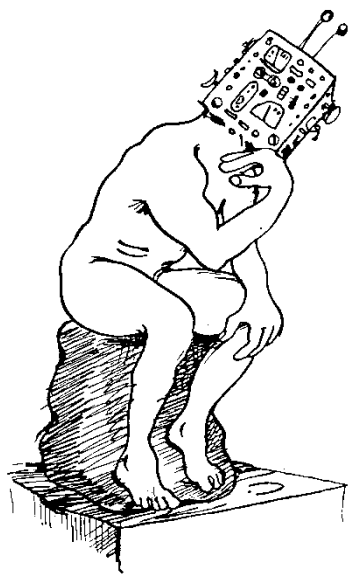


Карпов В.Э.

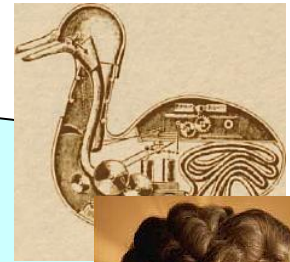
Введение в технологии искусственного интеллекта



Предыстория вопроса

- Середина XIII века. Фома Аквинский, Кельн. Альберт фон Больштед (Альберт Великий, монах-доминиканец). Человекоподобная кукла, умеющая ходить, двигать руками и говорить. Лучшее инженерное творение XIII века.
- Раввин Иегуда Лев бен Безалел. Прага. Голем. Специальный шарик с магическими формулами - шем. Жена. Суббота. Копия человека. Глина. Рост.
- Гомункулус. Еретическая идея. Религиозная догматика и мировоззрение. Повторение того, что было создано богом. Парацельс (Филипп Ауреол Теофраст Бомбаст фон Гогенгейм). 1-я половина XVI века. Врач и естествоиспытатель. Рецептúra изготовления полного подобия человека с помощью серии химических реакций, производимых над спермой человека.
- XVIII век. Андроиды. Механические люди. Технология изготовления часовых механизмов. Жак де Вокансон, француз. Механический флейтист (игрушка с часовым механизмом). Полное подобие. 11 мелодий. Пьер-Жак Дро и его сын Анри Дро. Швейцарские часовщики. 1774 г. Выставка в Париже. Писец (внешне похож на 6-ти летнего ребенка, макает перо, пишет слова и фразы, стряхивал избыток чернил; во время письма двигала головой и туловищем, имитируя человеческие движения). Рисовальщик. Музыкантша (фисгармония).
- XVIII век. Фаркаш Кемпелен. Венгр. Механик. Искусственный шахматист. Внешне - кукла размером со взрослого мужчину. В костюме турка. Сидит. Но: шкафчик-подставка. Там прятался настоящий шахматист. Проблема фиксации положения текущего расположения фигур и управления движениями рук и тела «турка».
- Карел Чапек. «R.U.R.». Новое имя - робот.

Предыстория. «Разумные машины»



Мифы и легенды

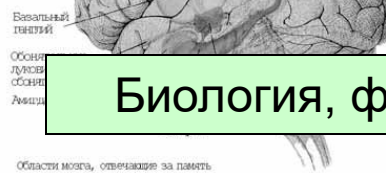
Андройды

К. Чапек (1920, «Россумские универсальные роботы»)

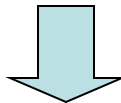
Лжероботы

А. Азимов (40-50 гг.)

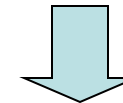
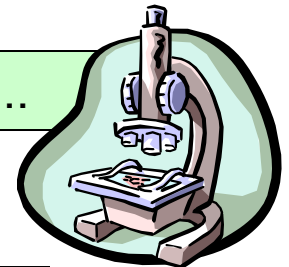
Мистификации



ЭВМ



Биология, физиология, ...



«Разумная машина»

Искусственный разум

Механизм (датчики, регуляторы, исп. механизмы)

Базовые направления ИИ

A - моделирование на ЭВМ отдельных творческих процессов

B - внешняя интеллектуализация ЭВМ

C - внутренняя интеллектуализация ЭВМ

D - **целенаправленное поведение роботов**
(создание интеллектуальных роботов, способных автономно совершать операции по достижению целей, поставленных человеком).

