



Наука связи

ISSN: 0954-0091 (печатать) 1360-0494 (онлайн) Домашняя страница журнала: <https://www.tandfonline.com/loi/ccos20>

Принципы робототехники: регулирование роботов в реальном мире

Маргарет Боден, Джоанна Брайсон, Дарвин Колдуэлл, Керстин Даутенхан, Лилиан Эдвардс, Сара Кембер, Пол Ньюман, Вивьен Гэрри, Джефф Пегман, Том Родден, Том Соррелл, Мик Уоллис, Блей Уитби и Алан Уинфилд

Для цитирования этой статьи: Маргарет Боден, Джоанна Брайсон, Дарвин Колдуэлл, Керстин Даутенхан, Лилиан Эдвардс, Сара Кембер, Пол Ньюман, Вивьен Гэрри, Джефф Пегман, Том Родден, Том Соррелл, Мик Уоллис, Блей Уитби и Алан Уинфилд (2017) Принципы робототехники: регулирование роботов в реальном мире, Connection Science, 29:2, 124-129, DOI: 10.1080/09540091.2016.1271400

Ссылка на эту статью: <https://doi.org/10.1080/09540091.2016.1271400>



© 2017 Автор(ы). Опубликовано Informa UK Limited, торгующей как Taylor & Francis Group.



Опубликовано онлайн: 19 апреля 2017 г.



Отправьте свою статью в этот журнал



Просмотров статьи: 8997



Просмотр связанных статей



Просмотр данных кросс-марка



Цитирующие статьи: 34 Просмотр цитирующих статей





Принципы робототехники: регулирование роботов в реальном мире

Маргарет Боден^а, Джоанна Брайсон^б, Дарвин Колдуэлл^с, Керстин Даутенханд^д,
 Лириан Эдвардс^е, Сара Кемберф^ж, Пол Ньюманг^з, Вивьен Парри^и, Джефф Пегмани^к,
 Том Родденж^л, Том Соррелл^м, Мик Уоллис^н, Блэй Уитби^о и Алан Уинфилд^п

^аДепартамент информатики, Университет Сассекса, Брайтон, Великобритания; ^бКафедра компьютерных наук,
 Университет Бага, Бага, Великобритания; ^сИтальянский технологический институт, Генуя, Италия; ^дШкола компьютерных наук,
 Университет Хартфордшира, Хертс, Великобритания; ^еЮридический факультет Страткландского университета, Глазго, Великобритания; ^жДепартамент
 Медиа и коммуникации, Ювелиры, Лондонский университет, Лондон, Великобритания; ^зИнженерный факультет
 Science, Оксфордский университет, Оксфорд, Великобритания; ^иНезависимый ученый; iRU Robotics Ltd, Манчестер, Великобритания; ^кШкола
 Информатики, Ноттингемский университет, Ноттингем, Великобритания; ^лКафедра философии, Университет
 Бирмингем, Бирмингем, Великобритания; ^мШкола перформанса и индустрии культуры Лидсского университета, Лидс, Великобритания;
 Бристольская лаборатория робототехники, Университет Западной Англии, Бристоль, Великобритания

РЕЗЮМЕ

Этой статье предлагается набор из пяти этических принципов вместе с семью общими
 с общими принципами в качестве основы для ответственной робототехники. Принципы
 робототехники были разработаны в 2010 году и опубликованы в Интернете в 2011 году.
 С тех пор принципы повлияли и продолжают влиять на ряд инициатив в области этики
 роботов, но до настоящего времени официально не публиковались. Данная статья
 исправляет это упущение.

ИСТОРИЯ СТАТЬИ

Поступила в редакцию 17 августа 2016 г.
 Принята 9 октября 2016 г.

