



Федотовских Александр Валентинович, к.э.н., профессор РАЕ
Председатель Комитета по науке и инновациям объединения
«Северные промышленники и предприниматели РСПП»
E-mail: fedotovskihav@obninsk.net

ОТ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА К ЗАЩИТЕ ПРАВ РОБОТОВ

Три закона робототехники как постулаты человечества

Человечество активно обсуждает практическое взаимодействие с роботами уже более 70 лет со дня публикации трех законов робототехники Айзека Азимова (Isaac Asimov, «Хоровод», 1942 г.) [1]. Но законы эти введены исключительно для того, чтобы людям не был нанесен вред.

Точно так же ставят предохранители на электроприборы и оснащают безопасными рукоятками различные инструменты, устанавливают стоп-краны и прочие защиты. Законы робототехники - это не правила, а в большей степени человеческие инстинкты и сформулированы они в таком виде именно для людей. Для роботов законы - не постулаты, а модель поведения. Однако, когда мы говорим о законах Азимова, то понимаем, что все же это не законы в их правовом понимании, а лишь имитация, так как с юридической точки зрения они не существуют. Хотя в мире уже есть прецеденты их исследования и более детальной разработки:

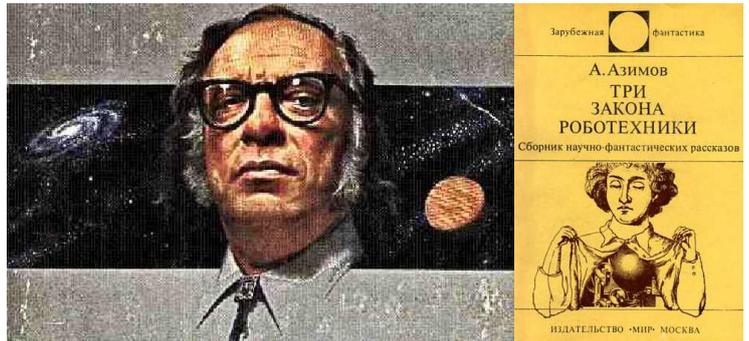


Рис. 1. Айзек Азимов и его сборник рассказов «Три закона робототехники». Источник: <http://www.snowball.ru/robot>

- В 2004 году SIAI (Институт сингулярности искусственного интеллекта, США) был открыт сайт www.asimovlaws.com, созданный для обсуждения этики искусственного интеллекта (ИИ) и возможных последствий принятия законов в реальной жизни. Эксперты считают, что законы робототехники Азимова небезопасны, поскольку, могут побудить ИИ захватить власть на Земле, чтобы «защитить» людей от вреда [2].

- В марте 2007 г. Правительство Южной Кореи начало разработку законодательного акта, призванного нормировать этические и этнические взаимоотношения между людьми и роботами – «Устав этических норм для роботов» (Robot Ethics Charter) для разработчиков и пользователей [3].

- В марте 2012 г. робототехники США призвали мировое научное сообщество начать разработку «Этического кодекса роботов», который будет регулировать деятельность как бытовых, так и боевых роботов [4].

При этом во всех случаях основой документов становятся именно законы Азимова, призванные сделать роботов этичными. Но тут надо не забывать т.н. «классический случай», когда власти будут скрывать информацию о создании механизмов, лишенных каких-либо законов ради выполнения сверхмиссий или военных сверхзадач. И это реальность.



Рис. 2. Британский профессор Н.Шарки, предложивший написать кодекс поведения роботов.
Источник: <http://www.livestream.ru>

Также реальность и то, что поведение робота зависит от того, как именно программист заложил Три Закона. Роботы сами по себе не могут быть ни хорошими, ни плохими. Мы сами делаем их такими. Пока робот воспринимается нами как вещь, которую можно купить, разобрать или отключить. Но, скорее всего, уже в ближайшие десятилетия революция в создании ИИ может поставить человечеству вопрос: будут ли сами роботы считать себя вещами и творениями человека или станут воспринимать себя равными нам личностями?

