

ЭТИЧЕСКИЕ АСПЕКТЫ СОЦИАЛЬНОЙ РОБОТОТЕХНИКИ



Середкина Елена Владимировна — кандидат философских наук, доцент кафедры философии и права. Пермский национальный исследовательский политехнический университет.
Российская федерация, 614990 Пермский край, Пермь, Комсомольский проспект, д. 29.
Электронная почта: elena_seredkina@pstu.ru

Аннотация: В статье рассматривается этика робототехники как ветвь техноэтики, а также история ее становления и развития. Предлагается три варианта значения в понимании этики робототехники. Первое значение имеет отношение к профессиональной этике инженеров и разработчиков роботов и в этом смысле совпадает с термином Д. Веруджио «робоэтика». Второе значение («этика робототехники») связано с разработкой морального кода в виде программы, которая устанавливается в робота. Третье значение («этика роботов») связано с пониманием роботов как полноценных моральных агентов и в этом смысле имеет больше отношения к научной фантастике и теоретическим спекуляциям, нежели к реальному предмету исследования. На основе анализа схемы развития автономных моральных агентов С. Аллена и В. Уоллаха изучается зависимость между степенью автономности роботов и их этической чувствительностью, под которой понимается способность робототехнических систем отбирать морально релевантные факты, чтобы осуществлять моральный выбор. Рассматривается возможность перехода от операционной морали, согласно которой действия машин полностью зависят от инженеров и потребителей, к так называемой функциональной морали, в рамках которой машины будут способны сами оценивать свои действия и отвечать на моральный вызов. Описывается два основных подхода в программировании социальных роботов: подход

«сверху – вниз» на основе правил, в основе которого лежат такие этические теории как деонтология и консеквенциализм (утилитаризм), а также подход «снизу–вверх» с акцентом на этику добродетели. Предпринимается попытка антиимперативистской критики нисходящего подхода на примере категорического императива Канта, трех законов робототехники А. Азимова и утилитаристского принципа максимального счастья для максимального числа людей. Рассматриваются возможные преимущества этики добродетели и особенно трактовки морали как отражения социально-культурных ценностей для программирования автоматизированных средств вождения и роботов.

Ключевые слова: социальная робототехника, этика робототехники, робоэтика, автономные моральные агенты, операционная мораль, функциональная мораль, деонтология, консеквенциализм (утилитаризм), этика добродетели.

Ссылка для цитирования: Середкина Е.В. Этические аспекты социальной робототехники // Человек. 2020. Т. 31, № 4. С. 109–127. DOI: 10.31857/S023620070010933-3.

Чем больше свободы у машин,
 тем больше необходимости в моральных стандартах.

Р. Пикард

Современные роботы и робототехнические системы хорошо выполняют функции типа навигации и манипулирования в относительно небольшом физическом замкнутом пространстве. Но готовы ли роботы действовать в реальном мире, в сложном и постоянно меняющемся социальном контексте? В настоящее время над этой проблемой работают ведущие научные лаборатории мира. Объем рынка сервисной робототехники быстро растет, и социальные роботы становятся частью повседневной жизни человека в таких сферах как образование, медицина, туризм, торговля, индустрия развлечений. В этой связи на первый план выходит новая сложная задача — научить робота эффективно взаимодействовать с людьми. Усиление *коммуникационных* характеристик роботов ведет к изменению социальных практик, что неизбежно ставит этические проблемы, связанные с будущим развитием робототехники и совместным сосуществованием людей и машин. Другими словами, появляется потребность в создании новой отрасли прикладной этики, такой как «этика робототехники», или «робоэтика».

Этика робототехники как ветвь техноэтики впервые была инициирована итальянским профессором морального богословия и

техноэтики Папского университета Святого Креста в Риме Хосе Марией Гальваном (José Maria Galvan). В частности, он прочитал доклад, посвященный этическому измерению технологий на семинаре «Гуманоиды: техно-онтологический подход», состоявшемся в 2001 году в Университете Васеда, Япония [19].

Термин «робозтика» был изобретен в начале 2000-х годов итальянским профессором Д. Веруджио, а в 2004 году в Италии прошел Первый международный симпозиум по робозтике [16]. Специфика данного мероприятия заключалась в том, что к диалогу были приглашены ученые из разных областей науки, в том числе гуманитарных, далеких от робототехники, таких как философия, социология, антропология, научная фантастика. Это был подлинно междисциплинарный диалог экспертов, мыслителей, визионеров по поводу возможных сценариев робототехнического будущего [19]. В частности, в рамках симпозиума обсуждались следующие вопросы: «В западном мире, в Японии, США и Европе в ближайшие десятилетия в нашу жизнь войдут антропоморфные роботы, которые станут компаньонами стариков и детей, помощниками медицинского персонала, пожарных, рабочих. У них будут глаза, человеческие голоса, руки и ноги; а также кожа и «мозг». В целом они будут умнее и проворнее, чем люди, которым они призваны служить. Имеют ли интеллектуальные роботы сознание? Умеют ли они «думать»? Чувствуют ли они эмоции, любовь, боль? Уже ведутся дискуссии о «сознании» интеллектуальных машин. Будут ли роботы представлять угрозу для человечества? Как только они получают от нас знания и поймут, что мы их слабее, попытаются ли они отнять у нас власть?» [16].

