Военно-экономический вестник / Military Economic Bulletin https://voenvestnik.ru

2019, №3–4 / 2019, No 3–4 https://voenvestnik.ru/issue-3-4-2019.html

URL статьи: https://voenvestnik.ru/PDF/01SCVV319.pdf

DOI: 10.15862/01SCVV319 (http://dx.doi.org/10.15862/01SCVV319)

Ссылка для цитирования этой статьи:
Дьяков, В. Ф. Некоторые вопросы этики искусственного интеллекта / В. Ф. Дьяков // Военно-экономический вестник. — 2019. № 3–4. — URL: https://voenvestnik.ru/PDF/01SCVV319.pdf DOI: 10.15862/01SCVV319

Дьяков Виктор Федорович
АНО ВО «Российский новый университет», Москва, Россия
Магистрант

Некоторые вопросы этики искусственного интеллекта

Аннотация. В статье затрагивается и анализируется проблема создания, внедрения и использования алгоритмов искусственного интеллекта с точки зрения морали и этики. Перечисляется ряд ключевых этических критериев при разработке и внедрении алгоритмов искусственного интеллекта, также описываются проблемы, с которыми сталкиваются компании и разработчики при использовании систем ИИ.

Ключевые слова: искусственный интеллект; машинное обучение; этика; мораль; алгоритм

Этика искусственного интеллекта — это система моральных принципов и методов, предназначенных для информирования о разработке и ответственном использовании технологий искусственного интеллекта. Поскольку искусственный интеллект стал неотъемлемой частью продуктов и услуг, организации начинают разрабатывать кодексы этики ИИ.

Айзек Азимов, писатель-фантаст, предвидел потенциальные опасности автономных машин задолго до их разработки и создал три закона робототехники как средство ограничения этих рисков. В кодексе этики Азимова первый закон запрещает роботам причинять вред людям или допускать причинение вреда людям, бездействуя. Второй закон предписывает роботам подчиняться людям, если только эти приказы не противоречат первому закону. Третий закон предписывает роботам защищать себя в той мере, в какой это соответствует первым двум законам.

Стремительное развитие технологий искусственного интеллекта побудило экспертов, инженеров, программистов вместе с крупными компаниями и предприятиями начать разрабатывать меры предосторожности для обеспечения безопасности людей и их окружения [1].

Предприятия сталкиваются с рядом этических критериев при разработке и использовании технологий искусственного интеллекта:

- *Прозрачность*. Когда системы искусственного интеллекта выходят из строя, разработчики должны иметь возможность отслеживать сложную цепочку алгоритмических систем и процессов обработки данных, чтобы выяснить причину. Организации, использующие искусственный интеллект, должны быть в состоянии объяснить на основе полученных данных, что делают их алгоритмы и почему они это делают.

- *Ответственность*. Общество все еще распределяет ответственность, когда решения, принимаемые системами искусственного интеллекта, имеют катастрофические последствия, включая потерю капитала, здоровья или жизни. Ответственность за последствия решений, основанных на искусственном интеллекте, должна быть определена в рамках процесса, включающего юристов, регулирующие органы и граждан. Одна из проблем заключается в нахождении надлежащего баланса в тех случаях, когда система искусственного интеллекта может быть безопаснее, чем деятельность человека, которую она дублирует, но все еще вызывает проблемы, такие как взвешивание достоинств автономных систем вождения, которые приводят к гибели людей, но гораздо реже, чем люди.

- *Справедливость*. В наборах данных, содержащих личную информацию, чрезвычайно важно обеспечить отсутствие предубеждений по признаку расы, пола или этнической принадлежности.

- *Злоупотребление*. Алгоритмы искусственного интеллекта могут использоваться для целей, отличных от тех, для которых они были созданы. Вишневский сказал, что эти сценарии должны быть проанализированы на стадии проектирования, чтобы минимизировать риски и ввести меры безопасности для уменьшения неблагоприятных последствий в таких случаях [2].

Алгоритмы искусственного интеллекта играют все более важную роль в современном обществе. Будет становиться все более важным разрабатывать алгоритмы искусственного интеллекта, которые были бы не только мощными и масштабируемыми, но и прозрачными для проверки.

Представьте себе, что в ближайшем будущем банк использует алгоритм машинного обучения, чтобы одобрять заявки на ипотеку своим клиентам. Клиент, чью заявку отклонит ИИ, подает иск против банка, утверждая, что алгоритм дискриминирует заявителей на ипотеку по расовому признаку. Банк отвечает, что это невозможно, поскольку алгоритм намеренно не учитывает расу заявителей. Действительно, это было частью обоснования банка для внедрения данной системы. Тем не менее, статистика показывает, что уровень одобрения банком чернокожих кандидатов неуклонно снижается. Подача десяти одинаково квалифицированных подлинных кандидатов показывает, что алгоритм принимает белых кандидатов и отклоняет чернокожих кандидатов.

Если алгоритм машинного обучения основан на сложной нейронной сети, то может оказаться практически невозможным понять, почему или даже как алгоритм оценивает кандидатов на основе их расы. С другой стороны, программа машинного обучения, основанная на древе решений, гораздо более прозрачна для программной проверка, которая может позволить аудитору обнаружить, что алгоритм искусственного интеллекта использует адресную информацию заявителей, которые родились или ранее проживали в преимущественно бедных районах, как причину для отказа.