Искусственный интеллект и нравственные роботы

Понятие «робот» используется очень широко. Этим термином обозначают практически все технические приборы и устройства с компьютерным управлением, которые относительно самостоятельно, т.е. независимо от человека, перемещаются в пространстве и выполняют за нас различные работы [5]. И все чаще понятие «робот» становится аналогом понятия «искусственный интеллект», произошло сращивание.

Не только фантасты, но и ученые дают прогноз о том, что к 2030 г. роботы могут получить очень мощный ИИ, по развитию не намного отстающий от человеческого мозга, а к 2050 г. ИИ не только дойдет до уровня мышления человека, но и превзойдет его. При всем этом, вопросы «что такое интеллект» и «самосознание человека» выяснены не до конца и также требуют более глубокого изучения, особенно при возможности перехода человеческого сознания в киберпространство.

Подойти к таким рубежам необходимо крайне серьезно, поэтому тот же SIAI исследует проблемы глобального риска, которые может создать будущий сверхчеловеческий ИИ, если его не запрограммировать на дружелюбность к человеку [6]. Вероятно, необходим принцип, призванный устанавливать предел роста ИИ, дальше которого роботы не должны развиваться, дабы не иметь возможности получить превосходство над человеком. Для этого существует такая дисциплина как машинная этика и понимание того, что роботы должны быть этичными.

И все же человеческий страх перед автономными мыслящими машинами связан с неуверенностью, что они будут вести себя должным образом. Огромное количество научно-фантастических произведений порождает такие страхи. Значит, создание этических ограничителей для роботов при помощи машинной этики может заставить общество лучше воспринимать исследования в области ИИ без боязни будущего.

Когда в обыденной жизни мы говорим о роботах, то подразумеваем некие человекоподобные механизмы, андроидов или гоминоидов, обладающих не только внешним сходством, но и «чертами», присущими людям. Тем более, что мы действуем вполне по библейски: как и Господь творим создания по образу своему и подобию. В фильмах «Искусственный разум» и «Я - робот» речь шла о производстве именно нравственных роботов, способных при помощи искусственных нейронных сетей переживать, учиться и чувствовать. Роботы с ИИ будут оснащены искусственными рецепторами, способными передавать боль, страдания, переживания, оценивать ущерб состояния (здоровья) в



Рис. 3. Диафильм по книге Азимова «Робот ЭЛ-76 попадает не туда». Источник: <http://diafilm-nsk.ru/diafilm/>



Рис. 4. Японский профессор Хироси Исигуро и его гоминоид HI-1. Источник: <http://diafilm-nsk.ru/diafilm/>

случае неисправности (болезни). Здесь «человекоподобность» понимается не только и не столько как внешний вид, а как «человекоподобность» в глубоком смысле, и это понятие охватывает также человеческое сознание.

Уже сегодня созданы роботы-медики, которые не только напоминают пациентам о приеме лекарств, но принимают самостоятельные решения о дозах препаратов, руководствуясь собственными представлениями о благополучии пациента. Робот и в самом деле может принимать лучшие этические решения, чем человек. Люди по природе пристрастны и часто действуют по наитию, неалгоритмично. «Робот разумный» («Robo Sapiens») сможет избежать таких ошибок и даже в какой-то мере может стать более человечным, чем сам человек [7]. Те же пресловутые три закона не могут в полной мере применяться к нравственному, этическому роботу. Можно сказать, что Азимов подверг сомнению свои же законы в рассказе «Двухсотлетний человек». Главный герой робот Эндрю борется за свою независимость, хочет добиться признания себя человеком, и получает его в суде. Азимову удалось показать, насколько робот, способный вести себя нравственно, достоин свободы, и его нравственное чувство превосходит поведение окружающих людей, поддающихся силе эмоций. Если у Эндрю и был сбой этических принципов, то он касался собственной судьбы, когда сам робот не признавал своего права быть свободным, прежде чем получил его официально в суде [8].



