

Мажей Ярослав Владимирович, студент бакалавриата 2 курс, МИРЭА-

Российский технологический университет (РТУ МИРЭА),

Институт информационных технологий

Свищёв Андрей Владимирович, старший преподаватель кафедры

практической и прикладной информатики

МИРЭА-Российский технологический университет (РТУ МИРЭА),

Институт информационных технологий, Россия, г. Москва

ЭТИКА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

Аннотация: Всё чаще в мире поднимается проблема внедрения искусственного интеллекта в повседневную жизнь. В данной статье будет описано, как разные компании применяют данную технологию, как она взаимодействует с людьми, что будет в случае внедрения, какие нюансы необходимо учитывать.

Ключевые слова: этика, информационные технологии, искусственный интеллект (ИИ).

Abstract: Increasingly, the question of introducing artificial intelligence into everyday life is being raised around the world. This article will talk about how different companies are applying this technology, how it interacts with people, what will happen in case of implementation, and what nuances need to be considered.

Key words: ethics, information technologies, artificial intelligence (AI).

18 марта 2018 года, 10 часов вечера. Элейн Херцберг возвращалась домой на велосипеде, внезапно её насмерть сбивает автомобиль. Транспортom управлял искусственный интеллект [3].

7 мая 2016 года Джошуа Браун ехал домой в своей машине. Фургонem

управлялись искусственный интеллект, но он врезается в фуру и человек погибает [2].

Как и другие инциденты, связанные с взаимодействием людей и технологий искусственного интеллекта, эти происшествия заставили людей задуматься об этических и правовых проблемах искусственного интеллекта. Какие этические обязательства должны выполнять системные программисты, чтобы их продукты исследований и разработок не подвергали опасности человеческие жизни? Кто несет ответственность за смерть Герцберга и Брауна? Это отдел тестирования компании по производству самоуправляемых автомобилей? Разработчик систем искусственного интеллекта или даже производитель бортового зондирующего оборудования?

Этические дискуссии об искусственном интеллекте продолжаются. С самого начала исследований в области искусственного интеллекта основное внимание уделялось теоретической работе, в которой обсуждались возможности и будущие последствия, но меньше исследований и дискуссий было посвящено практическому применению искусственного интеллекта. Хотя академическое сообщество десятилетиями обсуждало взаимосвязь между этикой и этикой искусственного интеллекта, оно так и не пришло к выводу, что такое общая этика искусственного интеллекта или даже как ее следует определять и называть. Единой стандартизации не существует. В последние годы, в связи с непрерывным развитием социальных наук и технологий, были сделаны крупные прорывы в развитии искусственного интеллекта. Этические исследования и дискуссии, связанные с искусственным интеллектом, становятся все более и более масштабными, влияя на нашу жизнь.

Искусственный интеллект и его обсуждение, и анализ на примере "Искусственный интеллект" разработан как система, которая получает факторы из окружающей среды и решает проблемы, оценивает риски, делает прогнозы и предпринимает действия на основе этих внешних входных данных. До эры мощных компьютеров и больших данных такого рода системы были реализованы людьми посредством определенного программирования и

сочетания с определенными правилами. С непрерывным прогрессом науки и техники продолжают появляться новые методы. Одним из них является машинное обучение, которое в настоящее время является наиболее активной и популярной областью искусственного интеллекта. Метод применения статистики позволяет системе “учиться” на данных и принимать решения. Обращая внимание на технический прогресс, мы больше озабочены этическими вопросами в экстремальных ситуациях. Например, использование технологии искусственного интеллекта в некоторых смертоносных военных беспилотных летательных аппаратах или риск того, что технология искусственного интеллекта может привести к краху мировой финансовой системы и так далее.

Для обобщения и анализа большого объема данных мы можем использовать технологию искусственного интеллекта, чтобы помочь проанализировать доверие к заявителям на получение кредита, решить, предоставлять ли кредиты и какую сумму денег, а также оценить кандидатов, решить, следует ли их принимать, и спрогнозировать вероятность повторного совершения преступлений преступниками и так далее. Эти технологические изменения глубоко повлияли на общество и изменили жизнь людей. Однако применение таких технологий также вызовет некоторые тревожные этические проблемы, поскольку системы искусственного интеллекта расширяют знания, полученные ими из реальных данных, и даже усилят расовые и гендерные предубеждения. Поэтому, сталкиваясь с неизвестными сценариями, система также будет делать неправильные суждения. Более того, поскольку многие такие системы являются “черными ящиками”, людям часто бывает трудно понять внутренние причины суждений системы, поэтому их трудно подвергать сомнению или исследовать, что создает риски для принятия решений людьми. Приведу несколько конкретных примеров: В 2014 году Amazon разработала инструмент подбора персонала для выявления набранных инженеров-программистов, но система совершала одну критическую ошибку. Это связано с тем, что компьютерные модели Amazon были обучены проверять кандидатов, ориентируясь на 10-летний опыт компании в отборе. Загвоздка была в том, что

большинство кандидатов прошлых лет были мужчинами, что вполне отражало мужское доминирование в технологической отрасли. Из-за этого система Amazon сама для себя выработала алгоритм, предпочитающий кандидатов-мужчин. Она также автоматически отвергала резюме, где вообще фигурировало слово «женский», например, в словосочетании «капитан женского шахматного клуба» [1].

