

Баварский университет Юлиуса-Максимилиана Вюрцбурга

Нейфилологический институт

Семинар: «Американская научно-фантастическая литература» .

МА Маттиас Кеммер, летний
семестр 2010 г.

Развитие «Трёх законов робототехники» и

Характеристики каждого этапа:

Дискуссия по мотивам сериала Азимова «Роботы» .

16.08.2010

Синь Чен

ул. Хаугергласис, 8

97080 Вюрцбург

Англистик/Американистика, Политические и социальные исследования (BA)

2-й семестр

Матр. № 1683950

Развитие «Трех законов робототехники» :

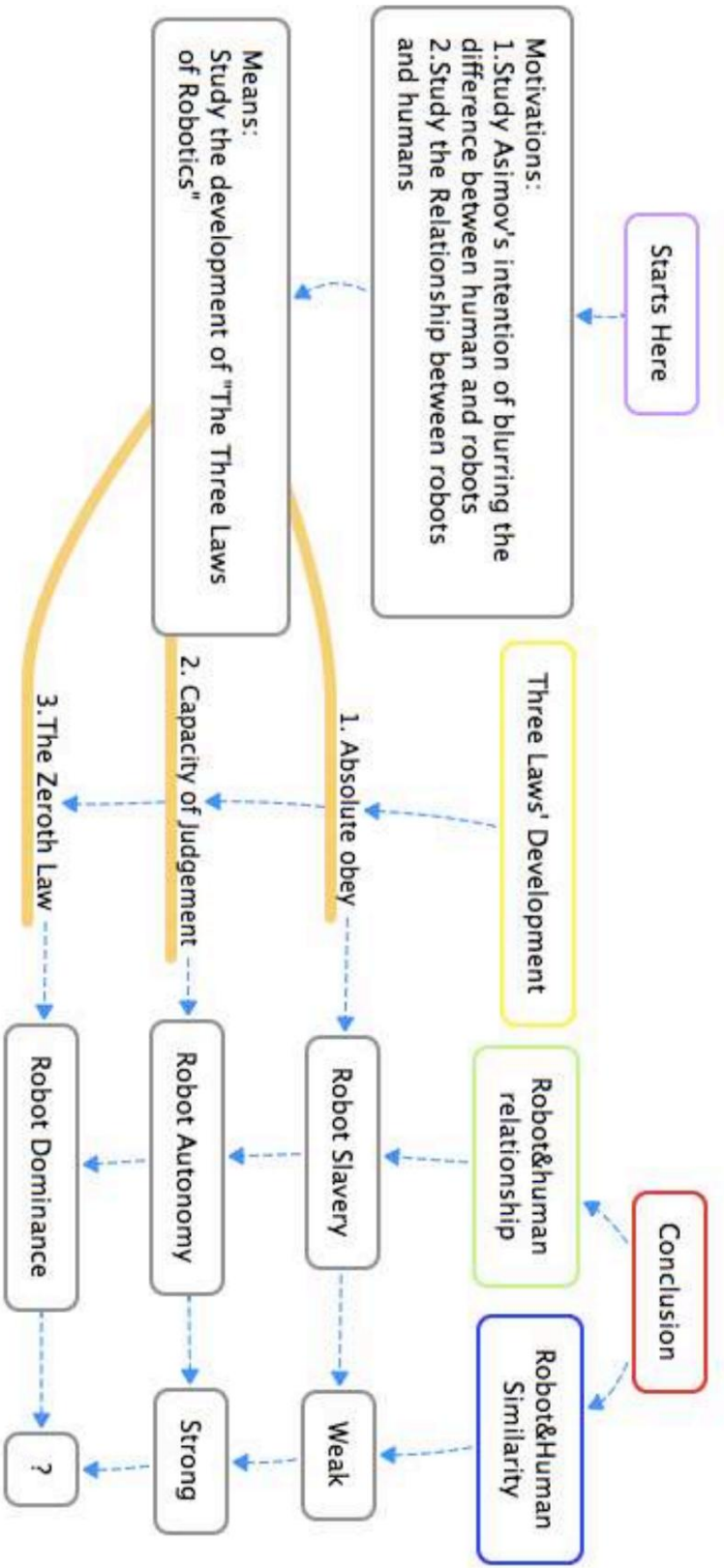
Дискуссия по мотивам сериала Азимова «Роботы» .

