

Доверие к искусственному интеллекту и факторы риска его применения в различных сферах жизни

Роман Болбаков, Михаил Коваленко, Антонио-Кристи Исаев (РТУ МИРЭА)

Стремительное развитие искусственного интеллекта (далее – ИИ) обусловило его широкое применение во всём мире, однако феномен искусственного интеллекта на данный момент недостаточно изучен, что подтверждается малым количеством исследований, проведённых в этой сфере, по причине её новизны. Данная работа призвана исследовать вопрос доверия к искусственному интеллекту, рассмотреть выделенные на сегодня факторы, на которые можно опираться при изучении данного вопроса, а также обсудить вопросы внедрения искусственного интеллекта в такие сферы, как повседневное использование, бизнес и военные структуры.

Введение

Исследование берёт своё начало в работе «Доверие в искусственном интеллекте» [1], собранной на основе изучения отношения населения пяти стран (США, Канады, Германии, Великобритании и Австралии), в котором приняло участие 6054 человека, к ИИ. Выводы оказались достаточно интересными: большая часть опрошенных (62%) слышали о существовании технологии ИИ, однако мало понимают, как именно она поможет в повседневной жизни. Кроме того, 81% респондентов ожидают, что ИИ будет регулироваться на законодательном уровне и что должен существовать новый независимый регулятор ИИ.

Выводы, полученные данным исследованием, подкрепляются также и основными тезисами из многочисленных работ по изучению этого вопроса, например, тезисы статьи Дашкова Андрея Александровича [2] о факторах, которые человечеству и государству необходимо принимать во внимание при решении столь трудной дилеммы, как изучение доверия и этичности ИИ. Видите ли, в отличие от заранее запрограммированных сценариев, ИИ способен обучаться и делать выводы, которые человеком могут трактоваться двояко (примеры будут рассмотрены в статье), в конце концов, интеллект – это качество психики, которое характеризуется способностью приспосабливаться к различным ситуациям, способностью к обучению и запоминанию информации и пониманию абстрактных концепций, которым нам ещё предстоит обучить машины, а что говорить о бесконечных научно-фантастических сценариях развития ИИ!

Можно также упомянуть о знаменитых законах робототехники Айзека Азимова, которые можно использовать и в применении ИИ. На их основе предлагается рассмотреть некоторые варианты планирования подхода к разработке и обучению ИИ.

Анализ проблемы в гражданской сфере

Рассмотрим пример работы с искусственным интеллектом на примере известной компании Tesla и их разработок в сфере управления транспорта на основе компьютерного зрения. С точки зрения законодательства подобное решение является лишь достаточно умной системой удержания автомобиля на полосе, однако как западные, так и отечественные разработки гражданских беспилотников в лице Яндексa стремятся создать достаточно мощный ИИ для решения задачи передвижения по дорогам общего пользования. В таких ситуациях стоит учитывать внештатное поведение, которое в состоянии обработать человек, однако подобные вопросы вводят ИИ и его разработчиков в ступор выбора вариантов действия. В ход вступают извечные дилеммы, такие как дилемма вагонетки [3] в слегка изменённом формате. Перед водителем резко появляется выбор препятствия: пешеход или бетонное ограждение, учитывая, что скорость автомобиля достаточно высока для полной остановки и, скорее всего, приведёт к неминуемой гибели одного из участников дилеммы. Этим вопросом задались и создатели ресурса «Moral Machine» [4], который предоставляет 13 вариантов выбора различных имитирован-

ных ситуаций. В ходе исследования результатов выбора было рассмотрено более 11 млн возможных сценариев поведения, на основании которых и предлагается обучение ИИ для понимания необходимого выбора. Также при разработке подобной системы стоит учитывать результаты краш-тестов на различных скоростях для понимания возможных повреждений самим ИИ в ходе решения подобного вопроса.

Не стоит забывать и о правовом аспекте подобных беспилотников. Прецедентом на данный момент стало поведение машины Tesla, которая хоть и остановилась на требование сотрудника полиции, однако, не дождавшись его, скрылась с места остановки. Этот эпизод считается феноменальным по причине того, что закону неизвестно, что делать с искусственным интеллектом, который, по сути, нарушил правила. Если рассматривать вопрос конкретной, кто должен понести наказание за превышение скорости: автомобиль (что в текущий момент звучит как бессмыслица), его владелец, не ответственный за поведение беспилотника, или же компания, создавшая его, не предусмотревшая подобный сценарий? Правовым ведомствам предстоит серьёзная работа по урегулированию всех аспектов подобной технологии [5].

Анализ вопроса в бизнес-сфере

Рассматривая возможности ИИ, невозможно задвигать его применение бизнесом на задний план. Начиная от задач, с которыми мы уже более или менее научились справляться, таких как анализ больших данных и введение систем распознавания речи и текста, заканчивая задачами, которые в перспективе могли бы лишиться множество людей их рабочих мест, что уже является вопросом этичности ИИ. Несмотря на полезность ИИ, не стоит забывать, что технологии подвержены кибератакам или нарушению данных, способных привести к неминуемым последствиям, например, ошибкам при анализе крови, ведущим к неправильному лечению пациентов.

Говоря о бизнесе, стоит понимать, что на текущем этапе развития невоз-

можно оставить систему, например, систему принятия решений, в подчинении ИИ, ведь недостаточно обученный механизм способен выдавать ошибки за счёт непонимания полученного результата вследствие нарушения данных. Системам искусственного интеллекта можно передать решения тех или иных бизнес-процессов, которые требуют анализа большого количества заранее известных факторов, при помощи которых система сможет анализировать решения.