В 2016 году ProPublica провела анализ коммерчески разработанной системы, которая прогнозирует вероятность повторного совершения преступлений преступниками и направлена на то, чтобы помочь судьям принимать более эффективные решения о вынесении приговоров. Также было установлено, что система дискриминировала темнокожих. Предвзятость. В последние два года автономные транспортные средства полагались на установленные правила и учебные данные для обучения, но, сталкиваясь с незнакомыми сценариями или входными данными, которые их системы не могут распознать, они не могут принимать правильные решения, что приводит к несчастным случаям со смертельным исходом [4].

Поскольку эти системы считаются интеллектуальной собственностью, частный коммерческий разработчик обычно отказывается предоставлять свой код для ознакомления. В то же время технологические достижения сами по себе не могут решить фундаментальную проблему, лежащую в основе ИИ — хорошо продуманные алгоритмы также должны принимать решения на основе конкретных данных реального мира. Однако эти входные данные будут ошибочными, несовершенными и непредсказуемыми.

Программисты быстрее других поняли, что после проектирования системы невозможно всегда решать эти проблемы впоследствии. Чем больше людей осознают, что этические вопросы следует рассматривать как проблему, которую необходимо рассмотреть перед развертыванием системы.

Обсуждение и размышления о безработице и дисбалансе:

Важные моральные и этические проблемы искусственного интеллекта находятся не только на переднем крае социальных рисков, но и на переднем крае

социального прогресса. Мы обсуждаем два нерешенных вопроса: безработицу и дисбаланс.

Безработица. В течение десятилетий, чтобы высвободить человеческий труд, мы производили машины, которые имитируют людей и позволяют машинам заменять нас для более эффективного выполнения повседневных задач. С быстрым развитием экономики и ростом степени автоматизации в нашей жизни появилось большое количество новых изобретений, которые делают нашу жизнь быстрее и проще. Когда мы используем роботов, чтобы заменить нас, людей, для выполнения задач, то есть для автоматизации ручных задач, мы высвобождаем ресурсы для создания более сложных ролей, связанных с когнитивным, а не ручным трудом. Вот почему уровень рабочей силы зависит от того, может ли работа быть автоматизирована (например, университетские профессора зарабатывают больше, чем сантехники). Согласно недавнему отчету McKinsey & Company, к 2030 году по мере ускорения глобальной автоматизации исчезнет почти 800 миллионов рабочих мест. Например, с появлением автономных систем вождения, технология искусственного интеллекта вызвала у людей беспокойство по поводу безработицы, и большое количество рабочих мест водителей грузовиков может оказаться под угрозой. Впервые в истории мы, люди, начнем конкурировать с машинами на когнитивном уровне. Самое страшное, что они обладают более мощными способностями, чем мы. Есть также некоторые экономисты, которые беспокоятся о том, что, будучи людьми, мы не сможем адаптироваться к такому обществу и в конечном итоге отстанем от машин [5].

Дисбаланс. Представьте, что будет в будущем без работы? Нынешняя экономическая структура общества очень проста: компенсация в обмен на взносы. Компания выплачивает зарплату своим сотрудникам в зависимости от определенного объема работы. Но с помощью технологий искусственного интеллекта компании могут значительно сократить свои человеческие ресурсы. Следовательно, его общий доход будет поступать меньшему количеству людей. Для компаний, которые широко используют новые технологии, небольшое

число людей будет получать более высокую долю заработной платы, что приведет к увеличению разрыва между богатыми и бедными. В 2008 году Microsoft была единственной технологической компанией, вошедшей в десятку самых ценных компаний мира. Apple заняла второе место, а Google – третье [6]. Однако к 2020 году в первую пятерку десяти самых ценных компаний мира вошли все американские технологические компании [7].

В современном мире Кремниевая долина продвигает экономику "победитель получает все", и независимая компания часто занимает большую часть рынка. Поэтому из-за сложности доступа к данным стартапам и небольшим компаниям трудно конкурировать с такими компаниями, как Alphabet и Facebook (больше пользователей = больше данных, больше данных = лучший сервис, лучший сервис = больше пользователей). Также обнаружен феномен, заключающийся в том, что эти технологические гиганты часто создают гораздо меньше рабочих мест, чем другие компании на рынке. Например, в 1990 году рыночная капитализация трех крупнейших компаний Детройта достигла 65 миллиардов долларов и насчитывала 1,2 миллиона работников. В 2016 году стоимость трех крупнейших компаний Кремниевой долины составляла 11,5 трлн долларов США, но в них работало всего 190 000 сотрудников. Итак, как выживут сегодня работники с избыточными навыками, вызовет ли эта тенденция социальные беспорядки и должны ли технологические гиганты брать на себя больше социальных обязанностей - все это вопросы, над которыми стоит задуматься.

