УДК 004.8

Милкова Эрика Геннадьевна
Преподаватель, кандидат экономических наук
Финансовый Университет при Правительстве РФ

Milkova Erika Gennadievna
Lecturer, Candidate of Economic Sciences
Financial University under the Government of the Russian Federation

СОЦИАЛЬНАЯ ЦЕНА ИСКУССТВЕННОГО ИНТЕЛЛЕКТА: ЭТИКА, КОНФИДЕНЦИАЛЬНОСТЬ ДАННЫХ И ДРУГИЕ ИЗДЕРЖКИ

THE SOCIAL PRICE OF ARTIFICIAL INTELLIGENCE: ETHICS, DATA PRIVACY AND OTHER COSTS

Аннотация: Этика искусственного интеллекта (ИИ) связана с вопросом о том, как должны вести себя разработчики, производители и операторы технологий, чтобы минимизировать негативные социальные последствия имплементации технологий ИИ в обществе. Сфера применения этики ИИ охватывает непосредственные, здесь и сейчас, проблемы, связанные, например, с конфиденциальностью данных и предвзятостью в современных системах ИИ; краткосрочные и среднесрочные проблемы, связанные, например, с влиянием ИИ и робототехники на рабочие места; и более долгосрочные опасения по поводу возможности достижения или превышения системами искусственного интеллекта эквивалентных человеческим возможностей. Растущее повсеместное использование смартфонов и управляемых ИИ приложений, на которые многие из нас теперь полагаются каждый день, тот факт, что ИИ все больше влияет на все сферы общественной жизни (включая промышленность, здравоохранение, судебную систему, транспорт, финансы и досуг), а также

кажущаяся перспектива «гонки вооружений» ИИ вызвали большое число национальных и международных инициатив со стороны НПО, академических и промышленных групп, профессиональных органов и правительств. Эти инициативы привели к публикации различных сборников этических принципов для робототехники и искусственного интеллекта (по меньшей мере 22 различных набора этических принципов были опубликованы с января 2017 года), появляются новые этические стандарты (в частности, от Британского института стандартов и Ассоциации стандартов IEEE), и все большее число стран объявили об имплементации стратегий, связанных с использованием ИИ, включая вложение крупномасштабных инвестиций, и создали национальные консультативные или политические органы. [4] [8]

Ключевые слова: искусственный интеллект, использование данных, этические принципы, неприкосновенность личных данных, машинное обучение.

Annotation: Artificial intelligence (AI) ethics is related to the issue of how developers, producers, and operators of the technologies should behave to minimize the negative social consequences of AI technology implementation in society. The scope of AI implementation includes the issues that arise right at this particular moment regarding various problems, related, for example, to data privacy and bias in modern AI systems; short and medium-term issues such as the influence of AI and robotics on employment; and also, long-term concern about the possibility of reaching and exceeding human abilities with help of AI. The widespread resort of smartphones and other gadgets with AI systems, which many of relying on nowadays, also the fact that AI has a growing influence on many areas of modern society, including manufacturing, healthcare, judicial system, transport, finance and leisure activities, and the possibility of perspective "armament drive" caused a great deal of national and international initiatives from NGOs, political and manufacturing groups and governments. These initiatives led to the publication of various sets of ethical principles for robotics and AI systems (at least 22 sets of them have been published since 2017), there are

also new ethical standards are being created, for example, one by IEEE, and the bigger number of counties announce the beginning of strategies implementation related to AI usage, including large scale financing and creation of national consultative or political organs.
Keywords: artificial intelligence, data usage, ethical principles, privacy, machine learning.