За подобными формулировками стоит не только воображение писателей-фантастов, но и реальная потребность в проведении социально-гуманитарной экспертизы робототехнических проектов, поскольку инженерные решения в области социальной робототехники оказывают серьезное влияние на жизнь людей.

С тех пор было проведено около 14 крупных международных конференций, посвященных этическим аспектам робототехники [см., напр.: 17]. На них обсуждались последние достижения в области искусственного интеллекта (ИИ), глубокого обучения, машинного обучения, робототехники, мобильных сетей 5G. Последнее мероприятие из этой серии состоялось в 2019 году. Речь идет о конференции «Робозтика: люди, машины и здоровье», которая была организована Папской академией жизни Ватикана [14]. Необходимо отметить, что при Папской академии существует специальная рабочая группа, посвященная робототехнике, целью которой является проведение исследований в области человеко-машинного взаимодействия с акцентом на антропологические

Е.В. Середкина
Этические аспек-
ты социальной
робототехники

и этические аспекты, связанные с «модификацией человека» и «антропоморфизацией роботов».

Большой вклад в развитие роботэтики внесла Европейская сеть исследований в области робототехники (European Robotics Research Network, EURON), которая стремится стать общеевропейской коммуникационной платформой по обмену знаниями. Основным достижением EURON является создание «Дорожной карты Роботэтики» (2006) [10]. В ней были прописаны этические ценности в области робототехники. В настоящее время в реализации программы участвует большое количество ученых, представителей естествознания, технических и гуманитарных наук. Основной целью «Дорожной карты роботэтики» является создание этики разработчиков, производителей и пользователей роботов.

В 2018 году Европейская группа по научной и технической этике опубликовала резюме по ИИ, робототехнике и автономным системам. В частности, в нем говорится: «Достижения в области искусственного интеллекта, робототехники и так называемых автономных технологий породили ряд сложных моральных вопросов. В настоящее время усилия специалистов, связанные с поиском ответов на этические, социальные и правовые проблемы с их ориентацией на общее благо, представляют собой «лоскутное одеяло» из разрозненных инициатив. В этой связи возникает необходимость в коллективном и многостороннем размышлении и диалоге, которые бы отражали общественные ценности, а также важную роль технологий. В настоящем заявлении содержится призыв начать процесс, связанный с созданием международной этической и правовой базы для проектирования, производства, использования и управления искусственным интеллектом, робототехникой и автономными системами. В заявлении также предлагается набор фундаментальных этических принципов, основанных на ценностях, заложенных в договорах Европейского союза, а также Хартии Европейского союза по основным правам» [15].

В резюме на первый план выходят вопросы, связанные с безопасностью, анализом рисков, распределением ответственности, управлением, регулированием, проектированием, сертификацией в области робототехники. Отдельно стоит вопрос о необходимости трансформации социальных институтов и законов, главной целью которых является благо общества, а также предотвращение угроз, связанных с негативными последствиями новых технологий.

Россия также принимает участие в обсуждении этических и правовых аспектов робототехники. Так, в 2017 году по инициативе Исследовательского центра проблем регулирования робототехники и искусственного интеллекта «Робоправо» начали разрабатывать «Модельную конвенцию о робототехнике и ИИ». Цель

данной конвенции заключается в том, «чтобы определить основные проблемы, которые могут возникнуть в обществе и правовой системе в связи с активным развитием киберфизических систем; обобщить ряд основных правил робототехники, высказанных в разное время различными экспертами; обозначить возможные направления решения существующих и (или) ожидаемых проблем; инициировать дискуссию между экспертами разных стран для выработки единых этических и правовых представлений о правилах создания и использования роботов и искусственного интеллекта» [5]. С 2019 года в России разрабатывается базовое законодательство о робототехнике и искусственном интеллекте. И хотя российские инициативы делают акцент в первую очередь на нормативно-правовых аспектах робототехники, тем не менее, они могут стать хорошей базой для проработки теоретических оснований робоэтики.

Что же такое робоэтика? Европейские эксперты понимают под робоэтикой современную междисциплинарную область исследования, лежащую на стыке прикладной этики и робототехники, которая изучает этические последствия внедрения интеллектуальных/автономных роботов в общество. Основная цель робоэтики заключается в том, чтобы разработать принципы морального дизайна в области робототехники [18]. Сам Веруджио, как основатель «робоэтики», определяет ее следующим образом: «Робоэтика является прикладной этикой, ориентированной на разработку научных, культурных и технических решений, которые могут быть общими для различных социальных групп и людей разных убеждений. Эти решения должны быть направлены на стимулирование развития робототехники для улучшения жизни как отдельных индивидов, так и человеческого общества в целом, а также на предотвращение ее незитичного и неправомерного использования против человека и человечества» [2, с. 90].

Три уровня этики в области робототехники

Чтобы охватить широкий спектр роботов и потенциальных роботизированных приложений, Веруджио выделял в робоэтике три уровня [19].

Уровень 1. Робоэтика (Roboethics). Этот уровень неразрывно связан с философскими вопросами, гуманитарными и социальными науками.

Уровень 2. Этика робототехники (Robot Ethics). Этот уровень относится в основном к науке и технике.

