

Отредактированная версия этого документа появилась в Ежеквартальном выпуске AISB 133 (весна 2012 г.), стр. 14-15.

Создание принципов робототехники EPSRC

Джоанна Дж.

Брайсон Группа искусственных моделей естественного
интеллекта (AmonI), факультет компьютерных наук,
Университет Бата <http://www.cs.bath.ac.uk/> jjb

5 декабря 2011 г.

В конце 2010 года EPSRC неожиданно пригласил меня посетить встречу в Нью-Форест на тему «Этика и роботы». Я время от времени пишу статьи на эту тему с 1996 года, в ответ на мой опыт участия в Cog (проект гуманоидного робота) и видя, как легко и даже настойчиво люди приписывают моральные обязательства совершенно нефункциональному (в 1993 году), но смутно гуманоидный робот. С 1998 года я также веду веб-страницу по этой теме. Тем не менее, это был первый раз, когда ко мне обратился государственный орган, и я, конечно, согласился.

Встреча была трехдневной выездной под председательством (профессионально) журналиста Вивьен Пэрри. Помимо меня, среди других участников были Маргарет Боден, Дарвин Колдуэлл, Керстин Даутенхан, Паула Даксбери, Лилиан Эдвардс, Энн Гранд, Хейзел Грайан, Сара Кембер, Стивен Кемп, Пол Ньюман, Джефф Пегман, Эндрю Роуз, Том Родден, Том Сорелл, Мик Уоллис, Ширер Уэст, Блей Уитби и Алан Уинфилд, а также при умелой помощи Яна Болдуина, Дениз Даббс и Пола О'Дауда. Участники пришли в основном из робототехники, а также из гуманитарных, юридических и социальных наук. Мы работали в основном в научных кругах, а также в промышленности и в исследовательских советах (включая тогдашнего главу AHRC Ширера Уэста).

Я был удивлен, что EPSRC так долго выплескивал столько людей в таком хорошем отеле на эту тему, но они довольно рано ясно обозначили объект своей озабоченности. EPSRC рассматривает робототехнику как критически важную технологию для Великобритании и не хочет, чтобы ее постигла та же участь, что и другие «футуристические» технологии, с точки зрения общественного отвращения, граничащего с истерией, которое больше не может быть преодолено с помощью любого количества взвешенных мер. научная оценка. EPSRC хочет с самого начала установить этику роботов, чтобы обеспечить как безопасность, так и признание робототехнических технологий.

Почти сразу стало очевидно, что им удалось подобрать очень прагматичную и социально заинтересованную группу экспертов. Группа заняла очень сильную позицию в отношении моральной и этической роли робототехники, и я бы не сказал, что она является доминирующей на типичных собраниях AISB.

В последний полный день профессор интернет-права Лилиан Эдвардс и я были в

небольшое совместное собрание в рамках сессии, предназначенной для разработки результатов в качестве итогов встречи. Мы решили сделать «настоящий» свод законов для роботов.

Лириан стремилась к тому, чтобы они четко следовали, но корректировали законы Азимова, в то время как я стремился включить несколько, которые я уже разработал, когда писал «Предложение для Лиги разработчиков гуманоидных агентов» (HAL) (Брайсон, 2000). В конце концов мы остановились на пяти, первые три из которых отражают и преломляют Азимова к интересам группы. Затем группа в целом усовершенствовала не только наши «законы», но и их версии на обычном языке, и разработала дальнейший список понятий, которые необходимо сообщить вам, нашим коллегам.

Полную версию этих документов теперь можно найти, погуглив Принципы робототехники EPSRC, и с апреля 2011 года они стали политикой EPSRC. Основные принципы воспроизведены на коробке. Ниже приведены семь идей высокого уровня, которые группа хочет сообщить вам, нашим коллегам. Подробные объяснения см. на веб-сайте, но основные моменты я привел здесь.

1. Мы считаем, что роботы могут оказать огромное положительное влияние на общество. Мы хотим поощрять ответственные исследования роботов. Мы не кучка луддитов, которые «не понимают» настоящего потенциала ИИ. Мы — заинтересованные профессионалы, которые действительно хотят, чтобы ИИ работал, а роботы — реальными.
2. Плохая практика вредит всем нам. Мы не можем игнорировать ситуацию, если некоторые из наших коллеги делают вещи, которые выставляют всех нас в плохом свете.
3. Решение очевидных общественных проблем поможет нам всем добиться прогресса.
4. Важно продемонстрировать, что мы, робототехники, привержены наилучшим возможным стандартам практики.
5. Чтобы понять контекст и последствия нашего исследования, мы должны работать с экспертами из других дисциплин, включая социальные науки, право, философию и искусство. Мы все были поражены тем, как много мы узнали от этой междисциплинарной рабочей группы.
6. Мы должны рассмотреть этику прозрачности: есть ли пределы тому, что должно быть открыто доступно? Все участники встречи были привержены решениям и подходам, основанным на программном обеспечении с открытым исходным кодом, но мы пришли к выводу, что с роботами и искусственным интеллектом в целом мы обязаны убедиться, что каждый «детский сценарий» не сможет взломать систему. Система, которая имеет информацию или память о частной жизни людей.
7. Когда мы видим ошибочные сообщения в прессе, мы обязуемся найти время, чтобы связаться с журналистами, сообщающими о них. Большинство научных репортеров действительно не хотят выглядеть глупо, сообщая об «эксперте», который оказывается саморекламой или любителем сенсаций. Негромкое слово или электронное письмо часто могут погасить истерию, вызванную безответственными заявлениями.

