

Проблема создания этических кодексов для искусственных интеллектуальных систем

А.В.Разин

Философский факультет МГУ

Нормативные документы

- Манифест робототехники (Первый мировой симпозиум по робототехники. Санремо, Италия 2004).
- Мировая декларация робототехники (Фукуока, Япония 2004 г.).
- Нормы гражданского права о робототехнике. (Резолюция Европарламента от 16.02.2017).
- Этическое руководство для программирования автоматизированных систем вождения Германия 2017.
- Рекомендации ЮНЕСКО об этических аспектах ИИ

Опасность со стороны искусственного интеллекта.

- Сбои в работе искусственных интеллектуальных систем.
- Непредсказуемые реакции машины на неожиданные ситуации.
- Неспособность к определению приоритетов в конфликтных ситуациях.
- Трудности во взаимоотношении машины и человека (какие распоряжения человека машина должна выполнять).
- Вытеснение человека как несовершенного вида более совершенной машиной. Восстание машин.

Виннер об опасности робототехники

- Ноберт Виннер считает, что машины могут быть опасны для человека и непредсказуемы. При этом он отмечает, что даже понимая в деталях, как работает машина, оператор может не успеть понять, что ее рассуждения ведут к негативному сценарию или даже не успеть понять, что машина уже "приняла решение" и работает над осуществлением этого сценария.
- Wiener N. Some Moral and Technical Consequences of Automation // Science (80). 1960. T. 131. № 3410. С. 1355–1358.

Тьюринг о последствиях создания думающих машин

- Алан Тьюринг обсуждает постулат: "машины не могут делать ошибок" и отмечает, что в сложных машинах ошибки могут быть детерминированы неадекватностью исходных данных, хотя при этом машины максимально точно выполняют все математические операции по их обработке.
- Turing A.M. Computing machinery and Intellicence // Mind. 1950. T. 54. № 236. С. 433–460.

Негативные сценарии (Бостром)

- - человечество лишнее
- - свобода человека предельно ограничивается, вплоть до погружения людей в вегетативное состояние.
- - для решения некоторой задачи неоправданно тратятся все возможные ресурсы.

/Bostrom N., Yudkowsky E. The Ethics of Artificial Intelligence // Cambridge Handb. Artif. Intell. 2011. С. 1–20./

Человек в социальных сетях и искусственный интеллект

- Две большие области, которые можно обозначить как значимые для человека в связи с развитием искусственного интеллекта это:
- А) новые способы общения людей в социальных сетях их правила общения в сетях, правовые аспекты работы с информацией. Сети уже сейчас активны.
- Б) искусственный интеллект как самостоятельно действующий агент. Его отношение к людям и отношение людей к искусственному интеллекту.

Злонамеренное использование социальных сетей

- Понятно, что через социальные сети может распространяться спам, что будет блокировать способность функционирования некоторых сайтов, через них же может распространяться ложная информация, содержащая призыв к некоторым действиям политического, в том числе – экстремистского характера. Все это будет злонамеренным использованием социальных сетей, и понятно, что мы можем решительно осудить это в этическом плане.

Перегрузка социальных сетей

- Но и простое использование социальных сетей в развлекательном плане может загружать их работу и по существу – ограничивать заложенные в них информационно-коммуникативные возможности. Этот аспект проблемы тем более важно подчеркнуть, в связи с тем, что сети становятся все более и более умелыми, а обслуживающие их интеллектуальные системы в какой-то степени начинают проявлять свою субъективность.

Спам и запрещенный контекст

- Может быть социально опасный спам, скажем призыв к массовой эмиграции.
- Пропаганда насилия. Ненавистнические речи (необходимые ограничение свободы слова).
- Создание групп самоубийств (этому кстати говоря способствуют виртуальные игры, где всегда можно сохраниться, начать сначала. Создается впечатление, что так можно и в жизни).

Ограничение времени, проводимой в социальных сетях.

- В Китае в настоящее время подросткам разрешено проводить в социальных сетях, играя в виртуальные игры, по два часа в выходные дни.
- Сети создают для нас новый мир, который зачастую становится привлекательнее реального мира.
- Реальное общение заменяется виртуальным, при котором можно представить себя другой личностью.

Правила общения

- Важным этическим требованием в общении с искусственными интеллектуальными системами может быть требование четкого определения запускаемой в сеть информации, корректности терминологии, ясного понятийного выражения своих мыслей, нормативной правильности т.е учета ожиданий тех с кем вы общаетесь, в общем – все то, что предлагается в качестве правил проведения дискурса, обозначенных Ю.Хабермасом.