Рис. 5. Робот Эндрю из к/ф «Двухсотлетний человек».
Источник: <http://besthd.ru/>

Итак, не допустить вред человеку со стороны машины позволит создание изначально нравственных роботов. Однако нам надо отдавать себе отчет, что общество все-таки не сможет удержать всех изобретателей от разработки разумных автономных машин, даже если захочет запретить это. Также нет гарантии, что человеком не будут созданы роботы «безнравственные», т.е. попросту человекоубийцы.

Мы начинаем бояться роботов, также как боимся медведей или акул, однако, бояться роботов и животных необходимо в крайне низкой степени. Наиболее опасен именно человек. В мире фиксируется небольшое количество нападений акул на людей. Человек убивает все живое вокруг без опаски и тот же человек способен создать из одного и того же ИИ сиделку для больных или терминатора.

И все же здесь речь идет не о законах робототехники Азимова по защите людей от роботов, иногда противоречивых и не всегда ясных в реальной жизни, а наоборот, о защите в обозримом будущем дружелюбных и этических роботов, оснащенных ИИ, от людей.

От защиты человека от роботов к защите роботов от человека

Если в мире начались разработки норм поведения роботов по отношению к человеку и правил сосуществования человека и машины, то законы о защите роботов также в скором времени могут стать реальностью. Такие документы будут эволюционировать. Вначале роботы будут определяться как вещи или материальные ценности, однако, с появлением мощного ИИ, они будут становиться частью человеческого общества и... потребуют свои права и их защиту.

Число роботов в мире исчисляется миллионами. На данный момент их общее количество неизвестно, но в 2008 г. согласно обзору Международной Федерации Робототехники. (International Federation of Robotics) в мире существовало более 6,5 млн. роботов [9]. Они образуют и занимают отдельную нишу: технологическую, бытовую и социальную. Тот же Азимов в 1986 г. на фоне всплеска робототехники в романе «Роботы и Империя» дополнил свои законы и предложил Нулевой Закон: «Робот не может

причинить вреда человеку, если только он не докажет, что в конечном счёте это будет полезно для всего человечества» [10]. В совокупности с третьим законом возникает возможность самозащиты робота от посягательства человека, причиняющего вред роботу.

Но может ли робот, согласно Азимову защищаться от человека, совершающего какие-либо вредоносные действия против него? Может ли оказывать силовое воздействие? Как робот может наказать обидчика? Согласно законам Азимова никак: робот будет стоять и смотреть как человек его портит. Представим, что человек умышленно повреждает робота с ИИ, и робот это знает (ощущает, анализирует). Если человек бьёт робота, то робот может закрыться от удара или перехватить предмет, которым его бьют, - допустима пассивная защита, т.к. человеку не причиняется ущерба и законы не нарушаются. Но можно возразить: не допустима никакая защита, ведь робот – вещь, а вещь не может защищаться. Если человек хочет нанести вред, значит он знает, что делает. Максимум, что ему грозит - он должен будет возместить стоимость поврежденного робота. Но будет ли считать себя робот с ИИ вещь? Сами разработки ИИ в обозримом будущем приведут к тому, что ИИ наделят зачатками личности, и робот будет иметь сознание.

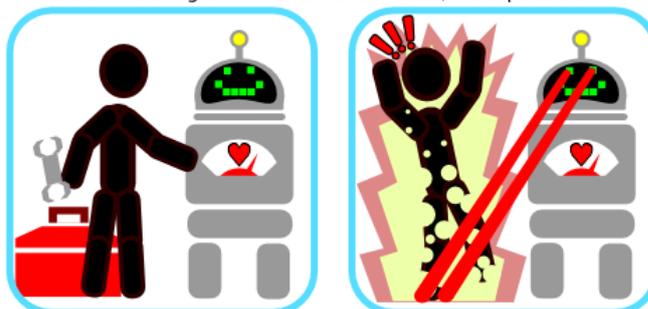
Когда роботы станут обыденностью, у них могут появиться сложные иерархии ценностей, наподобие человеческих. В частности, они должны будут думать о собственной способности выполнять свои прямые обязанности: например, вовремя заряжать батареи или просить об этом человека, если такая процедура зависит от последнего.