КЛЮЧЕВЫЕ СЛОВА

Этика роботов; принцип
 робототехники;
 ответственные инновации

1. Введение

В сентябре 2010 года группа представителей миротехнологий, промышленности, искусства, права и
 социальных наук встретилась на совместном семинаре EPSRC и AHRC Robotics Retreat, чтобы обсудить
 робототехнику, ее применение в реальном мире и возможные перспективы робототехники для общества.
 Роботы покинули исследовательские лаборатории и теперь используются во всем мире, в домах и на
 производстве. Мы ожидаем, что в краткосрочной, среднесрочной и долгосрочной перспективе роботы
 повлияют на нашу жизнь дома, на наш опыт работы в учреждениях, на нашу национальную и нашу мировую
 экономику и, возможно, на нашу глобальную безопасность. Однако реалии робототехники до сих пор
 относительно малоизвестны широкой публике, где доминируют образы роботов из научной фантастики и
 СМИ. Одной из целей встречи было выявить, какие шаги следует предпринять, чтобы исследования
 в области робототехники привлекли общественность, чтобы обеспечить интеграцию этой технологии в
 наше общество максимальной пользой для всех его граждан. Как и в случае с другими технологическими
 инновациями, мы должны постараться обеспечить внедрение роботов с самого начала таким образом,
 чтобы вызвать общественное доверие и доверие; максимизировать выгоды для общества и торговли; и
 заблаговременно предотвращать любые возможные непредвиденные последствия.

Учитывая их известность, невозможно обратиться к управлению робототехникой без учета
 знаменитых трех законов робототехники Азимова (Asimov, 1950). (Законы Азимова гласят

Контакты Алан Уинфилд alan.winfield@uwe.ac.uk Адрес в настоящее

время: Факультет политики и международных исследований, Уорикский университет, Ковентри, Великобритания

© 2017 Авторы. Опубликовано Informa UK Limited, торгующей как Taylor & Francis Group.

Эта статья в открытом доступе, распространяемая в соответствии с условиями лицензии Creative Commons Attribution (<http://creativecommons.org/licenses/by/4.0/>), которая
 разрешает неограниченное использование, распространение и воспроизведение на любом носителе при условии, что оригинальная работа правильно цитируется.

что 1 – робот не может причинить вред человеку или с воим бездействием позволить человеку причинить вред; 2 – робот должен подчиняться приказам, отдаваемым ему людьми, за исключением случаев, когда такие приказы противоречили бы первому закону, а 3 — робот должен защищать свое существование пока такая защита не противоречит первому или второму закону.)

Хотя они обеспечивают полезную отправную точку для обсуждения, правила Азимова — вымышленные устройства. Они не были написаны для использования в реальной жизни, и это было бы непрактично. Сделав это, не в последнюю очередь потому, что они просто не работают на практике. (Например, как может робот знать все возможные способы причинить вред человеку? Как может робот понимать и подчиняться всем приказам человека, когда даже люди путаются в том, какие приказы имеются в виду?) Расказы Азимова также показали, что даже в мире разумных роботов его законы все равно можно было обойти и найти лазейки. Но, наконец, и самое главное, Азимов законы неуместны, потому что они пытаются настоять на том, чтобы роботы вели себя определенным образом, как если бы они были людьми, тогда как в реальной жизни люди разрабатывают и используют роботов, которые должны быть фактически с объектами человеческого закона.

Когда мы рассматриваем этические последствия наличия роботов в нашем обществе, становится очевидным, что ответственность лежит не на самих роботах. Роботы — это просто инструменты различных видов, хотя и очень специализированных инструментов, и ответственность за то, чтобы они вели себя хорошо все равно должен лежать с людьми. Соответственно, правила для реальных роботов в реальной жизни должны быть преобразованы в правила, консультирующие тех, кто проектирует, продает и использует роботов, о том, как они должны действовать. Делегаты собрания разработали такой свод «правил» с целью спровоцировать более широкое и открытое обсуждение проблем. Они подчеркивают общие принципы обязанности, выделенные группой, с намерением проинформировать разработчиков и пользователей о роботов в определенных ситуациях. Эти новые правила для робототехники (не роботов) изложены ниже. Правила этических правил робототехники задуманы как живой документ. Они не предназначены как жесткие и быстрые законы, а скорее для информирования дебатов и для дальнейшего использования. Очевидно, по этим вопросам было проведено много размышлений, и этот документ не преследует подрывать любую из этих работ, но служить координационным центром для полезных дискуссий.

2. Принципы для проектировщиков, строителей и пользователей роботов

Пять правил представлены в полуплегальной версии вместе с более свободной, но более легкой для понимания версией. Эксперт, версия, которая улавливает мысли для неспециализированной аудитории. Каждое правило снабжается путем комментария рассматриваемых вопросов и пояснения важности правила.