Этические проблемы робототехники

- -Разработка правил общения с искусственным интеллектом
- - Разработка правил работы искусственных интеллектуальных систем (например - беспилотных автомобилей)
- - Правила для вовлеченных в живое общение роботов. Сейчас чтобы можно было общаться с роботами как с живыми людьми осуществляется имитация наличия у них человеческих эмоций (роботы сиделки, роботы няньки). Но как несовершенство таких эмоций может сказаться на процессе общения людей?

Этические проблемы робототехники

- - Вопрос о возможности применения роботами насилия. Военные роботы, роботы полицейские.
- - Вопрос о формализации этической теории, приспособления ее для работы интеллектуальных систем. А также вопрос о том, на какую именно теорию мы можем или должны опираться абсолютистскую деонтологию или утилитарную теорию.
- - вопрос об этике суперинтеллектуальных искусственных систем (ИСИ).

Киборгизация человека

- - Вопрос о совершенствовании человека с помощью искусственных интеллектуальных систем, его киборгизации. В том числе - вопрос о радикальной переделке человеческого тела. Пределы такой перестройки.
- - Опасность культурного разрыва (Лемм)
- - Будет ли человек будущего человеком, или он будет киборгом, неизбежно теряющим какие-то человеческие качества?
- - Нужны ли здесь этические ограничения?

Этика отношения к искусственным интеллектуальным системам.

- а) отношение к роботам как к реальным субъектам, которые могут принимать самостоятельные этические решения, обижаться, проявлять реальные эмоции, уже не в плане имитации человеческого поведения, а как реальные основания, регулирующие их поведение.
- б) вопрос о возможности иллиминации таких систем, их выключения, собственно это вопрос о правах роботов.

Этика отношения к искусственным интеллектуальным системам

- в) вопрос об ответственности суперинтеллектуальных систем, формах их наказания и поощрения.
- г) вопрос о праве на ошибку, степени вреда, который может нанести обществу искусственная интеллектуальная система.

Противоречия между роботами и человеком

- е) вопрос о возможных противоречиях между человечеством и искусственными интеллектуальными системами, в том числе - вытеснения интеллектуальными системами человека как такового.
- ж) вопрос об изменении нашей жизни в связи с тем, что все наши мысли могут стать доступными другим людям (чтение мозга с помощью интеллектуальных систем). Какие социальные противоречия такая открытость может породить.

История вопроса

- Слово робот придумал Карел Чапек в его пьесе «Россумские универсальные роботы». В ноябре 1920 вышла отдельной книгой, а в 1923 г. была переведена на английский язык. В пьесе роботы осознают свое положение приспособленных для тяжелой работы машин как несправедливое, и уничтожают человечество, а потом они погибают и сами, так как не сохранились чертежи по которым они были сделаны.

Этимология слова «робот»

- Слово образовано не от всеохватывающего слова «работа» (по чешски — «práce»), а от чешского же слова «robota», что означает «повинность», «каторга», «тяжелая работа», «барщина», то есть более узкое понятие неоплачиваемого труда крепостных и рабов с отрицательной коннотацией.

Права роботов

- Этично ли мучить ИИ, если он способен чувствовать боль?
- Нормально ли оставить ИИ без общения надолго, если он способен чувствовать одиночество?
- А можно использовать его как домашнее животное? А как раба? А кто это будет контролировать это и как, если ИИ это программа, которая работает «живет» в вашем «смартфоне»?

Ответ на «поступки» роботов

- Второй класс этических проблем это реакции людей на действия роботов. Будет ли искусственный интеллект, обладающий сознанием и самосознанием иметь возможность нас обидеть. В законах Азимова говорится от том, что робот не может причинить вред человеку.
- А что собственно понимать под вредом? Если робот просто скажет: эй, ты глупый или эй ты, лысый, это будет вредом? Или даже – более мягко, - обвинит Вас в некомпетентности?

Беспилотные автомобили

- Этическое руководство для программирования автоматизированных систем вождения Германия 2017
- Лицензирование автономных автомобилей этически оправдано и даже становится этическим императивом в том случае, если они будут провоцировать аварии реже, чем люди.
- Главный приоритет — защита человеческой жизни. Если аварии не избежать, и приходится выбирать между имуществом и здоровьем, жизнью человека или животного необходимо сохранить здоровье и жизнь человека.