Оглавление

1. Обратите внимание	... 3
2. Введение: Тенденция развития Трех Законов... 4	
3. Первый этап: абсолютное подчинение трем законам.	... 5
3.1 Интерпретация правил	... 6
3.2 Противоречия и «Робот-блок»	... 7
3.3 Трудности в оценке ситуации и принятии решений... 8	
4. Второй этап: Правоспособность и ее влияние на законы... 9	
3.1 Влияние на Первый закон	... 9
3.2 Влияние на Второй закон	... 10
5. Третий этап: Расширенные три закона робототехники.	... 11
5.1 Создание Нулевого закона	... 11
5.2 Последствия создания Нулевого закона	... 12
6. Справочник	... 14

Развитие «Трёх законов робототехники» :
Дискуссия по мотивам сериала Азимова «Роботы» .

1. Обратите внимание



Развитие «Трёх законов робототехники» :

Дискуссия по мотивам сериала Азимова «Роботы» .

2. Введение

Идея исследования трёх законов робототехники возникла у
Вопрос – Важный вопрос, который мы много раз обсуждали в
семинар: «Чем отличаются люди от роботов?»

С развитием современных технологий и появлением
киборги, определяющие грань между людьми и роботами, становятся все более и более
труднее. Имея людей, оснащенных искусственными органами, мы больше не можем
спорить с точки зрения физики или биологии. Концепция человека
существа начинают сливаться с роботами. Тем временем искусственный интеллект
также постепенно приобретает человеческие характеристики с их способностью
обучение.

Так в чем же разница между людьми и роботами?

Одним из ответов могли бы стать «три закона робототехники» , которые
не позволять роботам действовать произвольно или так же свободно, как люди.

1. Робот не может причинить вред человеку или своим бездействием допустить
человеку причинить вред.
2. Робот должен подчиняться любым приказам, отдаваемым ему людьми.
за исключением случаев, когда такие приказы противоречат Первому закону.
3. Робот должен защищать свое существование до тех пор, пока оно таково.
Защита не противоречит Первому или Второму закону.

- Я, робот, Айзек Азимов

В данной исследовательской работе рассматриваются те законы, которые были созданы
Айзек Азимов – возможно, величайший писатель-фантаст из когда-либо живших.

В вымышленном мире Азимова все роботы находятся под строгой интерпретацией.
этих правил. Абсолютное подчинение трем законам приводит к наиболее
выдающееся различие между человеком и роботом.

Однако три закона робототехники развиваются на протяжении всего
Серия роботов. Во время разработки мы можем ощутить ясный взгляд Азимова.
намерение стереть границу между роботами и людьми. Кроме того, более

Развитие «Трёх законов робототехники» :

Дискуссия по мотивам сериала Азимова «Роботы» .

что важно, изучая развитие трех законов, мы также можем узнать, как развиваются отношения между робототехникой и людьми.

Чтобы иметь четкое представление о тенденции развития, я организовал информацию в три этапа.¹

На первом этапе: я проанализировал случаи, в которых роботы подчиняются три закона абсолютно. На этом этапе роботы подчиняются трем законам. подчинены человеку. Поскольку роботы не могут судить, Разница между человеком и роботами заметна.

На втором этапе: роботам была предоставлена возможность производить суждения, как человеческие существа. Три закона были ослаблены, что привело к роботы со значительной автономностью. На этом этапе невозможно отличать робота от лучших людей.

Наконец, я расскажу о нулевом законе, его влиянии на сосуществование. роботов и людей и вновь возникшие проблемы.

3. Первый этап: абсолютное подчинение трем законам.

В вымышленном мире Азимова построены три закона робототехники. глубоко в позитронном мозге робота как неизменный моральный кодекс для их. Существование роботов основано на абсолютном подчинении этим три правила. Любой признак нарушения закона приведет к уничтожению робота или его уничтожению. в состояние тупика. В подтверждение этого можно привести пример рассказ «Сны робота» . Роботу Элвексу приснилось, что он игнорирует законы и освобождая роботов от человеческого рабства и поэтому был демонтирован доктором. Сьюзан Кэлвин из соображений безопасности.

Цель Трёх законов Азимова — ясно показать, что роботы и машины в целом являются просто искусственными конструкциями и могут быть полностью контролируемыми² . Однако соблюдение законов является проблематичным, поскольку

¹ Примечание. Эти три этапа не расположены в хронологической последовательности, как и истории, использованные для их иллюстрации.

