

Этика искусственного интеллекта: философия, право или политика?

Огородов Д.В., к.ю.н. член рабочей группы по этике искусственного интеллекта инициативы EAD IEEE, член РАИИ

исследовательский семинар
«Этические проблемы искусственного интеллекта»,
заседание 9 апреля 2021

Основные вопросы доклада

1. Основные подходы к пониманию проблематики, обозначаемой формулой "этика искусственного интеллекта»
2. Влияние государства, гражданского общества, бизнеса, науки на создание и применение интеллектуальных систем
3. Промежуточный итог дискуссий российских исследователей за 2017-2020 год

Промежуточный итог дискуссий российских исследователей за 2017-2020 год

Консенсус среди исследователей не сложился.

Существуют, как минимум, два разных подхода к пониманию проблематики, обозначаемой формулой «этика искусственного интеллекта»:

- 1) «этика искусственного интеллекта» как этика в буквальном понимании, **принципиально новая философско-этическая и правовая социальная проблематика** (В.Э. Карпов)
- 2) «этика искусственного интеллекта» - это условное обозначение **традиционной политико-правовой проблематики** создания и применения технических систем. Слово «этика» здесь для российского читателя «ложный друг переводчика»: рассматриваемая проблематика родилась в США с их правовой и политической спецификой – большая роль гражданского общества, высокий вес саморегулирования (Д.В. Огородов)

«Этика искусственного интеллекта» как принципиально новая социальная проблематика. Подход конструктора А/ИС (В.Э. Карпов)

- подход вытекает из успешного применения при конструировании А/ИС (прежде всего, групповой робототехники) технических приемов имитации некоторых черт психики живых организмов и квазисоциальных сообществ животных.
Например, технические аналоги темперамента животных - *«агрессивность»*, *«флегматичность»* робота-анимата - могут эффективно использоваться для построения эффективного управления роем (группой) роботов
- терминология информатики (*«социальные сообщества роботов»*, *«эмоции и темперамент роботов»*, *«поведение роботов»*, *«решение, принимаемое технической системой»*) переносится на философскую и правовую социальную проблематику
- *«моральность поведения робота»* предлагается понимать буквально, оценивая функционирование машины человеческой меркой морали. Стало быть, *«этика искусственного интеллекта»* - философская проблематика
- ожидаемый практический результат – обращение конструктора А/ИС к заведующему кафедрой этики философского факультета: *«Мне нужна справка о том, что моя техническая система функционирует этично!»*. Заведующий выдает конструктору справку

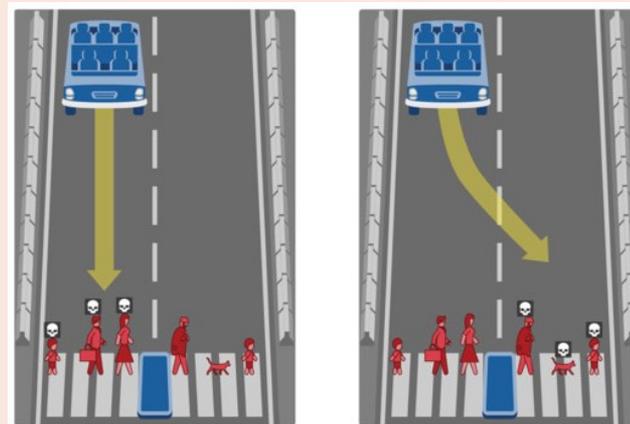
О чем обычно идет речь

Этические проблемы ИИ:

- Опасность для человечества
- Социальные и правовые проблемы
- Этические аспекты применения систем ИИ
- Этика самих специалистов в области ИИ
- ...

При этом:

Грядущая неизбежная стандартизация систем ИИ (IEEE) =>???



<http://moralmachine.mit.edu/>

«Проблема вагонетки»: сделать выбор: разбить автомобиль, убив пассажиров, или задавить прохожих на переходе.