Уровень 3. Этика роботов (Robot's Ethics). Этот уровень в основном касается научной фантастики.

Американский философ Кит Абни также выделяет три значения этики в области робототехники [8, р. 35]. В этой работе мы будем придерживаться его классификации.

Робозтика как прикладная этика профессионального сообщества

Первое значение (уровень 1) имеет отношение к профессиональной этике инженеров и разработчиков роботов и в этом смысле совпадает с термином Веруджио «робозтика». Но инженеры не могут самостоятельно формулировать этические принципы робототехнической деятельности, поэтому им на помощь приходят гуманитарные науки, в первую очередь, философия и этика. В широком смысле слова робозтика изучает отношения между наукой, техникой и обществом. В узком смысле она имеет дело с философским анализом и исследованиями по этическим вопросам, возникающим в результате интеграции роботов в общество. Таким образом, робозтика предполагает развитие прикладной этики, которая, подобно этическим исследованиям по биоэтике, затрагивает универсальные фундаментальные этические проблемы. Это связано с необходимостью защищать человеческое достоинство и личную неприкосновенность; обеспечивать права слабых социальных слоев; ограничивать социальное неравенство, которое может возникнуть в процессе использования социальных роботов.

Этика робототехники как моральный код

Второе значение (уровень 2) связано с разработкой морального кода в виде программы, которая устанавливается в робота. Роботы с различным уровнем автономности должны следовать некоему программному коду, согласно которому они будут вести себя определенным образом и делать то, что должно. Это позволит людям выносить суждения относительно того, действовал ли робот морально, то есть в соответствии с программой, или же аморально, совершая нечто непозволительное по причине, например, технического сбоя, (электромеханической ошибки или бага в программном обеспечении), неадекватного анализа рисков в отношении условий его использования или просто некомпетентного программирования. Внешний программный код исключает вопрос о «самосознании» робота как автономного морального агента.

Робот ничего не знает о морали и просто следует приказам, будь то плохо или хорошо.

Этика роботов как полноценных моральных агентов

Последнее соображение приводит нас к *третьему значению* (уровню 3) этики роботов. Оно связано со способностью роботов самостоятельно совершать этические рассуждения (онтологическая перспектива от первого лица), делать моральный выбор и нести персональную ответственность за свои действия. Такие способности делают робота полноценным моральным агентом, который действует сам по себе, а не потому, что его так запрограммировали инженеры и разработчики. В данном случае роботы подобны человеку и обладают самосознанием, способностью к рациональному выбору и свободой. Понятно, что в этом значении мы не можем говорить о реальном предмете исследования, поскольку с технической точки зрения такие роботы, по крайней мере, на современном этапе научно-технического развития, невозможны. Тем не менее, этические вопросы третьего уровня могут обсуждаться либо в научно-фантастической литературе, либо быть предметом теоретических спекуляций типа *проблемы вагонетки*, известного мысленного эксперимента в этике.¹

Нас будут интересовать первые два значения — робоэтика как прикладная этика профессионального сообщества и этика робототехники для разработки морального кода. Между ними существует тесная связь. Как только роботы становятся более автономными, неизбежно встает вопрос об их моральном коде.

Рассмотрим концептуальную схему развития автономных моральных агентов С. Аллена и В. Уоллаха [9, р. 26], представленную на рисунке. В ней выделяются две переменные — автономность и этическая чувствительность, (Под этической чувствительностью понимается способность робототехнических систем корректно

Е.В. Середкина
Этические аспек-
ты социальной
робототехники

¹ Проблема вагонетки (англ. Trolley problem) — это мысленный эксперимент в этике, впервые сформулированный в 1967 году английским философом Филиппом Фут. В классической формулировке речь идет о неуправляемой вагонетке, которая мчится по рельсам в направлении стрелки, которую можно переключить. Переключив стрелку, вы можете спасти пять привязанных к рельсам людей, но при этом пострадает один человек. Этот мысленный эксперимент породил бурные дискуссии в научном мире. Ему посвящено огромное количество научных и научно-популярных статей. В последние годы проблема вагонетки стала снова актуальной в связи с обсуждением этических аспектов использования беспилотных автомобилей [3, с. 47–49].

агентности с ответственностью и правами, сравнимыми с человеческими. Однако, как мы отметили выше, современные технологии этого не позволяют, поэтому в центре нашего внимания находится функциональная мораль.

Как же возможна функциональная мораль для роботов и робототехнических систем в контексте второго значения (этики робототехники как морального кода)? Какие этические теории могут стать основой для разработки программ? Попытаемся подробно ответить на этот вопрос.

Можно выделить два основных подхода в программировании социальных роботов — нисходящий («сверху — вниз») и восходящий («снизу — вверх»). Примером первого подхода служат деонтология и консеквенциализм, второго — этика добродетели. По большому счету эти подходы связаны с толкованием основного вопроса этики. Что значит быть моральным: совершать правильные поступки или быть хорошим?

Подход «сверху — вниз»: деонтология и консеквенциализм

В данном случае этика трактуется как изучение правильного поведения. Другими словами, если ты следуешь неким универсальным законам, значит ты моральный. Здесь можно провести аналогию с правовой системой. В этой связи основной задачей программирования становится поиск набора конечных правил. Ярким примером подобного подхода в философии является категорический императив Канта, а в робототехнике — три закона робототехники А. Азимова. Деонтологический подход в этике робототехники предполагает, что должным образом запрограммированные роботы действуют в согласии с некоторыми правилами, а принятие моральных решений происходит на основе исчисления результатов в соответствии с логикой первого порядка.