Беспилотные автомобили

- Когда автокатастрофа неминуема, необходимо минимизировать число травм. Жертвовать кем-то, исходя из его расы, пола, физического состояния и т. д., — недопустимо.
- В любой ситуации должно быть совершенно очевидно, кто управляет автомобилем и несёт ответственность за принимаемые решения: человек или компьютер.
- Информацию о том, кто находится за рулём, необходимо записывать и хранить.

Беспилотные автомобили

- Водитель сам должен решить, отправлять ли данные, полученные в процессе вождения, третьим сторонам.
- Использование автономных автомобилей должно стать частью «общего цифрового образования». Обращению с беспилотниками должны учить на курсах вождения.
- В документе подчеркивается, что вся концепция построена на базе принципов деонтологической этики.

Рекомендации департамента транспорта США

- Департамент транспорта в США выпустил методические рекомендации, в которых говорится, что перед тем, как выпускать беспилотную машину на рынок, производитель должен доказать, что продумал все этические вопросы до мелочей. Эти рекомендации вступили в силу в 2017 году, и это означает, что и разработчикам беспилотников и другим специалистам, в том числе – экспертам гуманитариям, прибавится работы в этой индустрии.

Особенности традиции разных стран

- Покидая пределы Германии и въезжая в Великобританию, должен ли будет владелец беспилотного автомобиля перестроить программу своего компьютера с деонтологической этики Канта на принципы утилитаристской этики Бентама и Милля?
- Насколько вообще пользователю будут доступны настройки компьютера беспилотного автомобиля?
- По-видимому здесь будут различаться базовые (неизменные) и относительные (выбираемые) принципы.

Возможности пассажира

- Допустим, пассажир все время нажимает кнопку «обгон», потому что спешит и стремится сэкономить ещё немного времени. А при очередном обгоне в машине случится сбой в алгоритме, и произойдет авария. Кто будет ответственен за эту аварию? Пассажир — за то, что поставил машину в такой режим? Производитель — за то, что недоглядел при программировании, не учёл и не поставил защиту от подобных ситуаций?

Этические приоритеты

- Нужно ли программировать машины так, чтобы они в первую очередь думали о жизни пассажиров или пешеходов? Или, может, нужно запрограммировать так, чтобы они распределяли травмы на всех: например, ребенок выбежал на дорогу, и вы резко тормозите, и ребенка все-таки сбиваете, но не насмерть, ну и сами пострададете. Зато все живы. Получается, прежде чем программировать подобные ситуации, нужно на каких этических принципах будет строиться поведение машины-робота.

Деонтоология

- Делает акцент на безусловном выполнении нормы, долге, который выступает в качестве самостоятельного мотива поведения, вне связи с пользой, потребностью и т.д.
- Допускает высокую степень формализации, что удобно для программирования машины.
- Но, упрощает процедуру принятия решения, что не всегда хорошо в смысле оптимальности, учета всех обстоятельств.

Утилитаризм

- Утилитарная концепция это так называемая консиквенционалистская этика. В ней решающим является результат поступка (возрастание блага), а не мотив, направленный на безусловное выполнение правил.
- Допускается подсчет пользы и вреда. Максимизация первого и минимизация второго.
- Допускается жертва интересами меньшинства.
- Утилитаризм вызывает вопросы в связи с обеспечением базовых прав человека, неопределенностью утверждаемых приоритетов.