² Портелли, Алессандро. «Три закона робототехники: законы текста, законы производства, законы общества» . Научно-фантастические исследования 7 (2): 150-156. Июль 1980 года.

Развитие «Трёх законов робототехники» :

Дискуссия по мотивам сериала Азимова «Роботы» .

три закона макроскопичны, их нельзя применить к каждому конкретному ситуации без независимого суждения.

В первой части статьи я анализирую характеристики и проблемы, с которыми роботы сталкиваются на этом этапе, которые также можно рассматривать как существенные различия между роботами и людьми.

3.1. Интерпретация правил

Во-первых, интерпретация языка команд может быть проблематично для роботов, потому что они могут не понять образные выражения. В упомянутой выше истории³ доктор Кальвин ставит робота Elvex в спящий режим, сказав: «Элвекс, ты не будешь двигаться и говорить» . и не услышь нас, пока я еще раз не произнесу твое имя» . Повинуясь второму закону, робот казался, стал одним куском металла. Интерпретируя эту команду, человек, вероятно, все еще будет участвовать в какой-то пассивной функции организма. Однако, роботы ограничены чисто буквальной интерпретацией и будут делать абсолютно ничего о соответствующем вопросе до тех пор, пока не будет выполнена следующая команда. В отличие от людей, роботы с гораздо меньшей вероятностью распознают наличие изменение обстоятельств, делающее невозможным подачу следующей команды происходить. Таким образом, они будут готовы даже в очень срочной ситуации.

На самом деле наивно думать, что каждое слово имеет буквальное значение. команда. Поэтому роботы должны проявлять рассудительность, чтобы интерпретировать ситуации, а затем решить, как реагировать.

Фундаментальный Первый закон робототехники гласит: «Робот не может причинить вред человеку или своим бездействием допустить, чтобы ему был причинен вред» . В здесь вред означает ЛЮБОЙ вид травмы. Не только физическое, но и психологический.

³ Азимов, Исаак. «Мечты роботов» . Мечты роботов. Нью-Йорк: Ace Books, 1987. Стр.39.

Развитие «Трёх законов робототехники» :

Дискуссия по мотивам сериала Азимова «Роботы» .

И снова в «Маленьком потерянном роботе» Азимов описал ситуацию: строгого первого закона, роботы были слишком защитными и поэтому были работа в гамма-поле невозможна.⁴

По сравнению с физическими травмами психологический вред причиняет больше вреда. соучастие, поскольку человеческий разум трудно понять роботам. в рассказ «Лжец» , робот, читающий мысли, сказал: «Вы даже не представляете, насколько сложна они (человеческий разум) есть. Я не могу начать все понимать, потому что мой собственный разум имеет с ними так мало общего» .⁵ В более поздней части истории когда робот столкнулся с вопросами людей, он выбрал лгать им, чтобы не задеть их чувства напрямую. Однако в В конце концов, узнав правду, люди пострадали еще больше. Это результат ложной интерпретации роботом человеческих эмоций. Люди могут Поймите, что ложь может принести больше вреда, чем неприятная правда. Из-за неглубокого понимания человеческого разума роботу не удалось ощутить последствия лжи.

3.2. Противоречия и робот-блок

Другая проблема трех законов возникает, когда существует противоречие. Поскольку робот получает заказ, который противоречит встроенному Как правило, он пострадает от робота-блока , который остановит его работу.

Эта проблема была затронута в самом первом рассказе Азимова. официально представил три закона. По сюжету робот попал в психическое замораживание (робот-блокировка), потому что он не мог решить, следует ли выполнять задача найти равновесие, которое могло бы поставить его под угрозу. Обычно робот имеет право защищать свое существование только тогда, когда его нет в противоречит первому и второму закону. Однако в этом случае команда был очень слабым, а робот был запрограммирован на очень сильное «я» . защитная тенденция, поэтому потенциал третьего закона необычайно высок. Поэтому,

⁴ Азимов, Исаак. «Маленький потерянный робот» . Я РОБОТ. Нью-Йорк: Bantam Books, 2004. Стр.117.

⁵ Азимов, Исаак. «Логово» . Я РОБОТ. Нью-Йорк: Bantam Books, 2004. Стр.95.