Правило	Полуплегальный	основная аудитория
1	Роботы — это многочисленные инструменты. Роботов не должно быть предназначенные исключительно или главным образом для убийства или причинения вреда людям, кроме как в интересах национальной безопасности	Роботы не должны проектироваться как оружие, за исключением случаев по соображениям национальной безопасности

Комментарий. Инструменты имеют более чем одно применение. Мы разрешаем разрабатывать оружие, которое фермеры используют для уничтожения вредителей и паразитов, но убивать с его помощью людей (вне боевых действий) недопустимо. Явно неправильно. Ножом можно размазывать масло или колоть людей. В большинстве обществ ни оружие, ни ножи не запрещены, но при необходимости может быть введен контроль (например, законы об оружии). для обеспечения общественной безопасности. Роботы также имеют множество применений. Хотя творческий конечный пользователь может верить, что использовать робота для насильственных целей, точно так же, как и с тупым инструментом, мы говорим, что роботы никогда не должны проектироваться исключительно или даже главным образом для использования в качестве оружия с

с мертонос ный или друг ой наступательный потенц иал. Это правило, ес ли оно будет приня то, ог раничивает коммерчес кие возмознос ти роботов, но мы рас с матриваем ег о как важней ший принц ип их признания в качес тве безопас ных в г ражданс ком общес тве.

Правило	Популярный	ос новная аудитория
2	Люди, а не роботы, я вля ются я ответс твенными аг ентами. Роботы должны быть с проектированс нас колько это возможно для с блюде ния сущес твую щих законов, ос новных прав и с вобод, вклю чая неприкос новеннос ть частной жизни	Роботы должны разрабатываться и эксплуатироваться в с ответс твии с дейс твующим законодательс твом, в том числе в отношении конфиденц иальнос ти.

Комментарий. Мы можем убедиться я , что дейс твия роботов предназна чены для с блюде ния законов с делали лю ди.

Здес ь ес ть два важных момента. Во-первых , конечно, вряд ли кто-то намеренно возьметс я за с оздание робота, нару шаю щег о закон. Но дизай неры не ю рис ты, и им нужно напомнить, что при с оздании роботов, к оторые выполня ют с вои задачи как можно лучше, иног да не обх одимо с опос тавля ть их с защитными законами и общеприя тыми с тандартами в облас ти прав человека. Конфиденц иальнос ть я вля етс я ос обенно с ложной проблемой, поэ тому она упоминаетс я . Например, робот, ис пользую емый для ух ода за уя звимым человеком, вполне может быть с пользой разрабо тан для с бора информац ии об этом человеке в режиме 24/7 и передачи ее в больницы для медиц инс ких ц елей. Но польза от этог о должна быть уравнове ше на правом этог о человека на неприкос новеннос ть частной жизни и право контролиро вать с вою жизнь, например, отказы ватьс я от лечения . Собранные данные с ледует х рани ть толькo в течение ог раниченног о времени; опя ть же правило вводит определенные г арантии. Разработчики роботов должны думать о том, как подобные правила мог ут быть с блюде ны в проц ес се проектирования (например, с помощью выключателей).

Во-вторых , это правило призвано прояс нить, что роботы — это все го лишь инс трументы, предназна ченные для дос тижения ц елей и желаний, заданных лю дьми. Пользователи и владель цы нес ут ответс твеннос ть так же, как дизай неры и производители. Иног да дизай неры должны думать наперед, потому что роботы мог ут учиться я и адаптировать с вое поведение. Но пользователи также мог ут зас тавить роботов делать то, что их разработчики не предус мотрели. Иног да обя заннос тью владельц я вля етс я наблюде ние за пользователем (например, ес ли родитель купил робота для иг ры с ребенком). Но ес ли дейс твия робота дейс твительно нару шаю т закон, ответс твеннос ть, ю ридичес кая и моральная , все г да будет лежать на одном или нес колько их лю дя х , а не на роботе (как узнать, кто нес ет ответс твеннос ть, мы рас с мотрим в правиле 5 ниже) .).