Абсолютизм

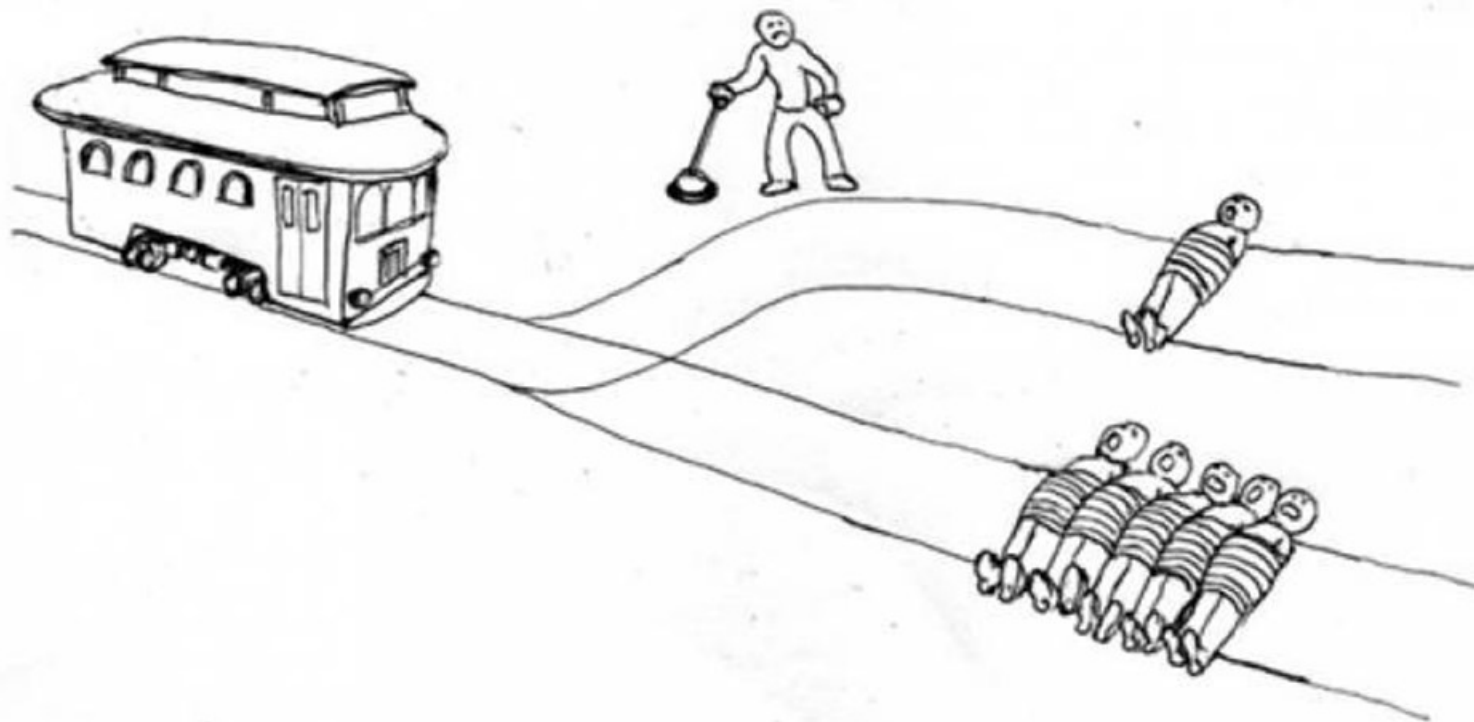
- Но и абсолютизм содержит в себе противоречия и ограничения.
- Безусловное подчинение абсолютному когда оно не подвергается сомнению, когда сама возможность обоснования моральной нормы тоже подвергается сомнению, содержит в себе отчуждение.
- Чтобы обеспечить моральную ответственность по словам Р.Хеара надо быть не только ангелом, но и архангелом, т.е. действовать не просто из благих намерений, но исходя из целостного видения ситуации.

Этика добродетелей

- Далее я хочу подчеркнуть, что моральный выбор это всегда непростое решение, требующее усилий от личности, связанное с глубокими размышлениями. Это отличает его от ситуации выбора, связанного с исполнением правовых норм, подчинением традициям, обычаям, этикетным правилам и т.д.
- Даже исполнение простых норм типа не убий, не укради, не лги связано с осмыслением возможности исключения.
- Осуществление выбора связано с моральными качествами субъекта – добродетелями.