Проанализируем категорический императив Канта в качестве возможного программного кода для роботов. Для начала напомним две формулировки категорического императива (КИ) [4]:

КИ (1): Поступай так, чтобы максима твоей воли могла бы быть всеобщим законом.

КИ (2): Поступай так, чтобы ты всегда относился к человечеству и в своем лице, и в лице всякого другого также, как к цели, и никогда не относился бы к нему только как к средству.

Кантовская деонтология до сих пор оказывает влияние на современные этические концепции, в том числе и на прикладную этику робототехники, но в ней содержатся существенные изъяны,

Е.В. Середкина
Этические аспек-
ты социальной
робототехники

такие как проблема применимости и игнорирование последствий. Например, робот, который не может лгать, теряет свою ценность, если попадает в руки врагов [8, р. 41]. К. Абни указывает на тот факт, что формулировка КИ (1) чересчур либеральна и допускает слишком широкое толкование [там же]. В свою очередь, это может вызвать конфликт с формулировкой КИ (2). Также может возникнуть конфликт долга, в случае если две максимы, возведенные в ранг всеобщего законодательства, вступают в противоречие между собой. В этом случае робот, запрограммированный на основе деонтологической этики, не сможет действовать. То же самое можно сказать о конфликте категорического императива с профессиональным долгом (в случае робота, речь идет о его функциональных возможностях, предоставлении услуг). Например, должен ли робот-сиделка сообщать пожилому человеку, о котором заботится, правду о его смертельной болезни? С одной стороны, категорически нельзя врать, с другой — главная задача робота-сиделки заключается в том, чтобы обеспечивать физический и душевный комфорт пациента. Таким образом, профессиональные обязанности вступают в противоречие с этикой долженствования. Если же мы обратимся к изучению кейсов, связанных с военной робототехникой, противоречия будут еще более серьезными [12]. Так, с точки зрения КИ (1) военные действия с участием роботов становятся невозможными. Особенно уязвимой является формулировка КИ (2). Например, в случае чрезвычайной ситуации роботу приходится предпринимать действия, затрагивающие волю и интересы людей без их явного согласия.

Другой пример деонтологического подхода в этике робототехники — знаменитые три закона робототехники А. Азимова. Они гласят:

1. Робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинен вред.

2. Робот должен повиноваться всем приказам, которые отдает человек, кроме тех случаев, когда эти приказы противоречат Первому Закону.

3. Робот должен заботиться о своей безопасности в той мере, в которой это не противоречит Первому или Второму Закону.

Итак, безопасность людей превыше всего. Однако в своих рассказах Азимов убедительно показывает, что эти простые, иерархически выстроенные законы, могут завести в тупик, когда робот, например, получает взаимоисключающие предписания от двух разных людей, или когда робот для сохранения жизни одного человека должен причинить вред другим. По сути, первый закон является неполным из-за проблемы «незнания». Робот ничего не знает об угрозах, рисках и негативных последствиях для

человека. В этом смысле вред, причиненный роботом, является непреднамеренным. «Например, в ответ на просьбу подать воды, робот может зачерпнуть воду из канавы, кишашую паразитами, толкнуть человека в бассейн или разбить об него лед...» [8, р. 42].

Другая сложность связана с определением степени приемлемого риска. Большой вопрос вызывает пункт в формулировке Первого закона о «бездействии». Что значит, что робот не может своим бездействием допустить, чтобы человеку был причинен вред? Не приведет ли это к тому, что робот будет постоянно вмешиваться в жизнь людей, спасая их от потенциальной угрозы? Будет ли робот при этом способен выполнять свою основную миссию? А что, если бездействие робота предотвращает сиюминутный вред для одного человека, но тем самым ставит под угрозу будущее многих людей?

Последний вопрос звучит особенно остро. Чтобы ответить на него, в 1986 году в романе «Роботы и Империя» Азимов предложил новую формулировку Нулевого закона:

Робот не может причинить вред человечеству или своим бездействием допустить, чтобы человечеству был причинен вред.

Казалось бы, ответ был найден. Но это привело к новым этическим противоречиям. Теперь страдал отдельный человек, которым должен был пожертвовать робот ради предотвращения вреда для всего человечества. И потом, как робот определит, насколько велика угроза, чтобы предпринять определенные действия? Насколько строгим должен быть «принцип предосторожности», если робот в принципе не может просчитать все возможные потенциальные и реальные риски?

Азимов не мог дать ответы на эти вопросы. Недаром японские ученые, исследующие влияние научной фантастики на развитие технологий ИИ и робототехники, отмечают: «Три закона робототехники оказали сильное влияние на многих авторов. Например, такие проекты как «Хартия роботов университета Чибэ» (2007, Япония), «Корейская хартия этических норм для роботов» (2005, Южная Корея) ориентировались как раз на законы робототехники Азимова. Тем не менее, следует отметить, что законы Азимова — всего лишь литературный прием для усиления драматургии. В сущности, большинство историй взаимодействия людей и искусственных созданий, изображенных в сборнике рассказов «Я, робот», не укладываются в установленные три закона» [7].