Правило	Популярный	ос новная аудитория
3	Роботы — это продукты. Они должны быть разработа ны с ис пользованием проц ес сов, обесп ечиваю щих их безопас ность и защищеннос ть.	Роботы — это продукты: как и друг ие продукты, они должны быть безопас ными и надежны ми.

Комментарий. Роботы прос то не лю ди. Это элементы тех нолог ий, которые их владель цы, безус ловно, х отя т защитить (точно так же, как у нас ес ть с иг нализац ия для наших домов и автомобилей и ох рана для наших заводов), но мы все г да будем ц енить человечес кую безопас ность выше безопас нос ти машин. Наша г лавная ц ель здес ь с ос тав лав том, чтобы убедиться я , что безопас ность и безопас ность роботов в общес тве будут обесп ечены, чтобы лю ди мог ли доверя ть им и быть уверенными в них .

Это не новая проблема в тех нике. У нас уже ес ть правила и проц ес сы, г арантирую щие, например, безопас ность покупки и ис пользования бытовой тех ники и детс ких иг рушек. Для обесп ечения этог о сущес твуют х орошо отработанные сущес твующие режимы безопас нос ти потребителей: например, отрас левые кайт-марки, британс кие и между народные с тандарты, методолог ии тестирования прог раммног о обесп ечения , чтобы убедиться я , что ошибки отс утс твуют, и т. д. Мы также знаем, что общес твеннос ть знает, что прог раммное обесп ечение и компью теры мог ут быть «взломаны» пос торонними, и также не обх одимо разрабо тать проц ес сы, пок азываю щие, что роботы макс имально защище ны от таких атак. Мы с читаем, что такие правила, с тандарты и тесты

должны быть публично приняты или разработаны для индустрии робототехники как можно скорее, чтобы заверить общественность в том, что были приняты все меры предосторожности того, как робот будет выпущен для рынка. Такой процесс также прояснит для промышленности, что именно они должны делать.

Это все еще оставляя открытыми дебаты о том, в какой мере те, кто владеет роботами или эксплуатирует их, должны разрешается защищать их, например, от кражи или вандализма, скажем, с помощью встроенных электрошокеров.

группа решила удалить фразу, которая обещала право производителей или владельцев на встроить в робота возможности «самозащиты». Другими словами, мы не думаем, что робот должен всегда быть «вооруженным», чтобы защитить себя. Это на самом деле идет дальше, чем действующее законодательство, где общий вопрос будет заключаться в том, совершил ли владелец прибор уголовные действия как нападение без уважительной причины.

Правило	Популярный	основная аудитория
4	Роботы — это искусственные артефакты. Они не должны быть разработаны обманным путем, чтобы использовать живых пользователей; вместо этого их машинная природа должна быть прозрачной	Роботы — это искусственные артефакты: иллюзия эмоций и намерения не должны использоваться для эксплуатации живых пользователей

Комментарий. Одно из больших обещаний робототехники заключается в том, что игрушки-роботы могут доставлять удовольствие, комфорт и даже своего рода товарищеские отношения людям, которые не в состоянии заботиться о себе. домашних животных, будь то из-за ограничений в их домах, физических возможностей, времени или денег. Однако, как только пользователь привязывается к такой игрушке, производители могут заявлять, что у робота есть потребности или желания, которые могут несправедливо дорого стоить владельцам или их семьям больше денег. Юридическая версия этого правила была разработана для того, чтобы сказать, что хотя допустимо и даже иногда желательно, чтобы робот иногда производил впечатление реального интеллекта, любой, кто владеет роботом или взаимодействует с ним, должен иметь возможность узнать, что это действительно так и, возможно, то, для чего оно действительно было создано. Интеллект роботов искусственный, и мы подумали, что лучший способ защитить потребителей — напомнить им об этом.

гарантируя им способ «приподнять занавес» (используя метафору из «Волшебника страны Оз»).

Это правило было труднее всего четко выразить, и мы потратили много времени на обсуждение формулировки. Достижение этого на практике потребует еще больше размышлений. Должны ли все роботы иметь видимые штрих-коды или что-то подобное? Должен ли пользователь или владелец (например, родитель, покупающий робота) для ребенка всегда можно найти базу данных или зарегистрироваться, где функционал робота указано? См. также правило 5 ниже.