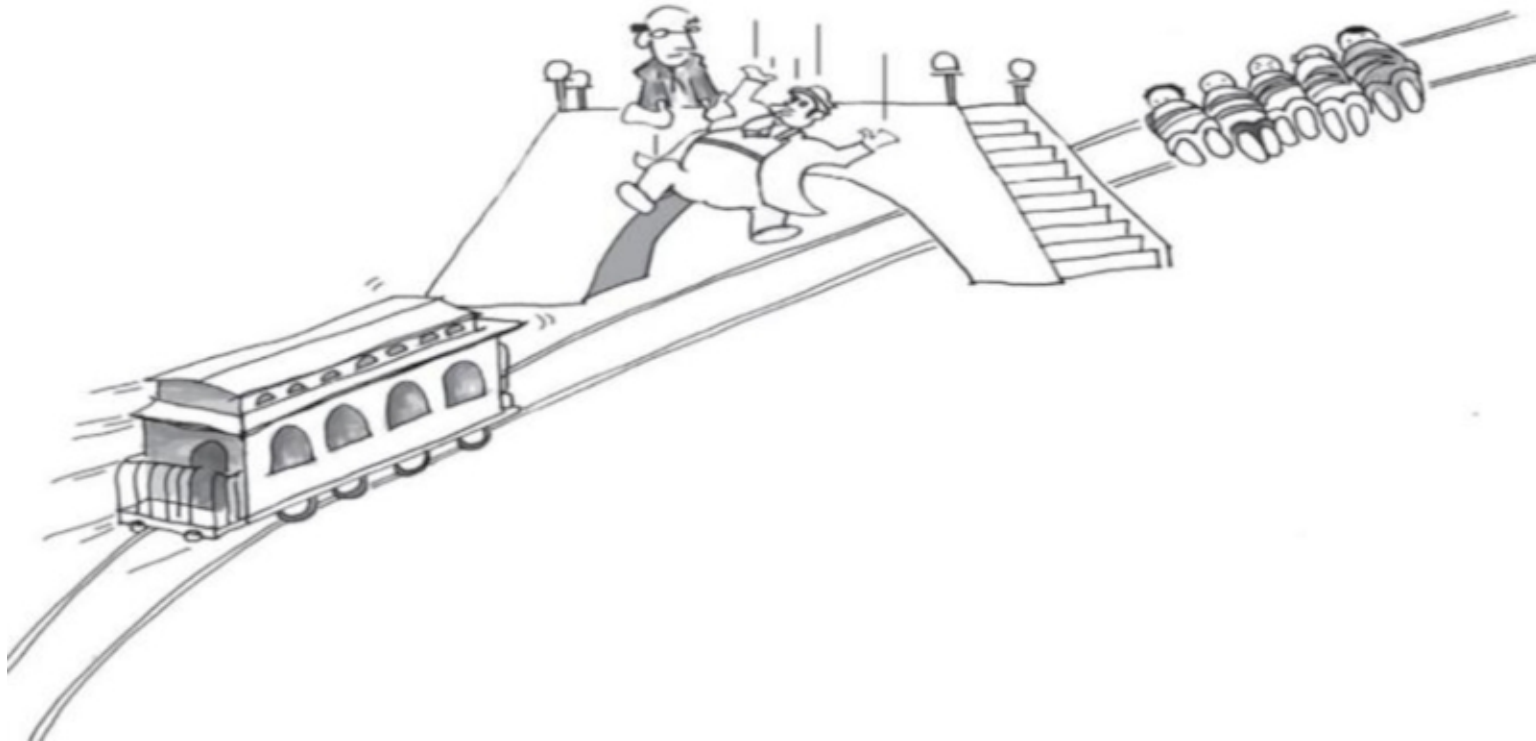
Вагонеткология

Вагонеткология (эпизод 1)



Вагонеткология

Вагонеткология (эпизод 2)



Движение по выделенной полосе

- Какой общий вывод я бы сделал из рассмотренных вопросов это вывод о том, что **там, где искусственный интеллект, если иметь ввиду ситуации на дорогах, - не должно быть людей.** То есть должны быть трассы, где ручное управление будет запрещено. Эти трассы должны быть огорожены сетками, заборами, так чтобы ни животные, ни люди не могли туда попасть. Пока я видел такую дорогу только в Польше между Жешувом и Краковом.

Моральная машина

- Пределы теоретического понимания морали обращают нас к практике.
- В этике существуют возражения против теоретического понимания морали, против процедуры ее обоснования (Шопенгауэр)
- Моральная машина - онлайн-платформа, разработанная группой масштабируемого сотрудничества Ияда Рахвана в Массачусетском технологическом институте, которая порождает **моральные** дилеммы и собирает информацию о решениях.

Институциональная этика

- «Этика добродетелей, связанная по преимуществу с мотивами поведения, сохраняет важное (быть может, даже возрастающее значение) в области личных отношений и во всех ситуациях, имеющих ярко выраженный личностный, индивидуализированный характер, т.е. говоря обобщенно, в зонах личностного присутствия. В системном (общественно-функциональном, профессионально-жестком) поведении она дополняется институциональной этикой».
[Гусейнов А.А. Философия, мораль, политика 2002. 123.]

Постконвенциональная этика

- «...Приходится учитывать, что многие вопросы этики иначе решаются на конвенциональном и постконвенциональном уровне развития морали, когда многие дилеммы уже дискурсивно закрыты и субъектам деятельности остается лишь быть «законопослушными» (Compliance) и следовать предварительно принятым ценностям и нормам... А если кому-то правила, нормы или стандарты не нравятся, то их совместно меняют, а не нарушают».
- Васильев Н. Управление ценностями – смена парадигм этики.//Облики современной морали. М.: МАКС Пресс, 2009. С. 205.

