

Три закона робототехники: вымысел и реальность

Мысль о том, что искусственное существо откажется подчиняться хозяину, давно тревожит учёных, писателей, мыслителей и простых обывателей.

Фантаст Айзек Азимов в 40-х годах XX века предложил решение этой проблемы в виде трёх законов робототехники, которые нужно «вшить» в искусственный мозг разумных машин.

Несложные правила накладывают на действия механических созданий ряд ограничений, поэтому человечество может отвлечься от размышлений о машинном бунте и смело шагать по дороге прогресса в сопровождении послушных и выносливых помощников.



Шло время, росла мощность вычислительной техники, появлялись и совершенствовались механизмы, разговоры об искусственном интеллекте выходили за пределы академических кругов. Но страх перед умными и своевольными машинами отнюдь не уменьшался. Азимовские законы стали всерьёз рассматривать учёные, которые разрабатывают машины с программным управлением. Кажется, что стоит понять, как донести их до роботов, и наступит новая эра, но так ли это?

Появление и формулировка

Многие предшественники Азимова описывали человекоподобных существ, которые созданы искусственно, как монстров с недобрыми намерениями. Немногочисленные авторы произведений о миролюбивых андроидах выделялись на общем фоне. Рассказы «Хелен О'Лой» Лестера дель Рея и «Я, робот» Отто Биндера вдохновили Азимова на работу над похожим произведением.

Результат не устроил Джона Кэмпбелла, редактора журнала «Astounding Science Fiction», где публиковался писатель: слишком велико было сходство с рассказом дель Рея. Азимов продолжал писать на эту тему и обсуждать её с редактором. Во время одной дискуссии Кэмпбелл изложил принципы, которые теперь весь мир знает как три закона робототехники. Эти правила работали в произведениях Азимова, издатель же просто указал на них писателю.

Законы робототехники предписывают машинам следующее:

1. Робот не может навредить человеку или позволить нанести ему вред в результате бездействия.
2. Робот обязан подчиняться командам человека за исключением тех, что не соответствуют Первому закону.
3. Робот должен следить за собственной сохранностью, если это не идёт вразрез с Первым или Вторым законом.

В произведениях Азимова эти принципы вносились в программное обеспечение почти всех машин. Если же происходило нарушение одного из правил, и робот это понимал, то его позитронный мозг получал серьёзное повреждение, случался так называемый «робоблок», из-за которого машина ломалась. Но художественную литературу любят за описание конфликтов, с которыми справляются действующие лица. Поэтому автор писал не об идеальных машинах, которые работают по безупречным алгоритмам, а о логических неувязках в законах. Нестандартные ситуации и вольные трактовки ведут за запретную черту, а вред от неукоснительного соблюдения законов порой превышает пользу.

Герои книг Азимова должны внимательно наблюдать за своими помощниками, а в случае беды разбираться в причинах, часто при участии специалиста-робопсихолога. Автор намеренно сделал три закона робототехники несовершенными, ведь иначе не получилось бы выстроить интересный сюжет. В ряде произведений он даже позволил роботам изменять строгие правила. Но целью писателя было не запугать читателей, а дать им возможность посмотреть на машины иначе. Азимов хотел, чтобы обыватели забыли о страхе перед механизмами, поняли, что те могут приносить пользу и стать людям верными товарищами. Писатель настаивал на том, что разработанные им принципы годятся не только для литературы, но и для воплощения в машинном коде.

Не навреди!

Жажда жизни перевешивает все остальные человеческие желания. Мощные механизмы с быстрой реакцией воспринимаются как явная угроза. Поэтому **Первый закон робототехники** ограждает людей от агрессии со стороны машин.

Если же человек попадёт в опасную ситуацию, верный робот спасет его. Здесь стоит забыть о двух других законах, ведь жизнь по умолчанию важнее, чем выполнение каких бы то ни было команд и исправность машины. Кажется, что здесь нет нестыковок и туманных толкований, но на деле всё обстоит куда сложнее.

Как понять Первый закон

Какой именно вред могут причинить роботы? Да, нападать на хозяев нельзя, но как быть, если автомат заденет человека нечаянно? Механизм подчиняется законам физики, он не может остановиться мгновенно, когда живой объект окажется рядом неожиданно. Каждую секунду мозг робота должен определять, как далеко находятся люди и с какой скоростью они перемещаются, а корпус робота должен быть усеян датчиками и камерами.

На обработку такого объема информации будет уходить слишком много времени и ресурсов машины, так что робот просто не выполнит свою основную задачу.

