки принятия решений, автоматическое машинное обучение, автономные программно-ориентированные интеллектуальные системы, автоматический дизайн физических объектов, алгоритмы обучения, функционирующие на основе данных с частичной разметкой, математические алгоритмы вычислений на основе новых типов вычислительных систем, интерпретируемые алгоритмы ИИ. Подчеркивается значимость бюджетного финансирования и участия бизнеса в рамках реализации государственно-частного партнерства. Явно упоминается значимость мирового лидерства в развитии ИИ, в то же время предполагается обеспечение защиты собственных технологий от внешнего вмешательства, защита суверенитета данных. Упоминается вопрос кадрового резерва, обучения, переквалификации и стимулирования специалистов.

Считаем важным подчеркнуть, что проблема кадрового потенциала и так называемая «утечка мозгов» не минула и нашу страну. Поэтому одним из важнейших приоритетов развития отрасли ИИ как одной из наиболее быстроразвивающихся и молодых с точки зрения активных участников должна быть работа с кадровым потенциалом, его стимулирование и развитие, выдерживание конкуренции перед зарубежными предложениями на рынке труда.

Говоря о юридической стороне регулирования отрасли, на основании приведенных выше 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). Примером подобного документа может служить действующая Конвенция о дорожном движении от 08 ноября 1968 г. [4], разработанная Экономическим и Социальным Советом ООН (ЭКОСОС).

Вопрос выработки универсальных этических принципов работы технологий и с технологиями значим в контексте стремительного развития алгоритмов на основе искусственного интеллекта. Высокий уровень общественной значимости вопроса выводит его на политический уровень, и можно говорить, что этика технологий приобретает общеполитическую значимость. Вопрос остро нуждается в детальной проработке, и мировому сообществу в лице государств, общественности и бизнеса еще предстоит объединить усилия в выработке единого взгляда на этические нормы работы алгоритмов искусственного интеллекта.

В общемировом информационном поле еще нет общепризнанного лидера мнений, задающего повестку развития темы. Можно выделить государства-лидеры и компании-локомотивы [12, с. 15], 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.

В заключение считаем важным добавить, что тема нуждается в дальнейшем изучении и популяризации в среде технических специалистов, которые имеют непосредственное отношение к технологиям и глубоко погружены в проблему.

Библиографический список:

1. Аналитическая философия: Становление и развитие (антология). Пер. с англ., нем. / Под ред. А.Ф. Грязнова. М.: «Дом интеллектуальной книги», «Прогресс-Традиция», 1998. С. 377-400.
2. В Сети продается биометрия и данные о кредитах клиентов Сбербанка [Электронный ресурс]. Режим доступа: https://www.cnews.ru/news/top/2019-10-24_sberbank_slil_v_set_biometriyu (Дата обращения: 26.04.2021).
3. Китай на пути к среднему классу [Электронный ресурс]. Режим доступа: <https://regnum.ru/news/economy/2981019.html> (Дата обращения: 26.04.2021).
4. Конвенция о дорожном движении (Вена, 08 ноября 1968 года).

5. Почему именно QWERTY? И за что нам такой ЙЦУКЕН [Электронный ресурс]. Режим доступа: <https://www.bbc.com/russian/vert-cul-38730675> (Дата обращения: 26.04.2021).
6. Роботизация складов Amazon привела к росту травм на 50% и усилению нагрузки на людей [Электронный ресурс]. Режим доступа: <https://vc.ru/story/163571-robotizaciya-skladov-amazon-privela-k-rostu-travm-na-50-i-usileniyu-nagruzki-na-lyudey> (Дата обращения: 26.04.2021).
7. Суд разрешил Google продолжить оцифровывать книги [Электронный ресурс]. Режим доступа: https://www.bbc.com/russian/society/2015/10/151016_google_scanning_books_court (Дата обращения: 26.04.2021).
8. Тьюринг А. Может ли машина мыслить? / Под ред. С.А. Яновской. М.: Государственное издательство физико-математической литературы, 1960.
9. Указ Президента РФ от 10 октября 2019 г. № 490 «О развитии искусственного интеллекта в Российской Федерации».
10. Facebook под суд: 5 резонансных исков против крупнейшей соцсети планеты [Электронный ресурс]. Режим доступа: <https://www.bbc.com/russian/features-45904994> (Дата обращения: 26.04.2021).
11. Artificial Intelligence in Medical Imaging: opportunities, applications and risks / **Ranschaert Erik R., Morozov S., Algra Paul R.** (Eds.). Springer International Publishing. 2019. 373 p.
12. 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 [Electronic resource] // Berkman Klein Center for Internet & Society at Harvard Law School. 2020. № 1. URL: <https://walton.uark.edu/business-integrity/images/PrincipledAIHarvard2020.pdf> (Access: 26.04.2021).

Demchenko O.Y. **Ethics and Artificial Intelligence Technologies: political aspect.**

Technological progress, in particular the development of AI-based technologies, underlines the importance of understanding the ethics of Human-AI interactions. The conceptualization and development of ethical principles to serve as the basis for universally accepted policies regulating advanced technologies and AI-based tools becomes an important task. Due to the nature of market economy and the increasing need for internal rules and regulations, the private sector tends to assume the role of the creators and end users of ethical rules, becoming the key stakeholder in the process of developing ethics-related rulesets and frameworks for AI-based technologies. The situation calls for broad cooperation between governments, communities and business aimed to process and formulate the common understanding of ethical principles regulating machine-to-machine and machine-to-human interactions. The document describes various national, international and corporate rulesets created as a possible framework for further development of policies regulating AI-based technologies, as well as current trends in the political thought concerning the concept of ethics and technology.

Keywords: technology ethics, artificial intelligence, technical regulation law, technological politics, science politics