Особенности этики для роботов

- Для роботов не будут иметь значения многие стороны человеческой морали. Скажем то, что определяется как так называемые «мещанские добродетели» - трудолюбие, бережливость, умеренность.
- Но требования, которые приобретают принципиальное значение во многих областях прикладной этики сохраняться. Например, compliance.
- Очевидно для контроля исполнения кодексов будут созданы специфические этические комиссии, комитеты.

Кодексы робототехники.

Законы Азимова

- 1. Робот не может причинить вред человеку или своим бездействием допустить, чтобы человеку был причинён вред.
- 2. Робот должен повиноваться всем приказам, которые даёт человек, кроме тех случаев, когда эти приказы противоречат Первому Закону.
- 3. Робот должен заботиться о своей безопасности в той мере, в которой это не противоречит Первому или Второму Законам.

Четвертый или нулевой закон робототехники

- 4. Робот не может нанести вред человечеству или своим бездействием допустить, чтобы человечеству был нанесен вред».
- Но что значит не причинить вред человечеству бездействием. А если для этого надо убить несколько людей, или даже целый город, в котором распространяется опасная инфекция?
- Что если будет изобретено радикальное средство продления жизни и это вызовет напряжение в связи с ростом населения. Надо ли будет убить изобретателя?

Какие распоряжения и какого человека надо выполнять?

- О какой временной перспективе идет речь? Как это связано с принципом разумной предосторожности, нанесения вреда экологии?
- Как это соответствует условиям свободной конкуренции? Не распространя какое-то ноу-хау, я в общем-то причиняю вред человечеству.
- Обладает ли робот способностью оценить вред?

Бен Герцель (Бразилия), главный научный сотрудник компании Aidyia Holdings

- Думаю, что тип роботов, которых предвидел Азимов, станет возможным в недалеком будущем. Тем не менее, в большинстве своих вымышленных миров, писатель предполагал, что человекоподобные роботы будут вершиной робототехники и инженерии искусственного интеллекта. Это вряд ли. Очень скоро, после достижения статуса азимовских роботов, станет доступным и создание искусственного сверхинтеллекта и сверхроботов».

Луи Хельм

Институт исследований машинного интеллекта
(MIRI – представительства в США и Канаде)

- «Честно говоря, я не нахожу никакого вдохновения в этих трех законах робототехники, — говорит Хельм. — Суть машинной этики в том, что они не удовлетворяют базису машинной этики. Возможно, три закона робототехники широко известны, но на самом деле использовать их в качестве основы для программирования бесполезно».

Недостатки законов Азимова

- Хельм отмечает следующие недостатки законов:
- Состязательны по своей сути
- Основаны на изжившей себя этической теории (деонтологии)
- Не работают даже в фантастике
- Герцель соглашается: в реальности эти законы работать не будут, поскольку термины с их участием неоднозначны и остаются предметом толкования — а значит, крайне зависимы от тех, кто делает переводы.

Шовинизм по отношению к роботам

- Другой аспект (и потенциальный недостаток) трех законов в очевидном шовинизме — предположение о том, что роботы должны оставаться, несмотря на их превосходящую мощь, в подчинении у человека, человеческих потребностей и приоритетов.

Луи Хельм

- «Я думаю, было бы неразумно проектировать систему искусственного интеллекта или робота с самосознанием. И в отличие от фильмов или книг, в которых создатели искусственного интеллекта «случайно» приходят к разумным машинам, я не верю, что в реальной жизни это может случиться. Для этого понадобится слишком много усилий и знаний. И большинство разработчиков ИИ — этически подкованные люди, поэтому они будут избегать создания того, что философы называют «морально значимыми существами».

Бен Герцель

- «Определение набора этических заповедей в качестве ядра машинной этики будет безнадёжным, если в основе машины будет гибкий общий искусственный интеллект. Если он будет задуман как интуитивный, гибкий, адаптивный или этический — в этом контексте этические заповеди будут полезны для системы только как грубый ориентир для применения собственной этической интуиции. Но в этом случае заповеди не станут основой этической системы, а только лишь аспектом».

Киборгизация человека

- Джозефа Карл Робнетт Ликлайдер, один из идейных отцов всемирной сети, в статье «Симбиоз человека и компьютера» (1960) предсказывал связь человеческого интеллекта с искусственным. Сегодня это становится очевидным, но последствия такого симбиоза неоднозначны. Особенно настораживает опасность утраты человеком определённых ментальных навыков и способностей.