Это ставит под сомнение правомерность использования деонтологического подхода в этике робототехники. Прежде всего, он не учитывает всех возможных последствий. Но как раз этим занимается консеквенциализм. Утилитаризм, как главная консеквенциалистская теория, предлагает максимизировать полезность.

Е.В. Середкина
Этические аспекты
социальной
робототехники

А полезность определяется как сумма всех хороших последствий действия минус сумма плохих. Основной принцип утилитаризма: максимальное счастье для максимально возможного числа людей.

Однако этот подход также не поддается вычислительной обработке. Невозможно просчитать полезность всех возможных действий даже в краткосрочной перспективе. Таким образом, утилитаризм делает моральную оценку невозможной. Конечно, можно было бы применить экономические методы оценки, например, анализ «затраты – выгода» (*cost – benefit analysis*), то есть преобразовать хорошие и плохие последствия в экономические ценности, а затем рассчитать, какой результат максимизирует ожидаемую прибыль (полезность). Но в этом случае этика превращается в ответвление экономики, а моральные ценности полностью сводятся к экономическим. Правомерно ли это? Разумеется, нет. Любовь, сострадание, преданность, самопожертвование, честь и достоинство не имеют цены.

В настоящее время активно идет процесс конвергенции робототехнических систем с технологиями ИИ, а это значит, что вычислительная мощность роботов в ближайшие десятилетия увеличится на несколько порядков. Но дело даже не в недостаточности вычислительной мощности современной техники. Нельзя чисто техническими средствами просчитать все последствия действий робота в режиме реального времени в сложной, постоянно меняющейся, социальной среде. Роботам и их разработчикам придется столкнуться с новыми вычислительными трудностями: например, как полезность должна быть представлена в вычислительной системе, как долго будут рассчитываться последствия, сколько данных должно быть введено, каким должен быть объем информации и т.д.

Проблема заключается в том, что утилитарный робот не сможет за ограниченное количество времени определить, какие действия будут наиболее приемлемыми, или использовать всю полноту информации для принятия правильного (с точки зрения утилитаризма, полезного) решения. Но если полезность на практике неисчислима, а суть как раз и заключается в том, чтобы максимизировать полезность, что же тогда остается от утилитаризма?

Кроме того, даже если проблема исчисления будет решена, возникает другая трудность, связанная с теорией фреймов. Напомним, что теория фреймов представляет собой научную проблему выражения результатов действия в терминах логики. С философской точки зрения, проблема фреймов связана с тем, как агент может эффективно фильтровать нерелевантную информацию в процессе принятия решения. Особенно актуальной становится теория фреймов в контексте проблемы предпосылочного знания как контекстуального, открытого и/или скрытого условия знания [6].

Но как передать предпосылочные знания машинам? Главная проблема заключается в том, как робот будет выбирать релевантную информацию для принятия морального решения? На основе чего он будет отделять релевантную информацию от нерелевантной? Сможет ли робот использовать весь объем необходимой информации, особенно в новых ситуациях, когда речь идет о терабайтах нерелевантных данных? Это отсылает нас к проблеме семантической релевантности в когнитивном моделировании. Под «семантической релевантностью» подразумеваются определенные логические отношения между приобретенными убеждениями, которые могут облегчить поиск информации, и практической аргументацией при эпистемологических ограничениях, например, дефиците знания, лимита времени и т.д. Традиционные подходы к изучению проблемы релевантности (на основе логики релевантности, байесовского подхода, а также логики описания) не в состоянии преодолеть вышеизложенные ограничения и в этом смысле не могут рассматриваться в качестве надлежащих инструментов для решения проблемы релевантности. И хотя в последнее время предпринимаются попытки разработать новый инструментарий для обработки проблемы фреймов, например, с помощью системы не аксиоматических рассуждений [21], тем не менее, до удовлетворительного решения проблемы еще далеко.

Е.В. Середкина
Этические аспекты
социальной
робототехники

Подход «снизу — вверх»: этика добродетели

Мы переходим к рассмотрению другого исторически значимого подхода к программированию роботов, который называется «снизу — вверх», базирующийся на обучении методом проб и ошибок. В основе такого подхода лежит, прежде всего, этика добродетели. Если моральные теории на основе подхода «сверху — вниз» обеспокоены главным образом действиями социального агента и пытаются ответить на вопрос «Что я должен делать?», то этика добродетели, напротив, имеет дело с вопросом «Каким я должен быть»? Данный подход открывает новые перспективы в программировании роботов [20]. В этике робототехники речь идет не о «хороших» или «плохих» поступках, а о «хороших» или «плохих» роботах. Интерпретация правильного поведения здесь находится за пределами простого послушания согласно набору конечных правил. При восходящем подходе царит моральный партикуляризм, здесь нет универсальных законов и правил, каждый кейс уникален.