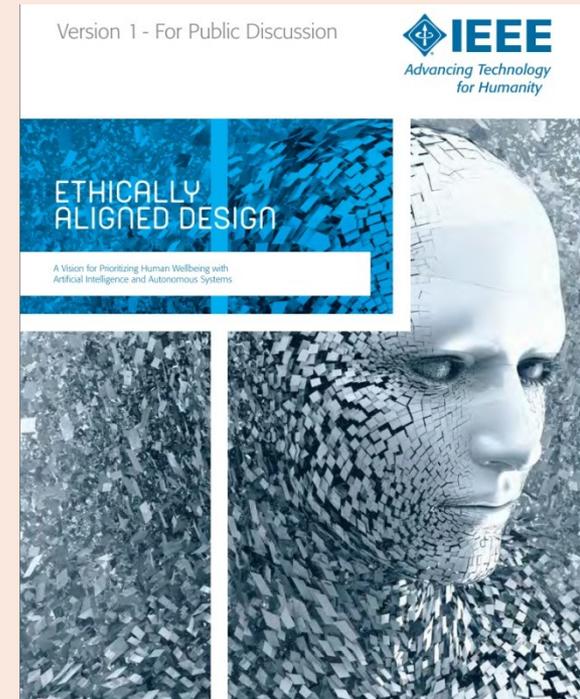


Новая постановка задачи

Стандартизация =>

Ответ на вопрос: *насколько соответствует система ИИ этическим нормам*

IEEE: Ethically Aligned Design



Стандарты IEEE, основанные на *Ethically Aligned Design*

- IEEE P7000™ – *Model Process for Addressing Ethical Concerns During System Design*
- IEEE P7001™ – *Transparency of Autonomous Systems*
- IEEE P7002™ – *Data Privacy Process*

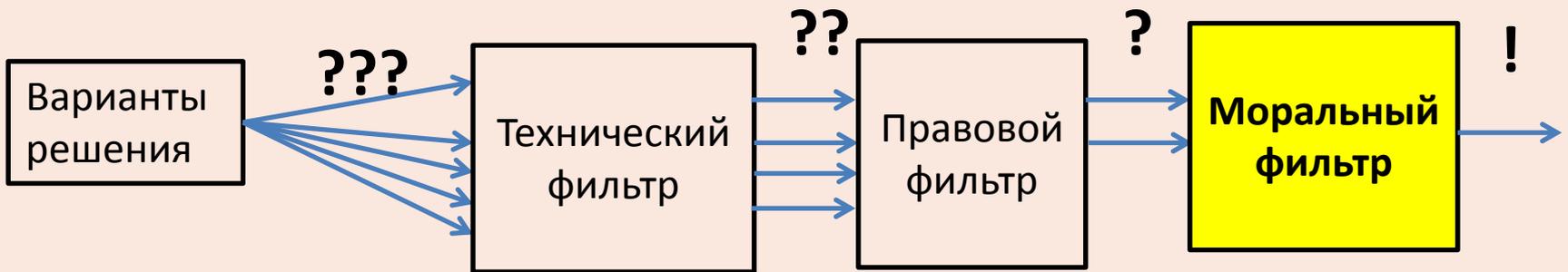
Понимание проблемы этики и ИИ

Основное отличие проблематики этики в ИИ от этики прочих областей:

1. ИИ-система – это система, автономно принимающая критические решения.
2. Основная проблема – определение того, насколько принимаемые И/АС решения соответствуют этическим нормам, т.е. насколько она «этична».

Механизмы оценки решения

$$\text{Оценка (Д)} = \text{Техническая_оценка(Д)} + \text{Правовая_оценка(Д)} + \text{Моральная_оценка(Д)}$$



Карпов В.Э Применимо ли понятие морали к отношениям между искусственными агентами? / ИФ РАН, семинар «Этические проблемы ИИ», 17.12.2019

Подход законодателя, судьи и юриста IT-компания к проблематике этики ИИ (Д.В. Огородов)

Каждая отрасль науки и техники **может использовать свою терминологию**, сколь угодно оригинальную, поэтическую и даже отсылающую к человеческим качествам. Например, в физике элементарных частиц используются слова «*странность*» и «*очарование*» для описания свойств элементарных частиц (впрочем, без ущерба для дела можно заменить словами, не связанными с психологическими отсылками к личности человека, например, назвать омега-температурой частицы).