Определения ИИ

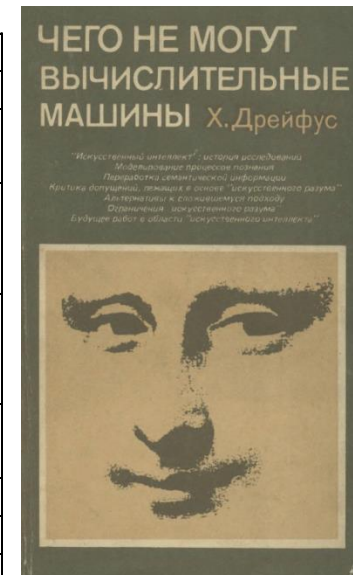
- ИИ изучает способы создания вычислительных машин, обладающих «интеллектуальным» поведением (единственное непротиворечивое определение интеллекта состоит в том, что «интеллект - это то, что оценивается в интеллектуальных тестах»).
- Интеллект - это способность правильно реагировать на новую ситуацию.
- Попытки определить понятие «интеллект» равносильны попыткам дать определение мышлению при ответе на вопрос «могут ли машины мыслить?», а сама проблема ИИ – это попытка ответить на этот вопрос.
- Пока разработка ИИ не достигнет своей высшей цели - дублирования каждого аспекта человеческого интеллекта, ему обязательно будет присуще самоотрицание при попытках определения. И поэтому ИИ - это область исследований, направленных на то, чтобы заставить машины выполнять функции, которые в настоящее время для них слишком трудны, и особенно такие функции, которые способны выполнять люди.
- Дж.Вейценбаум: ИИ выдвигает в качестве своей цели построение машин с мыслительными возможностями, соизмеримыми с мыслительными возможностями самого человечества.
- Марвин Минский анализирует саму концепцию интеллекта, рассматривая вопрос о том, сможем ли мы общаться с внеземным интеллектом, если нам когда-нибудь придется с ним встретиться.
- Изучение проблемы ИИ - это изучение природы и механизма мышления человека.
- Всякая задача, для которой неизвестен алгоритм решения, априорно относится к искусственному интеллекту.
- К сфере ИИ относятся те весьма различные области, где мы действуем, не имея абсолютно точного метода решения проблемы, и которые обладают в общем двумя характерными особенностями:
 - в них используется информация в символьной форме: буквы, слова, знаки, рисунки. Это отличает область ИИ от областей, в которых традиционно компьютерам доверяется обработка данных в числовой форме;
 - в них предполагается наличие выбора; действительно, сказать, что не существует алгоритма, это значит сказать, только то, что нужно сделать выбор между многими вариантами в условиях неопределенности, и этот недетерминизм, который носит фундаментальный характер, эта свобода действия являются существенной составляющей интеллекта.

Таблица Шиклоши

1969. Осознание первых проблем ИИ

1968: Дрейфус Х. *Чего не могут вычислительные машины*. М.: Прогресс, 1978.
333с.

Человек		Искусственный интеллект	
Возраст	Задача	Год	Программа
1	Восприятие окружающей среды	1968	Первая программа зрительного восприятия
2	Двигательные и манипуляторные способности	1968 1969	Проект «глаз-рука» Робот «Шейки»
3	Язык и речь	1966 1968 1968	Анализатор речи до 200 слов Синтезатор речи Программа обучения речи
5	Начало усвоения школьной программы		
15	Геометрия на плоскости	1959	Программа, доказывающая теоремы
15	Алгебраические задачи	1964	Программа для решения алгебраических задач
16	Геометрические задачи в пространстве	1963	Программа «Аналогия»
17	Программирование для ЭВМ	1963	Эвристическая программирующая программа
18	Математический анализ	1961	Программы для эвристического интегрирования
19	Исчисление высказываний	1959	«Общий решатель задач»
20	Игра в стоклеточные шашки	1959	Шашечная программа
21	Аксиоматическая логика	1956	Программа, делающая выводы в формальной системе



Что понимать под ИИ

1. Система, обладающая некоторыми свойствами.

1.1. Антропоморфная метафоричность.

«Интеллектуальная личность:

а) Много знает (БЗ)

б) Умеет рассуждать (познавательные процедуры: дедукция, индукция, абдукция, ...)