В этой связи необходимо отметить общую тенденцию в современной этике, которая связана с критикой принципа императивности как сущностной черты морали на примере двух версий —

кантианской и утилитаристской [1]. Обычно их рассматривают как оппозиционные этические учения, но у них есть общий знаменатель в понимании морали: моральным побуждениям приписывается принудительный характер, нормативное содержание морали сводится к законам, принципам, требованиям, правилам универсальным по своей природе. Это и стало предметом антиимперативистской критики. Нельзя сводить мораль к одной лишь императивности. «Мораль явлена не только через императивы, но и ценности, через добродетели, через поведенческие установки личности. Мораль не исчерпывается запретами, она невозможна без положительного утверждения ценностей. Наряду с принуждением в ней есть и свободное самоопределение личности» [там же, с. 10]. В этой связи можно говорить, что сторонники этики добродетели потеснили императивистскую этику. Этика добродетели прочно утвердилась в реестре современных этических парадигм.

Итак, этика добродетели не имеет отношения к оценке моральных поступков на основе результатов (консеквенциализм), или прав и обязанностей (деонтология). Теоретики добродетели со времен Аристотеля утверждают, что нравственно хорошие поступки вытекают из воспитания хорошего характера, заключающегося в реализации конкретных добродетелей. Как известно, Аристотель утверждал, что моральные добродетели отличаются от практической мудрости и интеллектуальных добродетелей, которым можно научить. Но как возможно создание робота с «добродетельным» характером? Это очень сложная, фактически нереализуемая задача. Реализация добродетелей в робототехнической системе осложняется и тем фактом, что добродетели представляют собой сложную структуру мотиваций и желаний, которые проявляются опосредованно. Автономный моральный агент, смоделированный по принципам этики добродетели, как и его деонтологические и утилитаристские аналоги, может застрять в бесконечном цикле итераций, проверяя, соответствуют ли его действия предписанным достоинствам [20]. Кроме того, человеческие добродетели характеризуются определенной стабильностью, в основе которой лежит эмоциональная составляющая. Задача разработчиков «добродетельных» роботов заключается в том, чтобы найти способ реализовать подобную стабильность в «холодной» и безэмоциональной машине.

Новые тенденции в этике робототехники

В последнее десятилетие наметился переход к так называемым гибридным подходам в программировании морально компетентных

роботов с использованием последних разработок в области технологий ИИ и когнитивных наук [13]. В любом случае, можно говорить о серьезном «эмпирическом повороте» в этике робототехники, который может иметь большое значение для дальнейших исследований в области искусственного интеллекта, а также философии когнитивного моделирования [11].

В конце статьи хотелось бы отметить еще одно перспективное направление в этике робототехники. Среди прочего мораль может трактоваться как ценностный феномен и соответственно, теория морали может развиваться в рамках общей теории ценностей, на основе которых обеспечивается согласование частных интересов, достижение и сохранение общности [1, с. 7]. В этом смысле исследования феномена морали представляют собой изучение социокультурных механизмов, формирующих особый тип индивидуальных установок и специфические способы межличностной коммуникации. На наш взгляд, было бы также интересно проанализировать влияние социокультурного контекста, а также национальных особенностей при разработке принципов проектирования морально компетентных роботов.

Рассмотрим эту проблему на примере разработки морального кода для беспилотных автомобилей. В августе 2017 года Германия первой в Европе представила этическое руководство для программирования автоматизированных систем вождения. Важно подчеркнуть, что этическая комиссия ориентировалась, согласно немецкой традиции, на кантовскую этику (деонтологию), что официально зафиксировано в документе. В этой связи возникает вопрос, возможен ли универсальный этический кодекс для использования беспилотных автомобилей? Или в будущем каждая страна начнет разрабатывать собственные этические принципы для программного обеспечения? «Образно говоря, покидая пределы Германии и въезжая в Великобританию, должен ли будет владелец автоматизированного транспортного средства перезагрузить программу с «этики Канта» на «этику Бентама»?» [3, с. 48].

Таким образом, в настоящее время этика в области социальной робототехники (прежде всего, в значении 1 и 2) является важной ветвью технической этики. Инженеры и дизайнеры все глубже осознают свою ответственность за проектирование и разработки социально устойчивой робототехники, поскольку их «творения» быстро распространяются в обществе и становятся все более автономными и рациональными. В этой связи этические аспекты социальной робототехники выходят на первый план. И не важно, что роботы до сих пор воспринимаются пользователями больше в качестве инструментов, нежели автономных моральных агентов.

Е.В. Середкина
Этические аспек-
ты социальной
робототехники

В любом случае, философы, специалисты по этике и инженеры должны работать сообща в этом направлении.