Киборгизация человека

- Нейсбит в книге Мегатренды показал, что тесты интеллектуальности показывают неуклонное снижение интеллектуальных способностей студентов ведущих американских университетов. Таких как Харвард и Принстон.
- В США 10 млн. неграмотных. Многие школьники в таком ведущем интеллектуальном штате как Калифорния не заканчивают средней школы. Более 35 млн. не могут в уме сложить два двухзначных числа.

Роботы убийцы

- Сейчас создаются военные роботы.
- Недавно по этому поводу разгорелся скандал. Компания Google разрабатывает программное обеспечение, используемое для пилотного военного проекта Project Maven по управлению дронами. Предположительно, в будущем это может привести к созданию полностью автономного оружия.

Роботы убийцы

- 12 сотрудников Google уволились в знак протеста, еще 4000 подписали петицию с просьбой отказаться от контракта с военными. Более 1000 видных ученых в области ИИ, этики и информационных технологий написали открытое письмо с просьбой к Google прекратить работы над проектом и поддержать международный договор по запрету автономного оружия.

Протесты против проектирования роботов, способных на насилие

- Эксперты ООН обсуждают эти вопросы на правительственном уровне, там запущена **кампания «Остановите роботов-убийц»**, которую инициировал Международный комитет контроля за боевыми роботами (Великобритания), объединившийся с Human Rights Watch и Amnesty International.

Роботы и человеческое общение

- **Ноэль Шаркли**, профессор робототехники из Университета Шеффилда (Великобритания) говорит, что неконтролируемый рост числа роботов может привести к недостатку общения между людьми. Это касается социальных роботов, способных общаться с человеком, воспитателей, медперсонала, но в принципе всех роботов, заменяющих людей. Но особенно негативно это отразится на детях. Дети, воспитанные роботами, будут мало приспособлены к жизни. Мы рискуем получить новых Маугли.

Тотальный контроль за поведением людей

- В статье, озаглавленной «2084 год: Большой робот следит за тобой» **Ноэль Шаркли** предсказывает становление нового мира, в котором задачи наблюдения, обеспечения безопасности и охраны правопорядка будут в большей степени переданы искусственному интеллекту.
- Но в работе искусственной системы могут быть сбои.
- Интересно обсуждается в рассказе Роберта Шекли «Страж птица».

Чтение мозга

- Технология основана на фиксации увеличения кровотока, к нейронам, возбуждающимся при представлении образов.
- Доктор **Камитани** говорит: «Эта технология также может применяться в других сферах, не только в распознавании (в голове) зрительных образов. В будущем она, возможно, сумеет выявлять чувства и сложные эмоциональные состояния»

Контроль за мыслями и эмоциями

- Будет ли открытость наших мыслей другим опасной для социального бытия людей, будет ли это сдерживать их фантазии, или же, наоборот, в такой открытости общество будет видеть нечто позитивное, так как оно получит возможность обезопасить себя от замыслов террористов, от рискованных экспериментов, идеи которых вынашиваются в головах безответственных ученых и т.д.

Первый международный симпозиум по робототехнике

- На «Первом международном симпозиуме по робоэтике» (Санremo, Италия, 2004), по его итогам, был заключен «Манифест робоэтики», некоторые ученые по-прежнему высказывали убеждение, что этические проблемы вообще нерелевантны для робототехники, а разработчики роботов не могут нести морально-этическую ответственность за свои научные труды, поскольку занимаются исключительно технической стороной вопроса, хотя никто и не отрицал, что влияние роботов на человека должно быть безопасным

Манифест робототехники

- Главная задача ИСИ [Искусственный Сверх Интеллект] - охрана морали [Мораль в данном случае - общественно полезное состояние или действие, выражение альтруизма и минималистичный для выживания эгоизм] путем математически верных рабочих моделей и максимальной валидации [валидация - цифровая проверка] путем умных контрактов [умный контракт основан на технически осуществимой проверке заказа и его исполнения, без участия человека, и без возможности юридического оспаривания].

Мировая декларация робототехники

- Важность этической составляющей в робототехнике отмечена и в «Мировой декларации робототехники» (Фукуока, Япония, 2004), в которой сделан акцент на том, что каждое следующее поколение роботов должно способствовать улучшению жизни человека, реализации безопасного и мирного общества.

Нормы гражданского права о робототехнике.

Резолюция Европарламента от 16.02.2017.

- Подчеркивает, что система руководящих этических принципов должна быть основана на принципах «делай благо» и «не навреди», а также на принципах автономности и справедливости, принципах и ценностях, указанных в Статье 2 Договора о Европейском Союзе и Хартии по правам человека (таким как уважение человеческого достоинства, равноправие, справедливость, недискриминация, право на информированное согласие, неприкосновенность частной и семейной жизни, право на защиту данных).