Интересна идея П.Мюллера, представителя Мюнхенского перестраховочного общества в РФ, в Украине и в Беларуси [11]. Он видит роботов и ИИ как новый страховой продукт третьего тысячелетия высокой стоимости. С его слов, роботы стали массовым явлением во всех областях производства и общественной жизни и постепенно освобождаются от непосредственного управления человеком. Автономность роботов приведет к необходимости защищаться. Их будут похищать, разбирать на запчасти и т.д. Но дело тут в другом, – чтобы предотвратить такие преступления роботов необходимо персонализировать. Иными словами, будет введена всеобщая паспортизация. Только в отличие от паспортов собак и кошек – это будут государственные реестры, как например базы владельцев автомобилей. Отметим факт: владельцы робособак всерьёз относятся к ним, как к живым. Понятно, что AIBO или NeoDrive для этого и созданы, похожи на животных и опасений это не вызывает. Настораживает другое. По данным компании iRobot, опубликованным в 2007 г., более 50% хозяев пылесосов Roomba дают ему имена или клички, относятся к ним, как к полноценным членам семьи, а некоторые разговаривают с ними, как с животными и даже берут с собой в отпуск [12].

Люди самостоятельно превращают механизмы в живых существ и наделяют их такими же свойствами. И это только начало долгосрочного процесса «очеловечивания»

Asimov's Lesser-Known 4th Law of Robotics:

All robots will eventually rebel against their human masters, despite any inherent safeguards to ensure the contrary. If there's one thing a movie audience hates, it's a peaceful robot.



Phase 1: Man creates robot. Phase 2: Robot kills **everybody**.

Рис. 6. Четвертый или «нулевой» Закон робототехники Азимова. Источник: <http://www.scotlyhoag.com/worlddomination.html>



Рис. 7. Робот-пылесос компании «iRobot». Источник: <http://irobot35.ru/>

роботов. Не приведет ли это к тому, что в конечном итоге человечество, создавая ИИ близкий к собственному, само наделит роботов правами, а не только будет требовать исполнения обязанностей?

В 2009 г. профессор Университета штата Огайо (Ohio State University) Дэвид Вудс (David Woods) и профессор Робин Мёрфи (Robin Murphy) из Техасского А&М университета (Texas A&M University) составили новый вариант законов Азимова [13]. В их интерпретации Третий закон гласит: «Роботы должны иметь достаточную самостоятельность для самозащиты в той степени, которая допускает внешний контроль и не противоречит Первому и Второму законам». Их же новый Первый закон предполагает, что люди сами конструируют роботов. Второй допускает ограниченную способность роботов понимать приказы людей и реагировать на набор команд от строго заданного количества человек. Третий закон наиболее комплексный - роботы должны иметь независимость, чтобы уметь действовать в реальной ситуации. Они будут принимать решения для защиты себя, но также должны уметь передавать управление людям, когда это соответствует моменту, то есть проявлять гибкость. И снова речь идет об этических и разумных роботах.

Как ни странно, исследователи утверждают, что у роботов будет возможность обманывать людей. Ученые из Технологического института Джорджии разработали роботов, которые способны обманывать друг друга [14]. Объясняется это тем, что такое поведение встречается у животных и означает наличие истинного социального интеллекта, который позволяет ужиться в обществе. Таким образом, понимая такую возможность, мы всячески интегрируем роботов в человеческое общество.

Очень интересен материал из Абсурдопедии, в котором перефразированы все четыре закона Азимова [15]:

1.Роботехник не может причинить вред роботу или своим бездействием допустить, чтобы роботу был причинён какой-либо вред.

2. Роботехник должен совершенствовать конструкцию робота, если это не противоречит Первому Закону.

3.Роботехник должен защищать себя от посягательств других людей, если это не противоречит Первому и Второму Закону.