Литература

1. *Апресян Р.Г., Артемьева О.В., Прокофьев А.В.* Феномен моральной императивности. Критические очерки. М.: ИФ РАН, 2018.
2. *Введенская Е.В.* Актуальные проблемы робоэтики // Научно-исследовательские публикации. М.: ИНИОН РАН, 2019. С. 88–101.
3. *Грунвальд А., Железняк В.Н., Середкина Е.В.* Беспилотный автомобиль в свете социальной оценки техники // Технологос. 2019. № 2. С. 41–51.
4. *Кант И.* Критика практического разума / пер. с нем. СПб.: Наука, 1995.
5. Модельная конвенция о робототехнике и искусственном интеллекте. 2017. [Электронный ресурс]. URL: https://robopravo.ru/modielnaia_konvientsiia (дата обращения: 12.06.2020).
6. *Николина Н.В.* Применение теории фреймов в решении проблемы предположительного знания // Вестник ТГУ. Философия. Социология. Политология. 2017. № 39. С. 40–47.
7. *Осава Х., Хасе С., Миямото Д.* и др. Влияние научной фантастики на представления о будущем искусственного интеллекта // Технологос. 2020. № 2.
8. *Abney K.* Robotics, ethical theory, and metaethics: A guide for the perplexed // Robot ethics: The ethical and social implications of robotics. Cambridge: MIT Press, 2012. P. 35–52.
9. *Allen C., Wallah W.* Moral Machines: Teaching Robots Right from Wrong. Oxford: Univ. Press, 2009.
10. EURON Roboethics Roadmap / ed. by Verruggio G. Genoa, 2006. [Electronic resource]. URL: <https://www.roboethics.org/atelier2006/docs/R0B0ETHICS%20ROADMAP%20Rel2.1.1.pdf> (date of access: 12.06.2020).
11. *Guarini M.* Particularism, Analogy and Moral Cognition // Minds and Machines. Vol. 20. N 3. P. 385–422. DOI: 10.1007/s11023-010-9200-4.
12. *Lin P., Abney K., Bekey G.* Ethics, War, and Robots // Ethics and Emerging Technologies / ed. by Sandler R.L. London: Palgrave Macmillan, 2014.
13. *Malle B.F.* Integrating robot ethics and machine morality: the study and design of moral competence in robots // Ethics and Information Technology. 2016. Vol. 18. P. 243–256. DOI: 10.1007/s10676-015-9367-8.
14. Roboethics: Humans, Machines and Health: the Workshop (25th–26th February 2019, the Pontifical Academy for Life, Rome, Italy). [Electronic resource]. URL: <https://www.academyforlife.va/content/pav/en/news/2018/humans--machines-and-health--workshop-2019.html> (date of access: 12.06.2020).
15. Statement on the Artificial Intelligence, Robotics and ‘Autonomous’ Systems / European Group on Ethics in Science and New Technologies. Brussels: Publ. Office of the EU, 2018. [Electronic resource]. URL: https://ec.europa.eu/research/ege/pdf/ege_ai_statement_2018.pdf (date of access: 12.06.2020).
16. The ethics, social, humanitarian and ecological aspects of Robotics: The First International Symposium on Roboethics (30th – 31st January 2004, Villa Nobel, Sanremo, Italy). [Electronic resource]. URL: <https://www.roboethics.org/sanremo2004/> (date of access: 12.06.2020).
17. *Tzafestas S.G.* Roboethics: Fundamental concepts and future prospects // Information. 2018. N 9 (6): 148. DOI: 10.3390/info9060148.
18. *Tzafestas S.G.* Roboethics: A Navigating Overview. Berlin: Springer, 2015.

19. *Veruggio G.* The birth of roboethics / Proceedings of the IEEE International Conference on Robotics and Automation: Workshop on Robot Ethics. Barcelona, Spain. 2005. 18 April. P. 1–4.

20. *Wallach W., Allen C., Smith I.* Machine morality: bottom-up and top-down approaches for modelling human moral faculties // *AI & Soc* 22, 2008. Vol. 22. N 4. P. 565–582. DOI: 10.1007/s00146-007-0099-0.

21. *Xu Y., Wang P.* The frame problem, the relevance problem, and a package solution to both // *Synthese*. 2012. Vol. 187. P. 43–72.

Е.В. Середкина
Этические аспек-
ты социальной
робототехники

Ethical Aspects of social Robotics

Elena V. Seredkina

PhD in Philosophy, Associate Professor.

Department of Philosophy and Law.

Perm National Research Polytechnic University.

29 Komsomolsky Prospekt, Perm Territory, Perm 614990, Russian Federation.

E-mail: Selena36@mail.ru

Abstract. The paper explores the ethics of robotics as a branch of technoethics and the history of its formation and development. Three interpretations of ethics in the field of robotics are outlined. The first meaning deals with the professional ethics of robot engineers and designers and coincides with the term “roboethics” proposed by Gianmarco Verrugio. The second meaning (“ethics of robotics”) is related to the development of moral code programmed into robots. The third meaning (“robot’s ethics”) refers to considering robots as fully moral agents and has more to do with science fiction and theorizing rather than with the actual scope of research. The framework for analyzing the development of autonomous moral agents by Colin Allen and Wendell Wallach is used to investigate the dependence between autonomy of robots and their ethical sensitivity, which refers to the capability of robotic systems to be sensitive to morally relevant facts and to make moral choices. In this regard, we consider the option of progressing from operational morality, according to which the actions of machines are completely dependent on engineers and consumers, to the so-called functional morality, when machines will be able to evaluate their own actions and deal with moral challenges. Two main approaches to programming social robots are described: a top-down rule-based approach relying on ethical theories such as deontology and consequentialism (utilitarianism), as well as a bottom-up approach with an emphasis on virtue ethics. The top-down approach is considered from the perspective of anti-imperative criticism drawing upon Kant’s categorical imperative, Asimov’s three laws of robotics, and the utilitarian principle of maximum happiness for the maximum number of people. The paper investigates the potential benefits of virtue ethics and morality as a reflection of socio-cultural values for programming autonomous social robots.

Key words: social robotics, ethics of robotics, roboethics, autonomous moral agents, operational morality, functional morality, deontology, consequentialism (utilitarianism), ethics of virtues.