Анализ проблемы в оборонной промышленности

Системы ИИ в оборонной промышленности способны принести как массу пользы, так и большие потери за счёт ошибок в интерпретации данных. Если даже человеку в экстремальной ситуации порой сложно применять принцип «свой-чужой», а решение зачастую приходит благодаря неожиданным факторам, таким как жесты, мимика, речь или внешний вид, то ИИ и подавно не справится с подобной задачей сразу. Кроме того, следует учитывать, что летальное оружие с применением технологий ИИ может неправильно интерпретировать заранее изученные данные и атаковать гражданских, что не позволяет полностью довериться подобным системам в боевом режиме. Однако, несмотря на все риски, с точки зрения человеческой этики спасти жизнь рядового, не дав ему погибнуть на поле боя, за счёт применения беспилотников, анализирующих ситуацию на основе холодного расчёта, гораздо более приемлемый вариант.

Готово ли общество к принятию этических решений с помощью ИИ?

Современные технологии в области ИИ являются примерами тенденции к повышению уровня сложности и автономии в самых разных процессах [6]. Уже давно существуют опасения по поводу этики ИИ, которые обозреваются и обсуждаются в различных трудах и работах, включая современную художественную литературу и исследования [7, 8].

Доверие касается вопросов в области компетенции человеческого познания. Также доверие касается и задач, которые человек может выполнить самостоятельно. Человек делегирует задачи, если у него нет желания их выполнять или он неспособен приступить к их

выполнению и решить их самостоятельно. В этом случае человек может положиться на общепринятое мнение или на авторитетное лицо [9].

Вопросы, рассматриваемые в статье, поднимались до этого в исследованиях [10], где авторы утверждали, что этические рассуждения и принятие этических решений различны: первое может быть достигнуто людьми, а второе (на сегодняшний день) едва ли может быть выполнено передовыми машинами. Формально в работе иностранных коллег [10] изучаются эти два понятия, но мы (люди) воспринимаем это различие интуитивно. Учитывая современные технологии и наличие чёткой границы между этическими способностями людей и ИИ, встаёт вопрос, зачем нам их сравнивать? На данный вопрос есть ответ: в некоторых простых сценариях алгоритма может быть достаточно для достижения почти (если не точно) одинаковых результатов между стороной этических рассуждений и человеком, принимающим решения исходя из морали.

Также важным принципом этического принятия решений является верность. Верность касается ответственности человека к самому себе: быть верным и правдивым в своих отношениях с другими. Это также включает в себя выполнение обещаний, выполнение взятых на себя обязательств и надёжность [11].

Современные исследования [12] показывают, что когда людей просят оценить этические решения человека или водителя ИИ, люди не отдают однозначного предпочтения ни тому, ни другому; этот нулевой результат однороден по странам. Однако когда спросили об их явном мнении, следует ли разрешать водителю принимать этические решения на дороге, испытуемые высказались более решительно против автомобилей с искусственным интеллектом. Кроме того, мы используем структурную оценку, чтобы показать, что несоответствие между этими двумя результатами вызвано комбинацией двух элементов: люди считают, что общество не хочет, чтобы ИИ принимал этические решения, и оправдывают свои убеждения об обществе (общепринятые или нет), когда их спрашивают об этом.

Вывод

На текущий момент не может быть однозначно трактован вопрос дове-

рия к ИИ и его этичности, споры об этом могут растянуться на весь жизненный цикл этой технологии, но нельзя отрицать важность влияния подобных систем на развитие понимания глобальных процессов за счёт считывания и обучения машин человеческим моральным и этическим действиям в повседневных ситуациях. При помощи технологий можно быстрее анализировать полученные результаты, искусственный интеллект позволит обработать множество медицинских исследований и провести среди них метаанализ для выявления или же опровержения закономерностей.

Литература

1. Studie: Bürgerinnen und Bürger haben wenig Vertrauen in künstliche Intelligenz // URL: <https://home.kpmg/de/de/home/themen/2021/05/studie-buergerinnen-und-buerger-haben-wenig-vertrauen-in-ki.html>.
2. Дашков А.А., Нестерова Ю.О. Построение доверия при использовании искусственного интеллекта // E-Management. 2021. Т. 4. № 2. С. 28–36.
3. Next Stop: “Trolley Problem” // URL: <https://www.merriam-webster.com/words-at-play/trolley-problem-moral-philosophy-ethics>.
4. Moral Machine // URL: <https://www.moralmachine.net/hl/ru>.
5. Синицын С.А. Искусственный интеллект в координатах права // URL: https://izak.ru/img_content/events/iskusstvennyj-intellekt-v-koordinatah-prava.pdf.
6. Pratt G.A. Is a cambrian explosion coming for robotics? // Journal of Economic Perspectives, 29 (3) (2015), pp. 51–60.
7. Wiener N. God and Golem, Inc: A Comment on Certain Points where Cybernetics Impinges on Religion vol. 42, MIT press (1964).
8. McCarthy J., Hayes P.J. Some philosophical problems from the standpoint of artificial intelligence Readings in artificial intelligence, Elsevier (1981), pp. 431–450.
9. Большой психологический словарь / под ред. Б.Г. Мещерякова, В.П. Зинченко. М.: Прайм-ЕВРОЗНАК, 2003. 632 с.
10. McDermott D. What matters to a machine? Machine ethics (2011), pp. 88–114.
11. Bersoff D., Koeppl P. The relation between ethical codes and moral principles. Ethics and Behavior 3 (3 & 4): 345–357 (1993).
12. Caro-Burnett J., Kaneko S. Is Society Ready for AI Ethical Decision Making? Lessons from a Study on Autonomous Cars. Journal of Behavioral and Experimental Economics, v. 98, Elsevier 2022.