Предложения по этическим вопросам искусственного интеллекта. Как видно из вышесказанного, недостаточная осведомленность об этике приведет к определенным рискам для общества и человеческой жизни. Поэтому, чтобы усилить правильную руководящую роль этических факторов ИИ в практических приложениях, мы должны начать со следующих аспектов:

Четко определить этическое поведение. Исследователи ИИ и специалисты по этике должны выражать этические ценности в виде количественных параметров. Другими словами, они должны предоставить машине четкие

ответы и правила принятия решений для любых потенциальных моральных дилемм, с которыми она может столкнуться. Это потребует от человечества достижения консенсуса относительно наиболее этичного образа действий в любой конкретной ситуации, что является сложной, но не невыполнимой задачей. Например, Немецкий комитет по этике автономного вождения и подключенного вождения предложил: «Рекомендуется программировать этические ценности в автономных транспортных средствах с приоритетом защиты человеческой жизни в качестве главного приоритета». В случае неизбежной смертельной аварии автомобиль не должен выбирать, убивать ли человека, основываясь на личных характеристиках, таких как возраст, пол, физическое или психологическое строение.

Краудсорсинг человеческой этики. Инженерам необходимо собрать достаточное количество данных о четких этических стандартах, чтобы должным образом обучить алгоритмы искусственного интеллекта. Даже после определения конкретных показателей этических ценностей, если для обучения модели недостаточно объективных данных, то выбрать систему искусственного интеллекта все равно может быть сложно. Получение соответствующих данных является сложной задачей, поскольку этика не всегда может быть четко стандартизирована. Разные ситуации требуют разных подходов, и в некоторых случаях может вообще не быть единого этического курса действий. Один из способов решения этой проблемы - собрать и упаковать потенциальные решения моральных и этических дилемм миллионов людей. Например, проект в Массачусетском технологическом институте показывает, как использовать краудсорсинговые данные в контексте автономных транспортных средств для эффективного обучения машин принятию более правильных этических решений. Однако результаты исследований также показывают, что могут существовать сильные межкультурные различия в глобальных этических ценностях, и этот фактор также следует учитывать при проектировании систем искусственного интеллекта, ориентированных на человека.

Сделать системы искусственного интеллекта более прозрачными.

Директивным органам необходимо внедрить руководящие принципы для того, чтобы решения ИИ по этике, особенно по этическим показателям и результатам, были более прозрачными. Если система искусственного интеллекта допустит ошибку или будет иметь нежелательные последствия, мы не сможем принять “алгоритм сделал это” в качестве оправдания. Но мы также знаем, что требование полной прозрачности алгоритма технически не очень полезно. Инженерам следует подумать о том, как количественно оценить этические ценности, прежде чем программировать их, и рассмотреть результаты использования этих технологий искусственного интеллекта.

Например, для самоуправляемых автомобилей это может означать постоянное ведение подробного журнала всех автоматизированных решений для обеспечения их этической ответственности.

В качестве заключения:

Возникновение этических проблем является неизбежным требованием для развития инженерной деятельности. Современная инженерная деятельность, основанная на технологиях искусственного интеллекта, становится все более сложной, а ее влияние на природу и общество становится все более глубоким. В то же время, как ключевая роль в инженерной деятельности, группа инженеров обладает способностью в определенном смысле изменить мир.

Библиографический список:

1. The Bell //Интернет-издание, Режим доступа: свободный [Электронный ресурс] - URL - <https://thebell.io/amazon-sozdala-robotu-dlya-otbora-kadrov-on-okazalsya-seksistom> (Дата обращения 28.09.2021).

2. Meduza //Интернет-издание, Режим доступа: свободный [Электронный ресурс] - URL - <https://meduza.io/feature/2016/07/01/pervaya-avariya-avtopilota-tesla-glavnoe> (Дата обращения 28.09.2021).

3. Ведомости //Информационный ресурс, Режим доступа: свободный [Электронный ресурс] – URL -

<https://www.vedomosti.ru/technology/news/2018/05/08/768916-bespilotnik-uber-sbil>
(Дата обращения 28.09.2021).

4. BBC News Русская служба //Информационный ресурс, Режим доступа: свободный [Электронный ресурс] - URL - <https://www.bbc.com/russian/features-39879902> (Дата обращения 29.09.2021).

5. McKinsey & Company //Информационный ресурс, Режим доступа: свободный [Электронный ресурс] - URL - <https://www.mckinsey.com/featured-insights/future-of-work/jobs-lost-jobs-gained-what-the-future-of-work-will-mean-for-jobs-skills-and-wages> (Дата обращения 29.09.2021).

6. Wikipedia //Энциклопедия, Режим доступа: свободный [Электронный ресурс] – URL - https://ru.wikipedia.org/wiki/Список_самых_дорогих_брендов_мира (Дата обращения 30.09.2021).

7. Forbes //Онлайн журнал, Режим доступа: свободный [Электронный ресурс] – URL - <https://www.forbes.ru/biznes/405947-samye-dorogie-brendy-mira-2020-goda-po-versii-forbes> (Дата обращения 30.09.2021).