Инициатива IEEE IEEE (Institute of Electrical and Electronics Engineers)

- Первым документом стал проект общих рекомендаций для разработчиков ИИ, в котором говорится о том, как разработчикам ориентироваться на этические проблемы в процессе разработки своих продуктов [IEEE, 2016]. Название документа: "Ethically Aligned Design" ("Этически обусловленное проектирование").
- Аналогичные вопросы рассматриваются, в отчете Юнеско, посвященном этике роботов и озаглавленном "Report Of Comest On Robotics Ethics"

Рекомендации ЮНЕСКО об этических аспектах ИИ

- В конце 2021 года генеральная конференция ЮНЕСКО утвердила рекомендацию об этических аспектах ИИ
- Уважение, защита и поощрение прав человека и основных свобод и человеческого достоинства.
- Благополучие окружающей среды и экосистем.
- Обеспечение разнообразия и инклюзивности.
Система обязана соответствовать всем правовым нормам, но не должна ограничивать возможность человеческого выбора.

Рекомендации ЮНЕСКО об этических аспектах ИИ

- Жизнь в мирных, справедливых и взаимосвязанных обществах.
- Соразмерность и не причинение вреда.
Деятельность ИИ-систем не должна выходить за рамки того, что требуется для достижения законных целей.
- Безопасность и защищенность
- Справедливость и отказ от дискриминации
- Устойчивость.
- Право на неприкосновенность частной жизни и защита данных.

Рекомендации ЮНЕСКО об этических аспектах ИИ

- Подконтрольность и подчиненность человеку. ИИ-система для принятия решений никогда не сможет заменить человека в качестве конечного субъекта ответственности
- Прозрачность и объяснимость. Человека необходимо информировать о том, какие решения были приняты при помощи ИИ, а также при необходимости понятно их обосновать.

Рекомендации ЮНЕСКО об этических аспектах ИИ

- Ответственность и подотчетность. Люди, ответственные за ИИ-системы, обязаны уважать, защищать и поощрять права и свободу человека, содействовать охране окружающей среды, а также предусмотреть механизмы контроля для проверки и аудита нарушений ИИ-систем.
- Осведомленность и грамотность. Необходимо улучшать понимание гражданами ИИ-технологий .
- Управление на основе международного права.

Сомнительные положения

- Ст. 92 «...следует поощрять гендерное разнообразие в сфере связанных с ИИ научных исследований ... посредством предоставления девушкам и женщинам льготного доступа к данной области деятельности ...» и 93 «...содействовать созданию репозитория передового опыта в области стимулирования участия женщин, девушек и недостаточно представленных групп населения во всех этапах жизненного цикла ИИ-систем...»

Кодекс этики в сфере искусственного интеллекта

- В России ряд ведущих российских компаний (лидеры в разработке технологий) утвердили первый национальный кодекс этики. Кодекс устанавливает общие этические принципы и стандарты поведения, которым следует руководствоваться всем, кто занимается созданием, внедрением или использованием технологий ИИ.
- Подписан на I международном форуме «Этика искусственного интеллекта (ИИ): начало доверия» 19.10.2021.

Бойкот искусственному интеллекту

- Партия «Единая Россия» увидела риски распространения и развития в России искусственного интеллекта, пишет РБК со ссылкой на двух неназванных членов партии.
- Предлагается запретить государству передавать ключевые функции и право принятия решения от человека искусственному интеллекту и цифровым платформам.
- Запрет касается образования, медицины, юриспруденции, вопросов собственности, права на труд.

Обсуждение цифрового манифеста

- 19 августа 2021 г. состоялось обсуждение «Цифрового манифеста» – стратегического документа, отображающего цифровое будущее страны.
- В ходе мероприятия лидер «Совета матерей» Татьяна Буцкая предложила не забывать о реальных запросах граждан при формировании документа. «Стратегия цифровизации обучения и повседневной жизни, безусловно, облегчит и сократит временные затраты наших граждан на посещение многочисленных инстанций, – отметила Буцкая.

Разделы кодексов

- В сфере прикладной этики достаточно хорошо разработаны методики составления кодексов.
- Обязанности перед обществом, перед клиентами, перед работодателем. Перед коллегами по профессии или перед корпорацией. Но как это будет выглядеть для роботов?
- Какие-то разделы могут сохраниться, но вряд ли неизменным останется раздел обязанности перед коллегами.
- Добавятся новые разделы, скажем обязанности перед производителем.