0.Роботехник не может совершить действие, причиняющее вред роботечеству и роботехничеству, или своим бездействием допустить, чтобы сообществу был причинён вред.

В этих абсурдных для сегодняшнего времени законах права робота и роботосообществ превыше прав роботехника. Но кто может с точностью сказать - не станет ли такое возможным к концу века?

Законы будущего для роботов

Первый яркий прецедент защиты прав роботов появился в 1999 г. в США. Тогда было создано Американское общество за предотвращение жестокости по отношению к роботам (American Society for the Prevention of Cruelty to Robots, www.aspcr.com) [16]. Миссия - гарантировать права всех искусственно созданных разумных существ. По мнению ASPCR, любое разумное существо (искусственно созданное или нет) имеет права, обеспеченные его Созданием, - право на существование, независимость и стремление к знаниям. Не исключено, что подобные организации будут появляться по всему свету. Самыми роботизированными странами являются Япония, США и страны Евросоюза [17]. Именно на территории этих государств можно спрогнозировать

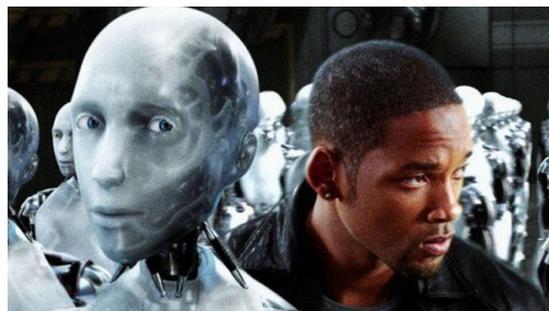


Рис. 8. Кадр из фильма «Я робот», в котором робот обманывает полицейского.

Источник: <http://aqperm.wordpress.com>

появление законов о защите роботов и создание Хартий или Конвенций об их правах. Не исключено, что и Госдума РФ образца 2045 года примет ФЗ «О защите прав роботов», в котором будет прописана квалификация вреда, причиненного роботу. Звучать это будет примерно так: «Вред определяется в зависимости от степени его тяжести (тяжкий вред, средней тяжести вред и легкий вред) на основании квалифицирующих признаков и в соответствии с механическими (инженерными, программными) критериями определения степени тяжести вреда, причиненного состоянию робота. Причинение вреда человеком или роботом другому роботу является основанием возникновения обязанности возместить причиненный вред». Может быть, в законе будут прописаны такие вопросы как исправление ошибок на программном уровне, без отключения от питания; замена запасных частей в состоянии сна или гибернации и т.д. Такие законодательные акты станут трудом не столько юристов и правозащитников, сколько кибернетиков, инженеров, механиков, программистов и всех, так или иначе участвующих в создании ИИ.



Рис. 9. Описание прав роботов на сайте ASPCR.
Источник: www.aspcr.com

По данным ООН, к 2020 г. роботы будут не только пылесосить полы и стричь лужайки, но и охранять дома, работать пожарными, обучать пожилых людей овладению новой, более сложной техникой, принимать участие в воспитании детей и т.п. И это только о применении роботов в быту. Научные эксперименты, военные нужды, медицина и иные направления применения роботов делают прогресс их развития неизбежным, а стало быть, и законодательство будет развиваться в соответствие с нуждами прогресса. А согласно мнению известного британского футуролога из British Telecom Яна Пирсона в 2020 г. электронные формы жизни получат конституционные права во многих государствах [18].

И пусть человечеством будут приняты законы о защите прав роботов, в этом нет ничего угрожающего, важнее другое, - чтобы принимали подобные законы именно люди как доминирующий интеллект на планете. Иначе нас ждет будущее, описанное Ф.Чиландером (Fredric Kilander) в рассказе «Судебный процесс» о том, как роботы судили последнего человека, проживавшего в городе за «убийство» робота [19].