For citation: Seredkina E.V. Ethical Aspects of social Robotics // Chelovek. 2020. Vol. 31, N 4. P. 110–128. DOI: 10.31857/S023620070010933-3.

References

1. Apressyan R.G., Artemyeva O.V., Prokofyev A.V. *Fenomen moral'noy imperativnosti. Kriticheskiye ocherki* [The phenomenon of moral normativity. Critical essay]. Moscow: Institute of Philosophy, Russian Academy of Sciences Publ., 2018.
2. Vvedenskaya E.V. Aktual'nyye problemy roboetiki [Actual problems of roboethics]. *Naukovedcheskiye issledovaniya*. Moscow: Institute of Scientific Information for Social Sciences RAS Publ., 2019. P. 88–101.
3. Grunvald A., Zheleznyak V.N., Seredkina E.V. *Bespilotnyy avtomobil' v svete sotsial'noy otsenki tekhniki* [Driverless cars in the context of technology assessment]. *Tekhnologos*. 2019. N 2. P. 4–51.
4. Kant I. *Kritika prakticheskogo razuma*: per. s nem. [Critique of Pure Reason: transl. from German]. St. Petersburg: Nauka Publ., 1995.
5. *Model'naya konventsia o robototekhnike i iskusstvennom intellekte* [Model Convention on Robotics and Artificial Intelligence, 2017]. [Electronic resource]. URL: https://robopravo.ru/modielnaia_konventsia (date of access: 12.06.2020).
6. Nikolina N.V. *Primeneniye teorii freymov v reshenii problemy predposylochnogo znaniya* [Applying of the theory of frames to solve the problems of premised knowledge]. *Vestnik TGU. Filosofiya. Sotsiologiya. Politologiya*. 2017. N 39. P. 40–47.
7. Osawa H., Hase S., Miyamoto D., et al. *Vliyaniye nauchnoy fantastiki na predstavleniya o budushchem iskusstvennogo intellekta* [The Impact of Science Fiction on the Future of Artificial Intelligence]. *Tekhnologos*. 2020. N 2.
8. Abney K. *Robotics, ethical theory, and metaethics: A guide for the perplexed. Robot ethics: The ethical and social implications of robotics*. Cambridge: MIT Press. 2012. P. 35–52.
9. Allen C., Wallah W. *Moral Machines: Teaching Robots Right from Wrong*. Oxford: Univ. Press, 2009.
10. *EURON Roboethics Roadmap*, ed. by Verruggio G. Genoa, 2006. [Electronic resource]. URL: <https://www.roboethics.org/atelier2006/docs/R0B0ETHICS%20ROADMAP%20Rel2.1.1.pdf> (date of access: 12.06.2020).
11. Guarini M. Particularism, Analogy, and Moral Cognition. *Minds and Machines*. Vol. 20. N 3. P. 385–422. DOI: 10.1007/s11023-010-9200-4.
12. Lin P., Abney K., Bekey G. *Ethics, War and Robots. Ethics and Emerging Technologies*, ed. by Sandler R.L. London: Palgrave Macmillan, 2014.
13. Malle B.F. Integrating robot ethics and machine morality: the study and design of moral competence in robots. *Ethics and Information Technology*. 2016. Vol. 18. P. 243–256. DOI: 10.1007/s10676-015-9367-8.
14. *Roboethics: Humans, Machines and Health: The Workshop* (25th–26th February 2019, the Pontifical Academy for Life, Rome, Italy). [Electronic resource]. URL: <https://www.academyforlife.va/content/pav/en/news/2018/humans--machines-and-health--workshop-2019.html> pdf (date of access: 12.06.2020).
15. *Statement on the Artificial Intelligence, Robotics and 'Autonomous' Systems* / European Group on Ethics in Science and New Technologies. Brussels: Publ. Office

of the EU, 2018. [Electronic resource]. URL: https://ec.europa.eu/research/eye/pdf/eye_ai_statement_2018.pdf (date of access: 12.06.2020).

16. *The ethics, social, humanitarian and ecological aspects of Robotics*: The First International Symposium on Roboethics (30th – 31st January 2004, Villa Nobel, Sanremo, Italy). [Electronic resource]. URL: <https://www.roboethics.org/sanremo2004/> (date of access: 12.06.2020).

17. Tzafestas S. G. (2018). Roboethics: Fundamental concepts and future prospects. *Information*. 2018. N 9 (6): 148. DOI: 10.3390/info9060148.

18. Tzafestas S.G. *Roboethics: A Navigating Overview*. Berlin: Springer, 2015.

19. Veruggio G. *The birth of roboethics*. Proceedings of the IEEE International Conference on Robotics and Automation: Workshop on Robot Ethics. Barcelona, Spain. 2005. 18 April. P. 1–4.

20. Wallach W., Allen C., Smith I. Machine morality: bottom-up and top-down approaches for modelling human moral faculties. *AI & Soc* 22. 2008. Vol. 22. N 4. P. 565–582. DOI: 10.1007/s00146-007-0099-0.

21. Xu Y., Wang P. The frame problem, the relevance problem, and a package solution to both. *Synthese*. 2012. Vol. 187. P. 43–72.

Е.В. Середкина
Этические аспек-
ты социальной
робототехники