Прозрачность — не единственная желательная особенность искусственного интеллекта. Также важно, чтобы алгоритмы искусственного интеллекта, выполняющие социальные функции, были предсказуемы для тех, кем они управляют. Чтобы понять важность такой предсказуемости, рассмотрим аналогию. Правовой принцип прецедентного права обязывает судей следовать прошлому прецеденту, когда это возможно. Инженеру и программисту ИИ такое предпочтение прецеденту может показаться непонятным — зачем связывать будущее с прошлым, когда технологии постоянно совершенствуются. Но одна из важнейших функций правовой системы — быть предсказуемой, чтобы, например, законы можно было составлять, зная, как они будут исполняться. Задача правовой системы не обязательно состоит в том, чтобы

оптимизировать общество, а в том, чтобы обеспечить предсказуемую среду, в которой граждане могут оптимизировать свою собственную жизнь [3].

Также уделяется больше внимания устойчивости к манипуляции у алгоритмов искусственного интеллекта. Система машинного зрения для сканирования багажа авиакомпании на предмет запрещенных предметов должна быть уверено распознавать людей, намеренно ищущих уязвимые недостатки в алгоритме. Устойчивость к манипуляциям — это обычный критерий информационной безопасности и актуальное требование к системам искусственного интеллекта.

Еще одним важным социальным критерием при работе с организациями является способность найти человека, ответственного за выполнение ИИ его работы. Когда система искусственного интеллекта не справляется с поставленной задачей, должен быть тот, кто возьмет ответственность в свои руки. Современные бюрократы часто прибегают к установленным процедурам, которые распределяют ответственность настолько широко, что невозможно определить ни одного человека, виновного в катастрофах, которые в результате происходят. Даже если система искусственного интеллекта разработана с переопределением пользователя, необходимо учитывать карьерный стимул бюрократа, которого лично обвинят, если переопределение пойдет не так, и который предпочел бы обвинить ИИ в любом сложном решении с отрицательным результатом.

Ответственность, прозрачность, проверяемость, неподкупность, предсказуемость — все критерии, применимые к людям, выполняющим социальные функции, все это необходимо учитывать в алгоритме, предназначенном для замены человеческого суждения о социальных функциях [4].

Хотя на некоторые современные этических проблем уже можно найти ответ, подход алгоритмов искусственного интеллекта к более гуманному мышлению предвещает предсказуемые осложнения. Социальные роли могут быть заменены алгоритмами искусственного интеллекта, что подразумевает новые требования к дизайну, к прозрачности и предсказуемости новых алгоритмов. Если развивается общество, то должны и развиваться системы, обеспечивающие жизнь людей. Соответственно если общество развивается в моральном и этическом плане, то искусственный интеллект отставать никак не должен. Это необходимо для поддержания достойной жизни общества.

Перспектива ИИ со сверхчеловеческим интеллектом и сверхчеловеческими способностями ставит перед нами экстраординарную задачу — сформулировать алгоритм, который выводит суперэтическое поведение. Эти проблемы могут показаться незначительными, но это довольно предсказуемо, что мы столкнемся с ними по мере того, как будет развиваться наше общество.

ЛИТЕРАТУРА

1. Карпов В.Э., Готовцев П.М., Ройзензон Г.В. К вопросу об этике и системах искусственного интеллекта // Философия и общество. 2018. № 2(87).
2. Резолюция Европейского Парламента 2017 г. «Нормы гражданского права о робототехнике». — Режим доступа: http://robopravo.ru/uploads/s/z/6/g/z6gj0wkwhv1o/file/oQeHTCnw.pdf (Дата обращения 18.10.2018).
3. Ройзензон Г.В. Проблемы формализации понятия этики в искусственном интеллекте // Шестнадцатая национальная конференция по искусственному интеллекту с международным участием (КИИ-2018). Труды конференции. В 2-х томах. Т. 2. М.: РКП, 2018.
4. Карлюк М. Этические и правовые вопросы искусственного интеллекта. — Режим доступа: http://russiancouncil.ru/analytics-and-comments/analytics/eticheskie-i-pravovyevoprosy-iskusstvennogo-intellekta (Дата обращения 18.10.2018).
5. Голованов Г. Этика ИИ должна включать нравственные ценности Востока. — Режим доступа: https://hightech.fm/2017/12/19/ai-ieee (Дата обращения 18.10.2018).

Dyakov Viktor Fedorovich
Russian New University, Moscow, Russia

Some questions of ethics of artificial intelligence

Abstract. The article touches upon and analyzes the problem of creating, implementing and using artificial intelligence algorithms from the point of view of morality and ethics. A number of key ethical criteria for the development and implementation of artificial intelligence algorithms are listed, and the problems faced by companies and developers when using AI systems are also described.

Keywords: artificial intelligence; machine learning; ethics; morality; algorithm

REFERENCES

1. Karpov V.E., Gotovtsev P.M., Roizenzon G.V. On the question of ethics and artificial intelligence systems // Philosophy and Society. 2018. № 2(87).
2. Resolution of the European Parliament 2017 "Norms of civil law on robotics". — Access mode: http://robopravo.ru/uploads/s/z/6/g/z6gj0wkwhv1o/file/oQeHTCnw.pdf (Accessed 18.10.2018).
3. Roizenzon G.V. Problems of formalization of the concept of ethics in artificial intelligence // The sixteenth National Conference on Artificial Intelligence with international participation (CII-2018). Proceedings of the conference. In 2 volumes. — Vol. 2. — Moscow: RCP, 2018.
4. Karlyuk M. Ethical and legal issues of artificial intelligence. — Access mode: http://russiancouncil.ru/analytics-and-comments/analytics/eticheskie-i-pravovyevoprosy-iskusstvennogo-intellekta (Accessed 18.10.2018).
5. Golovanov G. Ethics of AI should include the moral values of the East. — Access mode: https://hightech.fm/2017/12/19/ai-ieee (Accessed 18.10.2018).