Следующий вопрос – кого считать человеком? Люди долго учились различать себе подобных, они поймут, где живой соплеменник, а где мертвый, где манекен, чучело или другой предмет, который напоминает человека. Научить этому компьютер очень сложно, ни один алгоритм распознавания не будет работать без сбоев, а цена ошибки – жизнь, и не одна. И что делать, если перед роботом совершается убийство или другое преступление против личности. А иногда спасти одного человека можно, если пожертвовать другим. Сложных ситуаций чересчур много, в них законы робототехники стали бы причиной поломки огромного количества роботов: машины попросту выйдут из строя, когда осознают, что не исполнили Первый закон. Ещё одно уточнение: что считать вредом для человека? Ведь травмы бывают не только физические: моральный вред причиняет людям не меньшие страдания, а общаться без нанесения обид умеют далеко не все люди, не говоря уже о машинах.

Перевести этические нормы людей на язык математических формул пока невозможно, поэтому ущерб от роботов нельзя предотвратить целиком. Эти проблемы рассматривал в своих

книгах Айзек Азимов, но литературные герои всегда находили выход из трудного положения. В жизни предусмотреть все пути развития событий не получится.

Обходной путь

Если роботам удавалось безнаказанно попирать правила в фантастической литературе, то в действительности всё ещё сложнее. Три закона робототехники противоречивы, их можно толковать по-разному в зависимости от обстоятельств, что неизбежно приведёт к нарушениям. Раз человек далёк от совершенства, то машины и программы, которые он создает, тоже будут с изъянами. Ошибки и дефектные детали приведут к сбоям, из-за которых машины будут вести себя не так, как планировалось. Люди порой ведут себя нелогично и отдают путанные и необдуманные команды. Всё это действует на руку злоумышленникам, которые захотят изменить поведение роботов ради наживы, устрашения, власти или забавы.

Но не только люди могут задуматься о том, как обойти законы робототехники. Мощный и гибкий искусственный интеллект, который непрерывно анализирует потоки данных и опирается не на жесткие алгоритмы, а на свой опыт, теоретически способен преодолеть ограничения.

Такой робот будет действовать вопреки догмам, которые придумали его слабые и несовершенные создатели. К тому же люди сами часто нарушают общественные и этические нормы, так они подают роботам не лучший пример для подражания. А существование военных беспилотников и других подобных изобретений само по себе наталкивает на вывод о том, что машинам дозволено убивать людей.

Нулевой закон робототехники

Азимов не ограничился тремя правилами, которые изложил в ранних рассказах своего знаменитого цикла. Нулевой закон робототехники появился позже, но писатель поставил его над остальными, так как хотел подчеркнуть значимость этого догмата. Вот его формулировка:

0. Робот не может навредить человечеству либо позволить, чтобы вред был нанесен человечеству в результате бездействия.

Забота обо всей человеческой расе, а не только об отдельных её представителях, – благородная и ответственная миссия. С этой задачей трудно справиться, если учесть, что Нулевой закон противоречит трём другим, особенно Первому.

Понять, опасен ли какой-то фактор для всех людей, непросто, а справиться с глобальной угрозой одному роботу или целой группе вряд ли по силам. Если же опасность исходит от самих людей, можно ли позволять машине лишать жизни тех, кого в норме нужно оберегать? В конце концов, неумолимая логика искусственного разума способна принять за угрозу то, чего люди не опасаются. Поэтому о пользе этого важного правила можно говорить с натяжкой.

Мнения экспертов

Разработчики искусственного интеллекта скептически относятся к применению трех законов робототехники на практике. Человеческая этика базируется на устаревших принципах, а как их использовать в создании программного кода, неясно. Человек по Азимову – творец, он ставит себя выше роботов, которых сознательно ограничивает в правах, хотя машины куда выносливее, быстрее и рациональнее людей.

Вероятно, взгляд с позиции хозяина искажает картину: если воспринимать роботов как бесправных рабов, то бояться восстания и накладывать запреты естественно.

Нужно также помнить, что искусственный разум, который обладает самосознанием, невозможно создать случайно, как это происходит в фантастике. Разработчики вряд ли рискнут

дать жизнь существу, которое совершает морально значимые поступки, но лишено этических принципов.

Вместо того, чтобы придумывать оковы для искусственного интеллекта, стоит углубиться в практику: создать прототип и поэкспериментировать с ним. О машинной этике лучше узнавать из реальных опытов, а не опираться на мысленное моделирование и страшные теории. Хотя пока рано говорить о разумных роботах, готовиться к их появлению только полезно.

Автоматизация проникает во все сферы жизни, поэтому думать о влиянии машин на человечество и даже планету нужно уже сейчас. Вероятно, законы робототехники не получится использовать в первоначальном виде, зато они могут лечь в основу стандартов, которые помогут людям и роботам жить в мире и трудиться на общее благо.

(Материалы интернета)