Я хотел бы поблагодарить EPSRC, а также наших коллег, которые выступали за эту встречу. Двумя из последних были Алан Уинфилд и Том Родден.

Лично я очень горжусь и счастлив за свою профессию и нацию, что в Великобритании теперь есть официальный набор принципов робототехники, которые касаются таких важных вопросов. Но мы всего лишь одна страна, и предстоит еще много работы и пропаганды, чтобы обеспечить надлежащее использование интеллектуальной робототехники в нашем обществе.

Коробка!

1. Роботы — это многоцелевые инструменты. Роботы не должны разрабатываться исключительно или в первую очередь для убийства людей или причинения им вреда, кроме как в интересах национальной безопасности. Признавая, что творческие люди могут использовать даже мертвую рыбу в качестве оружия, мы были обеспокоены тем, чтобы запретить создание и использование автономных роботов в качестве оружия. Хотя мы прагматично признали, что это уже происходит в контексте вооруженных сил, мы не хотим, чтобы они использовались в других контекстах.
2. Люди, а не роботы, являются ответственными агентами. Роботы должны проектироваться и эксплуатироваться, насколько это практически возможно, в соответствии с существующими законами и основными правами и свободами, включая неприкосновенность частной жизни. Мы были очень обеспокоены тем, что любое обсуждение «этики роботов» может привести к тому, что отдельные лица, компании или правительства перестанут нести ответственность за создание, приобретение и развертывание роботов. Мы чувствовали, что последствия этого беспокойства значительно перевешивают любые «преимущества» удовольствия от создания чего-то разумного и ответственного, что соизволит общество. Это был закон, который, как мы знали, больше всего защитит некоторых наших коллег из AISB — следовательно (с Дэвидом Ганкелем) Я провожу симпозиум на AISB 2012, чтобы проверить, является ли это разумным правилом. Симпозиум называется «Машинный вопрос: ИИ, этика и моральная ответственность» .
3. Роботы — это продукты. Они должны быть разработаны с использованием процессов, обеспечивающих их безопасность и защищенность. Этот принцип еще раз напоминает нам, что ответственность за то, чтобы роботы не причиняли вреда, лежит на нас, как на создателях роботов, а не на самих роботах.
4. Роботы — это искусственные артефакты. Они не должны быть разработаны таким образом, чтобы обманывать уязвимых пользователей; вместо этого их машинная природа должна быть прозрачной. Это было самым трудным для согласования формулировок правила. Идея состоит в том, что каждый, у кого есть робот, должен знать, что он не «живой» и не «страдающий» , но именно обман жизни и эмоциональная вовлеченность являются целью многих терапевтических или игрушечных роботов. Мы решили, что до тех пор, пока ответственное лицо, покупающее робота, имеет хотя бы косвенный (например, документацию в Интернете) доступ к информации о том, как работает его «разум» , это обеспечит достаточную информированность населения, чтобы люди не подвергались эксплуатации.

5. Должно быть указано лицо, несущее юридическую ответственность за робота. Всегда должна быть возможность узнать, кому принадлежит робот, точно так же, как всегда можно узнать, кому принадлежит машина. Это еще раз напоминает нам о том, что что бы ни делал робот, за его действия несет ответственность некий человек или человеческое учреждение (например, компания).

См. также (примечание для редактора, пожалуйста, объедините веб-страницы и ссылки в один список, всего пять пунктов):

- Веб-сайт Принципов робототехники EPSRC.
- Машинный вопрос: ИИ, этика и моральная ответственность.
- Этика: ИИ, роботы и общество.
- (Уилкс, 2010 г.)

Рекомендации

Брайсон, Дж. Дж. (2000). Предложение для Лиги строителей гуманоидных агентов (HAL). В Барнден, Дж., редактор, Симпозиум AISB'00 по искусственному интеллекту, этике и (квази) правам человека, страницы 1–6.

Уилкс, Ю., редактор (2010 г.). Тесное взаимодействие с искусственными компаньонами: ключевые социальные, психологические, этические и дизайнерские проблемы. Джон Бенджаминс, Амстердам.