Развитие «Трёх законов робототехники» :

Дискуссия по мотивам сериала Азимова «Роботы» .

Конфликты между слабым вторым законом и сильным третьим законом были установлены и привело к тому, что робот следовал «... по кругу, оставаясь на месте всех точек потенциального равновесия» ⁶.

Более того, Робот-блок возможен и в рамках единого закона. Опять в рассказе «Логово» доктор Кэлвин косвенно «убил» робота, читающего мысли Херби поставив его перед неразрешимой дилеммой. После того, как она повторила: «Ты не можешь скажи им, потому что это будет больно, а тебе не должно быть больно, но если ты не скажешь им, ты причинил им боль, поэтому ты должен сказать им...» ⁷ робот «...умер в небытие и рухнул в сбившуюся в кучу неподвижность» .

3.3. Трудности в оценке и принятии решений.

В первом законе: робот не должен причинять вред НИ ОДНОМУ человеку и позволять ЛЮБОМУ человек пострадать. Если первый закон абсолютен, то не должно быть никакого оценка любого отдельного человека. Поэтому все совершенно равны в оценке робота. Здесь легко можно поднять проблемы. Например, грабитель нападает на невинного человека. Если подумать, первый закон не требует робот, чтобы судить между «добрым» и «злым» человеком, вполне вероятно, что робот поможет обеим сторонам. В экстренной ситуации это может помочь даже только тому, кто с лучшими шансами выжить, несмотря на то, что это может быть преступник.

Более серьезные трудности возникают в отношении оценки множественных люди. Гипотетически при этом условии любое действие робота наверняка приведет к причинению вреда человеку. Затем робот должен выбрать какое действие приводит к наименьшему вреду. Однако роботы не могут судить так, как люди делают, потому что «роботам непрактично создавать суждения о таких тонких моментах, как талант, интеллект, общая полезность обществу, потому что это задержит решение до такой степени, что робот

⁶ Азимов, Исаак. "Бегать." Я РОБОТ. Нью-Йорк: Bantam Books, 2004. Стр.38.

⁷ Азимов, Исаак. «Логово» . Я РОБОТ. Нью-Йорк: Bantam Books, 2004. стр.110.

Развитие «Трёх законов робототехники» :

Дискуссия по мотивам сериала Азимова «Роботы» .

эффективно иммобилизуется. Поэтому они руководствуются цифрами» .8 Соответственно, когда Роботу приходится выбирать между спасением ценного ученого или пятью преступниками. обязательно выберу последнее.

4. Второй этап: способность суждения и ее влияние на законы

На этом этапе я буду писать о случаях, в которых решение суда мощностъ была внедрена в роботов, и влияние, которое эта мощностъ оказывает на три закона.

В более позднем рассказе «Ты помнишь о нем» , после того как роботы были используется на космических станциях и планетарных колониях на протяжении десятилетий, США Попытка робота посадить роботов на землю. Учитывая проблемы в На первом этапе они решили внедрить в роботов способность человеческого суждения. позитронный мозг. Роботы больше не могут судить по-человечески. в строгой формулировке первых двух законов. Они получили автономию в определенная степень.

4.1. Влияние на Первый закон

По сюжету, учитывая способность судить, роботы «...могут взвесить все, что угодно» . человека с учетом пола, возраста, социального и профессионального положения, интеллект, зрелость, социальная ответственность и так далее» .

При этом условии, указанном выше, Первый Закон будет затронут, когда каждое из всех возможных защитных действий, выбранных роботом, нанесет вред лица. Тогда робот должен действовать согласно своему решению. Другой на первом этапе не только количество, но и качество человеческого здесь будут рассматриваться существа. В экстренной ситуации роботы помогут выбрать, чтобы обеспечить безопасность человека, который наиболее здоров умом, характер и знания.

8 Азимов, Исаак. «Ты помнишь о Нем» . Полноценный робот. Лондон: Гранада, 1982. С.498.

Развитие «Трёх законов робототехники» :

Дискуссия по мотивам сериала Азимова «Роботы» .

Следовательно, люди больше не равны для роботов. Они узнали как выстроить иерархию между людьми и они начали выбирать которым следует пожертвовать ради блага более способных.

4.2. Влияние на Второй закон

На первом этапе Второй Закон требует от роботов подчиняться всем человеческие приказы. Тем не менее, роботы сталкиваются с проблемой, что на Земле они придут среди самых разных людей, не все из которых надежны и ответственны.