Люди веками были обеспокоены вытеснением рабочих технологиями. Изначально автоматизация, а затем и механизация, вычислительная техника, а в последнее время ИИ и робототехника должны были нанести необратимый ущерб рынку труда. Однако в прошлом автоматизация часто заменяла человеческий труд в краткосрочной перспективе, но приводила к созданию рабочих мест в долгосрочной. [1] Тем не менее, существует широко распространенное опасение, что ИИ и связанные с ним технологии могут создать массовую безработицу в течение следующих двух десятилетий. В одном из недавних документов был сделан вывод о том, что новые информационные технологии в ближайшем будущем поставят под угрозу «значительную долю занятости во многих профессиях» (Frey and Osborne, 2013). [7]

ИИ уже широко распространен в сфере финансов, космических исследований, передового производства, транспорта, развития энергетики и здравоохранения. Беспилотные транспортные средства и автономные дроны также выполняют функции, которые ранее требовали вмешательства человека. Автоматизация уже оказала влияние на рабочие места «синих воротничков»; однако по мере того, как компьютеры становятся все более сложными, и выполняют все более универсальные функции, все больше рабочих мест будет зависеть от развития технологий и все больше профессий будет устаревать.

В течение следующего десятилетия ИИ окажет глубокое воздействие на частную жизнь. Конфиденциальность и достоинство пользователей ИИ

должны тщательно учитываться при разработке роботов-помощников, предназначенных для работы в домах людей, поскольку такая работа означает, что они будут посвящены в очень личные моменты, такие как купание и одевание. Однако другие аспекты ИИ также будут влиять на конфиденциальность. Gрезидент Microsoft, недавно заметил: «Технология Intelligent 3 поднимает вопросы, которые лежат в основе фундаментальной защиты прав человека, таких как право на неприкосновенность частной жизни и свобода слова. Эти вопросы повышают ответственность компаний, которые создают эти технологии. Они также требуют вдумчивого государственного регулирования и разработки норм, касающихся приемлемых видов использования технологий». [6]

Еще один аспект искусственного интеллекта, влияющий на конфиденциальность, — это Big Data. Технологии сейчас находится на той стадии, когда долгосрочные записи могут храниться на всех, кто производит эти данные: электронные счета, договоры, учетные записи, истории операций, не говоря уже о любом электронном письме и использовании социальных сетей. Такие записи можно найти с помощью алгоритмов распознавания образов, что означает, что наличие анонимности по умолчанию более не является таковым. [5]

Любой может быть идентифицирован с помощью программного обеспечения для распознавания лиц или интеллектуального анализа данных о покупках или привычных действиях в социальных сетях. Эти привычки в интернете могут указывать не только отдельную личность, но и на политические или экономические предрасположенности и предпочтения. Машинное обучение позволяет извлекать информацию из данных и открывать новые закономерности, а также превращать кажущиеся безобидными данные в конфиденциальные личные данные. Например, модели пользования социальными сетями могут предсказывать личностные характеристики, политические предпочтения и даже характерные показатели уровня жизни.

Приложения ИИ, основанные на машинном обучении, нуждаются в доступе к большим объемам данных, но собственники данных имеют ограниченные права на их использование. Недавно ЕС принял новые общие правила защиты данных (GDPR) для защиты частной жизни граждан. Однако эти правила применяются только к персональным данным, а не к агрегированным «анонимным» данным, которые обычно используются для обучения моделей. [3]

Кроме того, персональные данные или информация о том, кто входил в обучающий набор, в некоторых случаях могут быть восстановлены из модели, что потенциально может иметь значительные последствия для регулирования этих систем. Например, в то время как люди имеют права на то, как их личные данные используются и хранятся, они имеют ограниченные права в отношении обученных моделей. Вместо этого, как правило, считается, что модели в основном регулируются различными правами интеллектуальной собственности, такими как коммерческая тайна. Например, в настоящее время нет никаких прав и обязанностей по защите данных в отношении моделей в период после их создания, но до принятия каких-либо решений об их использовании.

