Секция «Компьютерное право и информационная безопасность»

Проблема «свободы воли» робота с искусственным интеллектом

Научный руководитель – Батурин Юрий Михайлович

Де Апро Сона Вагановна

Аспирант

Московский государственный университет имени М.В.Ломоносова, Высшая школа государственного аудита, Кафедра информационной безопасности и компьютерного права, Москва, Россия

E-mail: s.deapro@yandex.ru

В философии еще со времен Сократа ведутся споры о существовании свободы воли, ее проявлении, природе. Наиболее яркими позициями являются инкомпатибилизм, где свобода воли и детерминизм несовместимы (Лютер, Гольбах, Рид), и компатибилизм, где детерминизм совместим со свободой воли (Шопенгауэр, Юм). Размышления о свободе воли были связаны с человеком, с предопределенностью его действий, с ответственностью за эти действия и осознанием этой ответственности, а также со свободой действий в соответствии с собственными мотивами и с волей как следствием человеческих желаний, не определенной внешними условиями. Попытки выяснить, что же такое свобода воли для человека, продолжаются до сих пор. С развитием информационных технологий и появлением новых потенциально возможных агентов - роботов с искусственным интеллектом (далее — роботы с ИИ) вопрос свободы воли становится еще более сложным.

Сегодня роботы с ИИ представляются в юридической литературе как «объекты» и как «субъекты». В первом случае роботов с ИИ можно причислить, согласно ст. 128 ГК РФ, к объектам гражданского права ввиду наличия имущественной ценности в качестве технологического решения, программы для ЭВМ (ст. 1261 ГК РФ). Однако роботов с ИИ нередко пытаются «очеловечить», внедряя в них качества и способности людей и представляя их в виде «рациональных агентов», «гуманоидов». И второй случай показывает, что робот с ИИ претерпевает определенную «трансформацию», в которой, согласно Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники на период до 2024 года, утвержденной Распоряжением Правительства РФ от 19.08.2020 № 2129-р, объект приобретает черты «субъекта». Говоря о проблеме «свободы воли» робота с ИИ, мы уже думаем о нем не как об объекте, а как о «субъекте». При этом важно помнить, что суть работы искусственного интеллекта заключается в заложенных в него математических алгоритмах, реализованных в кодах. Если робот с ИИ принимается как «субъект», то следует иметь в виду, что действия, совершаемые данным «субъектом», будут основаны на заложенных в него алгоритмах, что может говорить о «предопределенности» действий роботов с ИИ.

Есть ли у робота с ИИ «свобода воли»? Дадим определения понятиям «воля», «свобода» и «свобода воли» для более ясного понимания исследуемой проблемы.

Воля — это способность принимать решения автономно и направлять внутренние усилия на достижение поставленных целей.

Свобода — возможность деятельности и поведения в условиях отсутствия внешнего целеполагания.

Свобода воли — способность неалгоритмически выбирать между возможными способами действий в условиях отсутствия внешнего принуждения лишь на основе автономно сформированного принципа этического выбора.

Так, проблема свободы воли робота с ИИ представляет собой вопрос о способности такого робота осуществлять автономный «этический» выбор без внешнего «принуждения». Т.е. может ли робот с ИИ перейти из алгоритмически детерминированной позиции

в «автономную» и делать «самостоятельный» этический выбор? Важно отметить, что «самостоятельным» выбор нам может только казаться, а на самом деле быть результатом алгоритма, обработавшим изменившиеся данные с датчиков.

Следует уточнить следующее: что делает выбор реальным? И что может отличать выбор, сделанный человеком, от выбора, сделанного роботом с ИИ? Свобода воли представляет собой философско-этическую проблему, которая вбирает в себя такие понятия, как вина, совесть, ответственность. Подлинный выбор предполагает понимание собственной ответственности за сделанный выбор. В этике наличие свободы воли определяет моральную ответственность человека за свои действия. При этом важно подчеркнуть, что наличие свободы воли и осознание этой воли отнюдь не равнозначны друг другу. Может ли робот с ИИ осознавать свою способность служить причиной или источником собственных действий, осознавать свою «ответственность»? Робот с ИИ полагается на алгоритмы для поиска решения и не может даже догадываться о собственной «ответственности» до тех пор, пока в его же алгоритмы не будут заложены понятия, «объясняющие» ему, какое действие и какой результат от того или иного действия является «добром», а какое — «злом». Для того, чтобы принять этическое решение, нужно знать, что правильно, а что нет, а затем с помощью свободы воли принять решение относительно «правильного» и «неправильного» поступка. В случае с роботами с ИИ для достижения даже совсем условной их «свободы воли» для начала следует выработать для них алгоритмическое решение, которое позволит роботам с ИИ посредством самообучения и внутреннего распределения «хорошего» и «плохого» определять для себя, какая перед ними разворачивается ситуация и к какому «моральному» выбору или «аморальному» последствию в случае ошибки с выбором она ведет. Вопрос ошибки в случае с принятием решений роботом с ИИ имеет неоднозначный характер. С одной стороны, может действительно случиться, к примеру, технический сбой в заложенной в робот с ИИ программе. С другой же стороны, ошибка может случиться ввиду вероятностного и усредненного подхода заложенных алгоритмов в поиске тех или иных решений и осуществлении выбора между имеющимися накопленными данными для выдачи результата.