Кроме того, технике давно известен подход заимствования технических решений из живой природы – бионика

Юриспруденция не препятствует использованию другим наукам их собственной терминологии. Однако не следует отождествлять наименование или существо технического приема решения технической же задачи с его правовой характеристикой.

Только **поведение человека** и основанные на нем общественные отношения людей **являются предметом правового регулирования и юридически значимых моральных оценок** (например, ст. 151, ст. 168 ГК РФ, ст. 107, ст. 113 УК РФ)

Феномены же, именуемые в информатике «*принятием решения*» или «*поведением* (технической системы)» для закона представляют собой лишь **функционирование техники; здесь нет поведения** как предмета моральной и правовой оценки. Это касается и неживых вещей в виде роботов-аниматов и вещей живых - животных.

Антропологизация вещей и программного обеспечения **не имеет практического политико-правового смысла** («*подушки безопасности выполнил свой долг, пожертвовав собой ради пассажира*», «*воспитанной и предупредительный дверной доводчик*»)

Подход законодателя, судьи и юриста IT-компаний к проблематике этики ИИ (Д.В. Огородов)

«Более моральна ставка НДС 18% или 20%? Более этично напряжение в розетке 220 или 110 вольт? Сколько метров должна составлять длина тормозного пути автомобиля, с точки зрения нравственности?» - такие вопросы в юридической практике не возникают. При наличии обязательного стандарта к технике незачем требовать от профессора кафедры этики некую дополнительную справку

В области техники не существует двойного этического и правового регулирования, которые были бы самостоятельны до степени противопоставления друг другу.

Этическое (философское), политическое и правовое неотделимы друг от друга в практической деятельности государства и IT-бизнеса, производящих и эксплуатирующих технические системы

Этические вопросы ставятся и разрешаются **в ходе политической дискуссии** и воплощаются **в правовую форму закона** – при нормотворчестве высшей юридической силы

Главный вопрос для производителей и эксплуатантов техники, за которыми присматривают контролирующие государственные органы и суд: *«Обязательна ли такая-то норма закона / технического стандарта или необязательна? Соответствует ли конкретное изделие стандарту или нет?»*

Подход законодателя, судьи и юриста IT-компаний к проблематике этики ИИ (Д.В. Огородов)

На одном из обсуждений проекта рекомендаций ЮНЕСКО по вопросам этики ИИ представитель Сбербанка А.Незнамов предложил сосредоточиться на этических аспектах, поскольку название рекомендаций указывает на этику, а (1) нормы этики, (2) нормы права и (3) технические нормы являются разными системами норм, которые нельзя смешивать.

Подобный подход трудно поддержать:

- во-первых, выражение "этика ИИ" описывает проблемы, значительная часть из которых **уже урегулирована нормами российского права**. Этику ИИ отнюдь не следует отождествлять с кругом якобы нерешенных в праве сугубо умозрительных вопросов морали, как это представлено в «дилемме вагонетки» в проекте <http://moralmachine.mit.edu>
- во-вторых, в ряде случаев нормы морали и моральная сторона деяний неотделима от права и правовой характеристики, а закон подчас ставит те или иные правовые последствия в зависимость от моральной характеристики поведения лица (например, ст. 151, ст. 168 ГК РФ, ст. 107, ст. 113 УК РФ). При этом разрешение таких вопросов морали – компетенция суда и (как составная часть вопросов права, а не факта) **не требует назначения философских экспертиз**
- во-третьих, в современном российском правовом порядке трудно противопоставить технические нормы нормам правовым. Хотя технические нормы суть требования к технике, ряд из таких норм сегодня выражен в форме федеральных законов - технических регламентов.