Правило	Популярный	основная аудитория
5	Лицо, несущее юридическую ответственность за робота, должно быть приписано	Должна быть возможность выявить, кто несет ответственность за любого робота

Комментарий. В этом правиле мы пытаемся представить практическую основу для того, что все правила выше уже не зависят от: робот никогда не несет юридической ответственности ни за что. Это инструмент. Если он выйдет из строя и причинит ущерб, виноват будет человек. Выяснить, кто ответственный человек, не может быть легким. В Великобритании реестр ответственных для автомобиля («зарегистрированный владелец») принадлежит DVLA; напротив, никому не нужно регистрироваться как официальный владелец собаки или кошки. Мы посчитали, что первая модель больше подходит для роботов, т.к. будет интересно не только установить робота, действия которого причиняют вред, но и людей пострадавшие могут также пожелать получить финансовую компенсацию от ответственного лица.

Ответственность может быть практически решена несколькими способами. Например, одним из способов продвижения вперед может быть лицензия и регистрация (так же, как и для автомобилей), в которых фиксируется, кто несет ответственность за любота. Это может относиться к владельцу или работателю только там, где правособственность не очевидно (например, для робота, который может бродить вне дома или работать в общественном учреждении, таком как школа или больницы). В качестве альтернативы, каждый робот может быть выпущен доступной для поиска онлайн-лицензией, в которой записано имя дизайнера/производителя и ответственное лицо, которое его приобрело (такая лицензия также может указывать детали, о которых мы говорили в правиле 4 выше). Очевидно, что требуется больше обсуждений и консультаций.

Важно, чтобы по-прежнему оставалась возможность разделения или передачи юридической ответственности, например, как разработчик, так и пользователь могут разделить вину, когда робот выходит из строя во время использования из-за сочетания проблем конструкции и модификаций пользователя. В таких обстоятельствах уже существуют правовые нормы для распределения ответственности (хотя мы могли бы уточнить их или потребовать страхования). Но ресурс всегда позволит пострадавшему человеку начать с выяснения того, кто, по новым принципам, несет ответственность за расставание с роботом.

3. С емь с общ ений выс ко го о уровня

В дополнение к вышеперечисленным принципам группа также разработала всеобъемлющий набор сообщений, предназначенных для поощрения ответственности в исследовательском и промышленном сообществе в области робототехники и, таким образом, для завоевания доверия к своей работе. Дух ответственных инноваций, по большей части, уже существует, но мы считаем целесообразным сделать это явным. В следующей таблице приведены сообщения вместе с пояснительными комментариями.

	Сообщение	Комментарий
1	Мы считаем, что у роботов есть потенциал оказывать огромное положительное влияние на общество. Мы хотим поощрять ответственных исследователей роботов	Изначально это было «0-е» правило, которое мы придумали на полпути. Но мы хотим подчеркнуть, что весь смысл этого оптимизм: положительный, хотя некоторые из приведенных выше правил можно расценивать как отрицательные, ограничивающие или даже нагнетающие страх. Мы думаем, что разжигание страха уже произошло, и, кроме того, есть законные опасения по поводу использования роботов. Мы считаем, что работа здесь лучше, с пообещать реализацию потенциала робототехники для всех, избегая при этом ловушек. Легко
2	Плохая практика вредит всем нам	не заметить работу людей, которые кажутся настроенными на экстремизм или безответственность, но это может легко колоссально нас в положение, в котором сейчас находятся ученые GM, когда ничто из того, что они говорят в прессе, не имеет никакого значения. Нам необходимо взаимодействовать с общественностью и брать на себя ответственность за наш
3	Решение очевидных общественных проблем поможет нам всеобщий прогресс	общественный имидж. Предыдущее примечание относится также к проблемам, вызванным широкой общественностью и писателями-фантастами, а
4	Важно продемонстрировать, что мы, как робототехники, стремимся к наилучшему стандарту практики.	не только нашими коллегами. Как указано выше
5	Чтобы понять контекст и последствия нашего исследования, мы должны работать с экспертами из других дисциплин, включая социальные науки, право, философию и искусство.	Мы должны понимать, как другие воспринимают нашу работу, и какими могут быть правовые и социальные последствия нашей работы. Мы должны выявлять, как наилучшим образом интегрировать наших роботов в социальные, правовые и культурные рамки нашего общества. Нам нужно придумать, как вести разговор о реальных возможностях нашего исследования с людьми из разных культур, которые будут смотреть на нашу работу, опираясь на широкий спектр предположений, мифов и нарративов.

