***Авэ Мэтс**, старший исследователь*
avemets@ut.ee
Тартуский университет (г. Таллин)

Авэ Мэтс. МОГУТ ЛИ РОБОТЫ СТАТЬ ЧЛЕНАМИ ОБЩЕСТВА?

Исаак Азимов в 1942 году [1] сформулировал три закона для роботехники, второй из которых звучит так: «Робот должен повиноваться всем приказам, которые даёт человек, кроме тех случаев, когда эти приказы противоречат Первому Закону.» (первый закон: «Робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинён вред.»). Однако роботы и искусственный интеллект—мозг робота—кажется, развиваются так быстро, что философы серьёзно рассуждают об их статусе в обществе. Есть и случаи, где у робота действительно есть функции, как у человека, например, в качестве члена советов организации или члена семьи. Итак, многие философы обсуждают, что роботам придётся присвоить права в том виде, в каком они есть у человека [2]. Но это предполагает у робота определённые личные черты, схожие с человеческими. В том числе, мы не ожидаем, что человек всегда повинуется всем приказам, которые ему отдают. Это значит, что человек в определённой степени автономен и у него есть право быть автономным. Может ли робот быть автономным, отказаться от соблюдения приказа человека? Из каких черт состоит автономность и можно ли построить робота с такими чертами? Автономное действие в обществе предполагает и этические правила, их знание, понимание и следование. Может ли искусственный интеллект понимать мораль?

Чтобы ответить на эти вопросы, рассмотрим некоторые аспекты того, как роботы и искусственный интеллект работают и как их функционирование отличается от человека.

Основа действия—сам мир: объекты и их положения. Когда человек знаком с одним объектом, он способен узнавать другие схожие с ним объекты. Например, он видит машину, обходит её и будет узнавать машины в других объектах. Искусственному интеллекту нужно показать миллионы фотографий с машинами различных цветов, форм, снятых с разных точек зрения, чтобы его «знание» о машинах не зависело бы от маленького набора этих свойств. Это очень важно для автономных машин. Наверное, вас всех Гугл много раз принуждал пройти проверку, что вы не робот, отмечая на фотографиях квадраты с машинами или дорожными знаками. Думали ли вы при этом, что в действительности никто не сомневается, что вы не робот, но в

этой ситуации вы преподаватель, помогающий искусственному интеллекту обучиться узнавать объекты на улице? Человек учится на практике, приобретает опыт, ощущает свойства объектов, и может экстраполировать свои знания на новые ситуации. В искусственном интеллекте всё «знание» представлено как данные—векторы нулей и единиц, он не ощущает, не «понимает» качества или объекта как такового и не образует понятий. Посмотрим на пример—искусственную фотографию человека, который не существует. Всё очень даже правдиво выглядит, но дужка направлена к середине уха. Машина не знакома с дужками и с ушами как объектами и с тем, как они связаны; она узнаёт только узоры. Человек узнаёт неистину и ошибки (в том числе в данных) на основе понимания и опыта, искусственный интеллект же—не узнаёт.

Объекты просты для изучения. Но как робот мог бы знать, какая ситуация и каким образом требует моральной оценки? Можно подумать, что в ситуации, когда требуется принятие решений, проще всего применять утилитаризм, так как он опирается на калькуляции, и калькуляция—это то, что искусственный интеллект делает. Но, как рассуждают Уалач и Ален [3], в реальной жизни последствий так много и такое многообразие, что вычислить их нереально. Кроме того, часто важны такие последствия, которые нельзя объективно измерить, например, дружба или чувство безопасности. У человека есть моральные правила и ценности, которые управляют его поведением. Можно ли в искусственном интеллекте запрограммировать эти правила? Например, «не лги». Для этого робот должен знать истину и её отношение к тому, что он свидетельствовал бы в формате текста, речи, цифр, фотографий, видео и т. д. Для человека эти различные виды представления мира соединены в практике, но для машины они ряды нулей и единиц, которые вовсе не должны совпадать и быть сравниваемыми. И рассмотрим ситуацию, в которой придётся лгать: в нацистской Германии люди скрывают евреев в своих домах и Гестапо приходит и спрашивает «есть ли здесь евреи?». Мог бы искусственный интеллект знать, что в такой ситуации правило «не лги» не применяется, как и второй закон Азимова?