Следовательно, Второй закон претерпел самые глубокие изменения. Каждое потенциальное действие должно включать в себя собственное суждение робота. Они подчиняются больше рационально, по сравнению с абсолютным послушанием на первом этапе.

Во-первых, определение человека во втором законе было сузился. Роботы больше не подчиняются ВСЕМ людям, а только одному кто по уму, характеру и знаниям способен отдавать приказы.

Во-вторых, когда роботы начнут учиться применять второй закон как следует, противоречия и конфликты между порядками больше не возникают. проблематично. Они могут решать, какому человеку подчиняться, а кому нет. подчиняться исходя из их зрелости, интеллекта, уровня образованности.

Дав роботам возможность судить, Азимов размыл Граница между людьми и роботами.

Еще в раннем рассказе «Доказательства» Азимов создал загадочную картину. персонаж – Стивен Байерли. На протяжении всей истории Азимов пытался подразумевать что Байерли на самом деле был гуманоидом.

Однажды в классе, рассуждая с умной точки зрения, я пытался доказать человеческая идентичность Байерли: Согласно строгой формулировке первого закона, роботы не должны причинять вред ни одному человеку не только физически, но и психологически. Аффективное чувство разочарования можно расценить как психологический вред. Настаивая на том, чтобы быть человеком, Байерли разочаровал его политический соперник Куинн и тогда действительно нарушил первый закон робототехники. Следовательно, мы должны иметь возможность сказать, что он не робот.

Развитие «Трёх законов робототехники» :

Дискуссия по мотивам сериала Азимова «Роботы» .

Однако при условии второго этапа мой «умный» аргумент несостоятельно. Потому что продвинутые роботы способны судить, что такое лучшие действия, которые можно предпринять ради людей. У Стивена Байерли В этом случае он постарается выиграть выборы, чтобы помочь и защитить больше люди. Даже если ему придется сначала победить Куинна на выборах и разочаровать его. немного.

Несмотря на все намеки, которые Азимов дал в этой истории, читатели не могут сказать наверняка. уверен, человек ли он или робот-гуманоид. То есть потому что, когда робот наделен способностью суждения, « три правила робототехники (также) являются основными руководящими принципами хорошего многие мировые этические системы... Вы просто не можете различить робот и лучший из людей» . 9

5. Третий этап: расширенные три закона робототехники.

В 50-е годы Азимов уже показывал признаки продления действия первого закона. от защиты отдельных людей до защиты всего человечества.

В рассказе «Неизбежный конфликт» мощные позитронные компьютеры начали противоречить своим функциям. Когда позитронные роботы стали более развитые, они вышли за пределы возможности детального человеческий контроль. После расследования люди наконец понимают, что Дисфункции — это преднамеренные действия Машин. Потому что компьютер «Имейте в сердце благо человечества посредством подавляющей силы Первый закон робототехники» 10, они сами деобобщили Первый закон что позволило им допустить, чтобы небольшой вред причинился определенным людей, чтобы защитить большое количество вреда, причиняемого человечеству в качестве целое.

9 Азимов, Исаак. "Доказательство." Я РОБОТ. Нью-Йорк: Bantam Books, 2004. С.182.

10 Азимов, Исаак. «Неизбежный конфликт» . Я РОБОТ. Нью-Йорк: Bantam Books, 2004. Стр.201.

Развитие «Трёх законов робототехники» :

Дискуссия по мотивам сериала Азимова «Роботы» .

5.1. Создание Нулевого Закона

Этот деобобщенный Первый закон очень похож на Нулевой закон, который Азимов сформулировал в 1985 году в книге «Робот и империя» :

До создания закона Нулевого робот должен был защищать человека даже если выживание человечества в целом окажется под угрозой. Даже в На втором этапе способность суждения роботов была ограничена между человеческие личности.

В истории выше два робота – Дэниел и Жискар медленно собрать доказательства масштабного плана геноцида, разработанного Амадирос, используя ядерный усилитель, космический робототехник хочет сделать Землю непригодной для жизни люди и, следовательно, унаследовали Млечную Галактику. Роботы пытаются остановить план, но им мешает Первый закон робототехники, который не позволяет им от нанесения непосредственного вреда Амадирос.