Сообщение		Комментарий	
6	Мы должны учитывать этику прозрачности: есть ли ограничения на то, что должно быть открыто достаточно	Этот момент был проиллюстрирован интересной дискуссией о программном обеспечении с открытым исходным кодом и операционные системы в контексте, когда системы, которые могут использовать это программное обеспечение, обладают дополнительными возможностями, которыми обладают роботы. Что вы получаете, когда даёте роботов «сценарным детишкам»? Мы все очень поддерживали движение за открытый исходный код, но мы думаем, что нам нужна помощь в размышлении над этим конкретным вопросом и более широкими проблемами,	
7	Когда мы видим ошибочные сообщения в прессе, мы стараемся найти время, чтобы связаться с журналистами, общающимися с ними.	связанными с открытой наукой в целом. Многие люди расстраиваются, когда видят возмутительные утверждения в прессе. Но на самом деле научные репортеры не очень тохотят, чтобы их дурили, и вообще такие утверждения можно поправить и дискредитировать источниками простыми и тихими словом в адрес репортеров в подлин. Подобная кампания уже однажды успешно проводилась в конце 1990-х.	

4. Заключение

Введение, принципы и сообщения высшего уровня в разделах 1–3 представлены как первоначально опубликованные в 2011 г. (Boden et al., 2011) с небольшими редакционными исправлениями как по грамматике, так и по согласованности. Цель этой статьи не в том, чтобы пересматривать или расширять принципы и сообщения, которые здесь описаны неизменными.

С момента публикации в Интернете в 2011 году принципы робототехники распространялись различными способами и редствами массовой информации, в том числе в New Scientist (Winfield, 2011). Впоследствии принципы были процитированы в Википедии (2016 г.) и во влиятельной статье в журнале AI Magazine, в которой излагаются исследовательские приоритеты «Надежного и полезного искусственного интеллекта» (Russell, Dewey, and Tegmark, 2015 г.). Они также включены в британский стандарт BS 8611, Руководство по этическому проектированию и применению роботов и робототехнических систем (BS 8611:2016, 2016).

Благодарности

Семинар, на котором были разработаны принципы, проходил под умелым руководством Вивьен Пэрри. Мы также благодарны Яну Болдуину, Энн Гранд и Полу О’Дауду, которые сособщствовали обсуждению, а также руководителям исследовательского совета Ширеру Уэсту (AHRC) и Стивену Кемпу (EPSRC). Принципы робототехники воспроизведены здесь с любезного разрешения Исследовательского совета по инженерным и физическим наукам (EPSRC).

Заявление об открытии информации

Авторы не сообщали о потенциальном конфликте интересов.

Рекомендации

Азимов И. (1950). Я робот. Нью-Йорк: Гном Пресс.

Боден М., Брайсон Дж., Колдуэлл Д., Даутенхан К., Эдвардс Л., Кембер С., ... Уинфилд А. (2011). Принципы робототехники. Сундон, Великобритания: Исследовательский совет по инженерным и физическим наукам. Получено www.epsrc.ac.uk/research/ourportfolio/themes/engineering/activities/principlesofrobotics/. Посетили 16 августа 2016 г.

BS 8611:2016 (2016). Роботы и робототехнические устройства. Руководство по этическому проектированию и применению роботов и роботизированных систем. Лондон: Британский институт стандартов.

Раселл С., Дьюи Д. и Тегмарк М. (2015). Приоритеты исследований для надежного и полезного искусственного интеллекта. AI Magazine, 36(4), 105–114. Ассоциация развития искусственного интеллекта.

Википедия. (2016). Три закона робототехники. Получено en.wikipedia.org/wiki/Three_Laws_of_Robotics. Посетили 16 августа 2016 г.

Уинфилд А. (2011). Роботика – для людей. New Scientist, 210 (2811), 32–33. 7 мая 2011 г.