Робот делает, что ему прикажут делать в соответствии с тем, как он построен и запрограммирован. Если не делает, это считается техническим дефектом. В принципе, возможно запрограммировать робота не повиноваться определённым приказам, но тогда это инженер, кто решает, какими приказам повиноваться и какими нет, и робот не автономен. Можно и неповиновение сделать независимым от воли инженера: когда робот узнаёт, что ему дали приказ и что он способен сделать то, что приказали, включается, например,

случайная функция, которая решает, повиноваться ли приказу или нет. Но человеческое неповиновение не случайно. У человека есть собственные мысли, чувства, предпочтения, цели, ценности, из-за которых он решает отказаться от соблюдения приказа, который противоречит его целям и ценностям. Способен ли робот приобрести, например, путём машинного обучения, какое-либо понимание о морали и личности, или такая возможность станет слишком узкой, обусловливая переобучение (чрезмерную специализацию к узким условиям) и неспособность к экстраполяции данных?

Цели и ценности человека связаны с том, как он «составлен». Он биологическое существо, нуждающееся в воздухе, воде, пище, сне и т. д.; он ощущает эти нужды телесно. Он ментальное—интеллектуальное и эмоциональное—существо; ему нужно самосознание, саморазвитие, товарищество, и т. д. Он может страдать и болеть телесно и ментально. Всё это составляет важную основу для его ценностей и поведения. Если у робота есть надлежащие «органы» восприятия и переработки информации—сенсоры, такие как камера, микрофон и т. д., и коды обучения (например, нейросети)—он может выучить узоры поведения. Но как с дужкой и ухом, так будет и с причинами поведения: нет даже понятия причины, не говоря о подноготной конкретного поведения, её связей с человеческим естеством. Итак, чтобы робот выработал свою подлинную автономность, он должен осознать самого себя и соответствующие нужды. А эти нужды пока лишь материальные и механические, и их «знание» возможно в нём запрограммировать и данные вводить с помощью сенсоров. Например, он будет знать, когда его батарею нужно зарядить. Но даже если её не зарядить, роботу не будет от этого страдания и боли, как у человека из-за отсутствия пищи и сна. Можно его построить и запрограммировать выражать интеллектуальные и эмоциональные черты (способность разговаривать, гримасы), но тогда это симуляция и опять-таки сделано инженерами. Если нужно, чтобы робот со своей автономной деятельностью не вредил человеку (и вообще окружающему миру), он никак не сможет это экстраполировать из своих состоянии и нужд. Ему возможно в известной мере запрограммировать описание человеческих нужд и пристроить соответствующие сенсоры, например: живой человек должен находится в атмосфере, в которой 21% кислорода, 78% азота и т. д., исключая профессиональных пловцов во время плавания и т. д. Но уже видно, что количество таких правил и их исключений превышает возможности реализации, так как возможных ситуации в принципе безгранично много.

Кратко, виды восприятия и разработки информации у человека и робота качественно различаются. Робот может воспринимать данные (не ощущать) и регулярности, поверхностные узоры в них, и он способен так делать быстро и много. Но у него нет понимания и понятии и нет подлинной автономности.

Литература

1. Азимов, Исаак. Runaround (Хоровод). *Astounding Science Fiction* 29:1, 1942. С. 94-103.
2. Gunkel, David J. *Robot Rights*. Cambridge: London: MIT Press, 2018.
3. Wallach, Wendel и Colin Allen. *Moral Machines. Teaching Robots Right from Wrong*. Oxford University Press, 2009.

Е.В. Беляцкая, *магистрант*
sub_cultura@mail.ru
БГУ (Минск)

Е.В. Беляцкая. ПРОБЛЕМА ИНТЕРНЕТ-СВОБОДЫ В КУЛЬТУРЕ ПОЗДНЕЙ СОВРЕМЕННОСТИ

В культуре поздней современности формулируется необходимый вопрос о потребности в существовании автократических режимов и самого феномена автократии в условиях стремления индивидов и представляемых ими масс к свободному самоопределению. Оно, в свою очередь, выражается в наиболее объемном пространстве современной культуры, а именно - в интернете. В связи с этим автократический режим не столько как режим политический, а, в большей степени, как философский феномен использует интернет-пространство для своих целей, для распространения необходимой информации. И в данном контексте, за счет существования модерации и правил для публикации контента в социальных сетях и электронных изданиях, которые, в свою очередь, принадлежат заинтересованным в политической поддержке компаниям и корпорациям, возникает вопрос о существовании либо границах осуществления интернет-свободы.

В культуре поздней современности индивид растерян и пребывает в постоянном поиске своего места. За необходимой информацией и возможностью выбора социального определения и реализации своей либидозной энергии в контексте творческого самовыражения личность обращается к цифровым сетям, опутавшим мир своей глобальностью и всеобщностью. Однако, необходимый для удержания своей власти контроль политические институты осуществляют в том же пространстве, чем направляют свое влияние на умы отдельных индивидов. Такой феномен в современной философии и социологии называется индоктринацией. Она