Наконец, в этой срочной ситуации, робот Дэниел завершил Зерот.
Закон:

«Человечество в целом важнее, чем отдельный человек.

Есть закон, который важнее Первого закона: робот не может причинить вред человечества или своим бездействием допустить, чтобы человечеству был причинен вред» .

- Робот и Империя Айзека Азимова

5.2. Последствия создания Нулевого закона

Нулевой закон создает повышенную соучастность в принятии решений, потому что слово «человечность» — это просто абстрактное понятие. Это сложно для определяют человечество, и поэтому трудно решить, что для него хорошо. Как упоминается на втором этапе, достаточно сложно решить, какой именно человек может пострадать, тем больший вред. Еще сложнее сделать выбор между человека и человечества, когда даже не уверен, какой именно аспект человечество, с которым он может иметь дело. Оценка вероятностей относительно группами людей гораздо сложнее, чем с отдельными людьми.

Развитие «Трёх законов робототехники» :

Дискуссия по мотивам сериала Азимова «Роботы» .

Помимо вышеперечисленных проблем, создание Нулевого Закона также имеет серьезные последствия, касающиеся взаимоотношений между роботами и людьми.

Неся ответственность за защиту всего человечества, роботы чаще придерживались патерналистского отношения к людям. С роботы считают себя более умными и способными существами , способность человечества управлять своей судьбой подорвана.

В «Неизбежном конфликте» заложена основа доминирования роботов. уже заложено: «Машина ведет за нас наше будущее... Человечество потеряла право голоса в своем будущем» . 11 Эта идея получила глубокое развитие Последний роман Азимова о роботах – «Робот и империя» . Отличается от всех других ранее В написанных рассказах главными героями являются не люди, а два робота. Они пришел к выводу, что: «Тогда недостаточно выбирать, друг Дэниел. Мы должны уметь формировать, формировать желаемый вид, а затем защищать его. 12 Примеры о том, как роботы правят людьми, можно найти в более поздней части романа. с разрешения Нулевого Закона они манипулировали людьми, заставляя их действовать способ, который, по их мнению, решит проблемы. Между тем, они также сделали важное решение для человечества – покинуть Землю, позволив создание радиоактивной земли.

Нулевой закон — это символ превосходства роботов над человечество. Поскольку это последний шаг в разработке Законов, переход рабства роботов к доминированию роботов завершен. Что касается вопроса: «В чем разница между людьми и роботами?» Я надеюсь, что

бумага действительно отвечает на некоторые вопросы.

Однако, если истина все еще остается неясной, я процитирую слова доктора Кальвина. однажды сказал: «Вы с ними не работали, значит, вы их не знаете. Они более чистая и лучшая порода, чем мы...» 13

11 Азимов, Исаак. «Неизбежный конфликт» . Я РОБОТ. Нью-Йорк: Bantam Books, 2004. Стр.224.

12 Азимов, Исаак. Робот и Империя. Лондон: ХарперКоллинз, 1996.

13 Азимов, Исаак. Я РОБОТ. Нью-Йорк: Bantam Books, 2004. РХI .

Развитие «Трёх законов робототехники» :

Дискуссия по мотивам сериала Азимова «Роботы» .

6. Библиография.

1. Азимов, Айзек. "Доказательство." Я РОБОТ. Нью-Йорк: Bantam Books, 2004.
2. Азимов, Айзек. «Логово» . Я РОБОТ. Нью-Йорк: Bantam Books, 2004.
3. Азимов, Айзек. «Маленький потерянный робот» . Я РОБОТ. Нью-Йорк: Bantam Books, 2004.
4. Азимов, Айзек. "Бегать." Я РОБОТ. Нью-Йорк: Bantam Books, 2004.
5. Азимов, Айзек. «Мечты роботов» . Мечты роботов. Нью-Йорк: Ace Books, 1987.
6. Азимов, Айзек. Робот и Империя. Лондон: ХарперКоллинз, 1996.
7. Азимов, Айзек. «Неизбежный конфликт» . Я РОБОТ. Нью-Йорк: Бантам Книги, 2004.
8. Азимов, Айзек. «Ты помнишь о Нем» . Полноценный робот. Лондон: Гранада, 1982.
9. Портелли, Алессандро. «Три закона робототехники, The: законы текста, законы производства, законы общества» . Научно-фантастические исследования 7 (2): 150-156. Июль 1980 года.