Может ли робот с ИИ «быть уверенным» в своих расчетах, и насколько может быть велика эта уверенность, чтобы принимать решения робота с ИИ за бесспорные, адекватные и абсолютно точные? Может ли человек быть уверенным в своих расчетах при осуществлении выбора, и всегда ли он делает выбор по расчету? Понятие «неуверенность» присуще как человеку, так и роботу с ИИ с той точки зрения, что и человек, и робот с ИИ могут сделать выбор наугад. Только в случае с человеком такой выбор может быть сделан им намеренно, с осознанным расчетом и пониманием, что результат может быть как верным, так и ошибочным (например, при решении тестовой части экзамена), тогда как робот с ИИ не сможет сделать выбор с автономно выполненным расчетом. Выбор роботом с ИИ при его «неуверенности» будет осуществляться на вероятностной основе. Робот с ИИ не может выйти за пределы своей системы и принять решение вопреки алгоритмам и независимо от того, из какого объема данных он обогащал свой опыт, извлекал уроки и насколько «интеллектуальными» являются эти алгоритмы. Тогда как человек может просто выйти за рамки программы и принять решение, воспользовавшись «свободой воли», вопреки логике. Логика для робота с ИИ выражается в заложенных в него алгоритмах. Но интеллектом ИИ станет только тогда, когда автономно сможет принимать и нелогичные решения.

Для наглядности приведем следующий свежий пример. Нейросеть Midjourney, создающая изображение по текстовым сообщениям с момента запуска обработала несколько миллионов совершенно разных запросов по генерации идей пользователей. Первое время нейросеть выдавала довольно точные решения, отталкиваясь от использованных слов в сообщениях для генерации изображений. Однако с постепенным обогащением опыта нейро-

сети, наполнением ее базы все новыми и новыми изображениями и частыми повторениями в запросах определенного слова (допустим, «солдат»), стали «проявляться» изображения, не всегда соответствующие текстовым сообщениям. Несоответствие выражалось в том, что нейросеть стала генерировать изображения, учитывая как текстовое сообщение, так и наиболее частые слова из ранее обработанных запросов, тем самым искажая заложенную в текстовое сообщение первоначальную идею пользователя. Т.е. пользователь вместо запрошенного большого корабля с белыми парусами получил корабль с разрушенной солдатом палубой, чего не было в запросе. Можно ли в данном случае расценить выданное решение нейросети как сделанного ею выбора на основе «свободы воли»? Очевидно, нет, поскольку свобода воли предполагает способность без принуждения делать выбор между возможными способами действий. В описанном же примере в качестве «принуждения» сделать такой выбор выступили алгоритм и часто встречаемые слова в многочисленных запросах, что вошло в число данных, усредняемых нейросетью для получения результата. И такой результат стал попадаться тем пользователям, чьи запросы направлялись в момент, когда генерировались изображениями с учетом наиболее часто встречаемых слов в текстовых сообщениях. И это лишь совсем примитивный пример того, как может заложенный алгоритм повлиять на выдаваемые искусственным интеллектом решения.

Следовательно, «свободу воли» робота с ИИ следует рассматривать как метафору, с которой необходимо обращаться аккуратно. Сегодня роботы с ИИ, как и другие компьютерные системы и технологии, не способны обладать «свободой воли», как и не способны обладать собственным «интеллектом», поскольку являются детерминированными объектами с некоторыми признаками «субъекта» и зависят от заложенных в них алгоритмов. Понятия «интеллект» и «свобода воли» могут относиться к человеку, но не к роботу с «ИИ». Способен ли «искусственный интеллект» обучиться принятию этически безупречных решений? Этого не всегда удается сделать даже человеку. А в случае с роботами с «ИИ» алгоритмическое принятие решений приводит к усреднению имеющихся данных. Полученные результаты вновь подводят нас к вопросу этики, с которой робот с «ИИ» не способен работать, тогда как для человека при проявлении свободы воли этика становится руководящим началом и философским фундаментом, напоминающем об ответственности и моральных принципах, что еще предстоит заложить в роботы с «ИИ».

Источники и литература

1) Гражданский кодекс Российской Федерации (часть четвертая) от 18 декабря 2006 года № 230-ФЗ (в ред. ФЗ от 11 июня 2021 года № 213-ФЗ) // Собрание законодательства РФ. 2006. № 52 (1 ч.). Ст. 5496; 2021. № 24 (Часть I). Ст. 4231.
2) Гражданский кодекс Российской Федерации (часть первая) от 30 ноября 1994 года № 51-ФЗ (в ред. ФЗ от 28 июня 2021 года № 225-ФЗ)// Собрание законодательства РФ. 1994. № 32. Ст. 3301; 2021. № 27 (часть I). Ст. 5053.
3) Распоряжение Правительства РФ от 19.08.2020 № 2129-р «Об утверждении Концепции развития регулирования отношений в сфере технологий искусственного интеллекта и робототехники на период до 2024 г.»
4) Батурин Ю. М., Полубинская С. В. Искусственный интеллект: правовой статус или правовой режим? // Государство и право. – 2022. – Номер 10 С. 141-154. Режим доступа: URL: http://gospravo-journal.ru/s102694520022606-7-1/. DOI: 10.31857/S102694520022606-7 (дата обращения: 16.02.2023)
5) Chuan Hiang Teng Free Will and AI: Making a Clear Distinction Between the Two [Электронный ресурс]. – Режим доступа: URL: https://becominghuman.ai/free-will-and-ai-85adbb09ac07 (дата обращения: 16.02.2023)