В связи с этим возникает ряд этических вопросов. Какой уровень контроля будут иметь над данными те, о ком они собираются? Должны ли люди иметь право использовать эту модель или, по крайней мере, знать, для чего она используется, учитывая их вклад в ее обучение? Могут ли системы машинного обучения, ищущие закономерности в данных, непреднамеренно нарушать конфиденциальность людей, если, например, установление последовательности генома одного члена семьи выявило медицинскую информацию о других членах семьи? Еще один этический вопрос связан с тем, как предотвратить раскрытие личности или личной информации человека, участвующего в обучении модели, например, с помощью

кибератаки. М.Вил и другие утверждают, что людям, чьи данные были использованы для подготовки моделей, следует предоставить дополнительную защиту, например право доступа к моделям; знать, откуда данные были получены и кому они передаются; право убрать данные о себе из подготовленной модели; и право выразить пожелание, чтобы эта модель не использовалась в будущем. [2]

Исследователи обнаружили, что автоматизированные средства распространения рекламы с большей вероятностью показывают публикации хорошо оплачиваемых рабочих мест среди мужчин, чем среди женщин. Например, инструмент самообучения Amazon, используемый для оценки соискателей, значительнее благоприятствует мужчинам, высоко оценивая их. Система научилась определять приоритетность заявок, в которых подчеркиваются мужские характеристики, и понижать рейтинг заявок из университетов с большим процентом женщин. ИИ создан людьми, а это значит, что он может иметь тенденцию к предвзятости. Систематическое искажение может возникать в результате использования самих данных, необходимых для обучения систем, или в результате ценностей разработчиков систем и пользователей. Чаще всего это происходит, когда приложения машинного обучения обучаются на данных, которые отражают только определенные демографические группы или отдельные социальные предубеждения и установки. [9]

Поскольку многие модели машинного обучения строятся на основе данных, генерируемых человеком, человеческие предубеждения могут легко привести к искажениям в результатах обучения данных моделей. Если разработчики не будут работать над распознанием и противодействием этим предубеждениям, приложения и продукты ИИ могут увековечить несправедливость и дискриминацию. ИИ, который предвзято относится к определенным группам общества, может иметь далеко идущие последствия.

Концентрация технологической, экономической и политической власти между несколькими мегакорпорациями может позволить им оказывать чрезмерное влияние на правительства, однако использование ИИ может угрожать демократии и другими способами: поддельные новости, подтасовка голосов на выборах, манипулирование отдельными гражданами, - это лишь немногие из аспектов, которые можно создать благодаря использованию ИИ.

Список литературы

[1] Аутор Д.Х. Почему до сих пор так много рабочих мест? История и будущее автоматизации. Журнал экономических перспектив, 2015. 29(3), 3-30

[2] Вил М., Биннс Р., Эдвардс Л. 2018. Алгоритмы, которые помнят: модель инверсионных атак и защита данных. Philisophical Transactions of the Royal Society, Physical and Engineering Sciencies. 376(2133)

[3] Общий регламент по защите данных. 2016. Режим доступа: https://gdpr-info.eu/

[4] Покровский А.В, Кашкин С.Ю., Искусственный интеллект, роботехника и защита прав человека в Европейском Союзе. Вестник Университета имени О.Е. Кутафина, 2019. Режим доступа: https://cyberleninka.ru/article/n/iskusstvennyy-intellekt-robototehnika-i-zaschita-prav-cheloveka-v-evropeyskom-soyuze

[5] Селинджер Е, Хартзог В. Пространство для будущего: философия технологий. 2017, New York: Routledge

[6] Смит Б., Технология распознавания лиц: необходимость государственного регулирования и корпоративной ответственности. Ответ Microsoft. Режим доступа:

[7] Фрэй С.Б., Осборн М.А. Будущее занятости: насколько восприимчивы рабочие места к компьютеризации? Oxford Martin Programme on the impacts of Future Technology, 2013.

[8] Храмцовская Н., ИСО: разработка международных стандартов, спецификаций и технических отчетов в области ИИ, 2019. Режим доступа: http://rusrim.blogspot.com/2019/08/

[9] Datta A., Tschantz and M.C., Datta A., Automated Experiments on Ad Privacy Settings – A Tale of Opacity, Choice, and Discrimination. Proceedings on Privacy Enhancing Technologies. 1, 92-112, DOI: 10.1515/popets-2105-0007