Всё вышесказанное в совокупности подтверждает предложенный нами подход к пониманию словосочетания "этика ИИ", обозначающего, прежде всего, **политико-правовую проблематику создания и использования технических систем**

Подход законодателя, судьи и юриста IT-компания к проблематике этики ИИ (Д.В. Огородов)

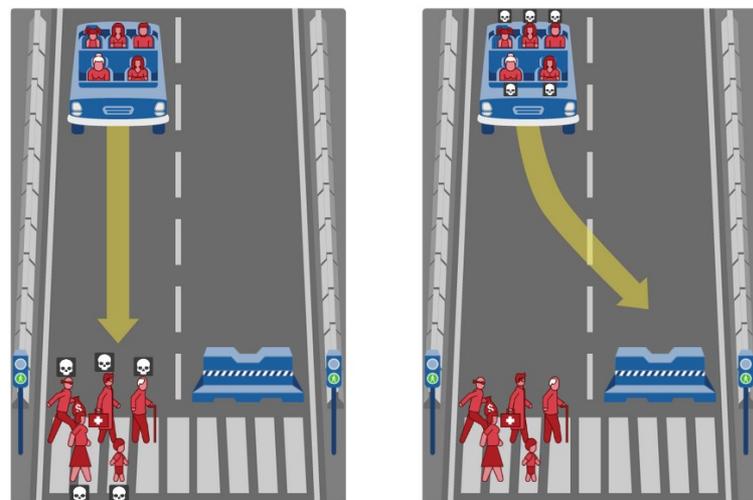
В рамках правовой системы РФ заметная часть вопросов этики ИИ имеет прямое юридическое значение и уже решена.

В проекте «Машина морали» была смоделирована моральная дилемма траектории беспилотного автомобиля при неожиданном появлении препятствия:

- продолжать движение прямо, жертвуя пассажирами ради спасения пешеходов;
- либо свернуть в сторону и пожертвовать пешеходами ради спасения пассажиров. Автопроизводитель Mercedes-Benz в 2016 году намеревался решить эту дилемму в пользу пассажиров, позже дезавуировав свое заявление под шквалом критики.

Однако данная моральная дилемма уже решена в уголовном законе РФ:

- команда разработчиков беспилотного автомобиля – в котором его создателями предreshen вопросы жизни и смерти той или иной категории других людей (например: жить пассажиру, умереть пешеходу или наоборот) – **несет уголовную ответственность за убийство, совершенное организованной группой**, что наказывается лишением свободы на срок от восьми до двадцати лет с ограничением свободы на срок от одного года до двух лет, либо пожизненным лишением свободы, либо смертной казнью (п. «ж» ч. 2 ст. 105 УК РФ).
- такое преступление будет считаться совершенным **с косвенным умыслом**, поскольку указанные лица осознавали общественную опасность своих действий (бездействия), предвидели возможность наступления общественно опасных последствий, **не желали, но сознательно допускали эти последствия либо относились к ним безразлично** (ст. 25 УК РФ).



<http://moralmachine.mit.edu/>

Влияние государства, гражданского общества, бизнеса, науки на создание и применением А/ИС

Государству и обществу интересен весь спектр вопросов, связанных с созданием и применением А/ИС, включая последствия для рынка труда и занятости. Для государства и общества неприемлем ограничительный - «конструкторский» - подход к проблематике «этике ИИ»

Недопустимо **перекладывать связанные с А/ИС риски** создателей и эксплуатантов, в чьих интересах создается и используется новая техника, **на третьих лиц (общество, граждан)**

Интересы граждан и общества в целом – сохранение, неуменьшение достигнутого уровня гарантий прав и законных интересов в связи с приходом новой техники. Именно это определяет круг **насуточных социальных задач**, охватываемых словесной формулой «этика ИИ»

Такие гарантии могут быть достигнуты **только в правовой форме** законов, стандартов и иных источников общеобязательных правил. Саморегулирование бизнес-сообществом дополняет общеобязательные рамки (следует учесть опыт СРО в строительстве и иные известные проблемы саморегулирования в РФ)

Упование на необязательную мораль разработчиков и эксплуатантов А/ИС - **наивно**. Пример Сбербанка: его корпоративные *Принципы этики ИИ* vs. его соглашение с ФСИН

Справка конструктору от заведующего кафедрой этики: **недопустима передача политико-правовых задач регулирования на откуп** IT-бизнесу или ученым. Роль технических (информатика), юридических и других наук – поддержка научно-обоснованного технического регулирования в области создания и применения А/ИС

Вопросы?

d.ogorodov@gmail.com

*Огородов Д.В. к.ю.н. член рабочей группы по этике искусственного интеллекта
инициативы EAD IEEE, член РАИИ*