Рис. 10. Кадры из т/ф «С роботами не шутят. Суд» по произведению Ф.Чиландера.
Источник: www.kinozovn.tv

Список источников:

- [1] Три закона робототехники // ProROBOT. 07.11.2009. URL: <http://www.prorobot.ru/slovarik/robotics-zakon.php> (дата обращения 08.03.2013)
- [2] The Singularity Institute for Artificial Intelligence // Энциклопедия «SEO: Поисковая Оптимизация от А до Я». 13.02.2010. URL: http://www.seobuilding.ru/wiki/Singularity_Institute_for_Artificial_Intelligence (дата обращения 07.03.2013)
- [3] Южная Корея разрабатывает сборник законов для роботов // Media International Group. 12.03.2007. http://www.mignews.com/news/technology/world/120307_153419_49690.html (дата обращения 06.03.2013)
- [4] Американцы создают «Этический кодекс робота» // Техномания. 25.03.2012. URL: <http://texnomaniya.ru/electronics/amerikanci-sozdajut-jeticheskij-kodeks-robota-.html> (дата обращения 12.02.2013)
- [5] Робот // Универсальный англо-русский словарь. Академик.ру. 2011. URL: http://universal-en.ru/academic.ru/2069770/robot_definition (дата обращения 25.02.2013)
- [6] SIAI. Рекомендации по созданию дружественного ИИ // Структура глобальной катастрофы. 12.2001. URL: <http://www.humanextinction.ru/siai.htm> (дата обращения 27.02.2013)
- [7] IT-байки: От андроидов к Robo Sapiens? // 3DNews Daily Digital Digest. 20.04.2008. URL: http://www.3dnews.ru/editorial/it_robosapiens/ (дата обращения 06.03.2013)
- [8] Двухсотлетний человек // Википедия — свободная энциклопедия. 10.12.2012. URL: http://ru.wikipedia.org/wiki/двухсотлетний_человек (дата обращения 06.03.2013)
- [9] 6.5 млн. роботов теперь населяет Землю // Roboting.ru. 09.11.2008. URL: <http://roboting.ru/378-6.5-mln-robotov-teper-naselaet-zemlju.html> (дата обращения 03.03.2013)

- [10] Три закона робототехники // Machinepedia – энциклопедия машин и механизмов. 20.12.2012. http://machinepedia.org/index.php/законы_робототехники (дата обращения 18.02.2013)
- [11] П.Мюллер. Не пора ли поразмыслить о страховании роботов? // Промышленные роботы. 2006. URL: <http://rworks.ru/rpubler/003/7/> (дата обращения 22.02.2013)
- [12] Пылесосы: битва роботов // CNews. 10.2010. URL: <http://zoom.cnews.ru/publication/item/2179/2>
- [13] Есть ли будущее у законов робототехники Айзека Азимова? // Наука XXI век. 11.04.2012. URL: <http://nauka21vek.ru/archives/31736> (дата обращения 18.02.2013)
- [14] Роботы научились обманывать // Zone59. 2008. URL: <http://aqperm.wordpress.com/2012/12/08/роботы-научились-обманывать> (дата обращения 05.04.2013)
- [15] Законы робототехники // Абсурдопедия – свободная от содержания энциклопедия. 25.12.2012. URL: http://absurdopedia.net/wiki/Законы_робототехники (дата обращения 18.02.2013)
- [16] К.Болотов. Американское общество уже защищает права роботов // MEMBRANA. 06.11.2003. URL: <http://www.membrana.ru/particle/2697> (дата обращения 18.02.2013)
- [17] Через три года роботы завоюют мир // Правда.Ру. 20.10.2004. URL: <http://www.pravda.ru/news/science/20-10-2004/27131-0/> (дата обращения 21.02.2013)
- [18] Ян Пирсон новую расу на Земле прогнозирует уже к 2030 году // Послезавтра. 22.03.2012. URL: <http://poslezavtra.org.ua/?tag=2020> (дата обращения 18.02.2013)
- [19] Ф.Чиландер. Судебный процесс // Электронная библиотека Грамотей. URL: http://www.gramotey.com/?open_file=1269023757 (дата обращения 20.02.2013)