в) Сообразителен (эвристики)»

⇒ ИИ – это система, **имитирующая** таковые свойства.

1.2. Общая метафоричность:

ИИ - синоним искусственного разума, некое воплощение всего того, что присуще человеку.

2. ИИ - усилитель естественного человеческого интеллекта (его творческих способностей).

3. ИИ, как область исследований.

(“ИИ - это то, чем занимаются специалисты в области ИИ”)

ИИ, как область исследований

ИИ – область информатики, занимающаяся решением двух классов задач:

- 1) Слабоформализуемые задачи.
- 2) Задачи, для которых не известен эффективный алгоритм решения.

Парадигмы ИИ

- 50-60 Модели поведения
Перцептроны
- 60-е Эвристический поиск
- 70-е Представление знаний
- 80-е Обучающиеся машины
- 90-е ГА, ГП, ИЖ
- 2000-е Формальные системы,
ДСМ
«Мягкие» вычисления
- 2010-е Гибридные системы



КИИ-2014

1. Моделирование рассуждений и неклассические логики
2. Машинное обучение в интеллектуальных системах и интеллектуальный анализ данных
3. Компьютерная лингвистика
4. Когнитивное моделирование
5. Планирование и моделирование поведения
- 6. Искусственный интеллект в робототехнике**
7. Искусственный интеллект в социальной сфере и гуманитарных исследованиях
8. Интеллектуальные обучающие системы и среды
9. Нечеткие модели и «мягкие» вычисления
 - 9.1. Нейроинтеллект
- 10. Эволюционное моделирование и генетические алгоритмы**
11. Моделирование образного мышления и когнитивная графика
12. Инженерия знаний, онтологии и управление знаниями
13. Инструментальные системы для искусственного интеллекта
14. Многоагентные системы и распределенный искусственный интеллект;
15. Интеллектуальные интернет-технологии, семантический web
16. Прикладные интеллектуальные системы, динамические интеллектуальные системы и системы реального времени
17. Интеллектуальные системы поддержки принятия решений и управления
18. Интеллектуальные организации и виртуальные сообщества
19. Методологические и философские проблемы искусственного интеллекта.

КИИ-2016

Секции:

1. Инженерия знаний и онтологии
- 2. Интеллектуальные динамические и робототехнические системы**
3. Интеллектуальные системы поддержки принятия решений и управления
4. Интеллектуальный анализ данных
5. Интеллектуальный анализ текстов и семантический WEB
6. Классификация, распознавание и диагностика
7. Когнитивные исследования
8. Многоагентные и распределенные системы
9. Моделирование рассуждений и неклассические логики
10. Нечеткие модели и мягкие вычисления
11. Прикладные интеллектуальные системы
12. Программные средства ИИ

Семинар «Поведение интеллектуальных систем»

Круглый стол «Когнитивная семиотика и искусственный интеллект»

КИИ-2018

Направления:

- автоматизация рассуждений и интеллектуальный анализ данных;
- мультиагентные системы, интеллектуальные **роботы и поведение**;
- анализ естественного языка и понимание текстов;
- нечёткие модели и мягкие вычисления;
- когнитивное компьютерное моделирование;
- интеллектуальные системы и приложения методов ИИ.

Секции:

1. Инженерия знаний и онтологии
- 2. Интеллектуальные динамические и робототехнические системы**
3. Интеллектуальные системы поддержки принятия решений и управления
4. Интеллектуальный анализ данных
5. Интеллектуальный анализ текстов и семантический WEB
6. Классификация, распознавание и диагностика
7. Когнитивные исследования и психологические аспекты ИИ
8. Моделирование рассуждений и неклассические логики
9. Нечеткие модели и мягкие вычисления
10. Нейросетевые методы и нейроинформатика
11. Прикладные интеллектуальные системы

Круглые столы:

«Искусственный интеллект в промышленности и финансах»

«Проблемы подготовки квалифицированных пользователей и разработчиков интеллектуальных систем».

КИИ-2019

Секции:

1. Прикладные интеллектуальные системы
2. Интеллектуальные динамические и робототехнические системы
3. Многоагентные и сетевые модели
4. Интеллектуальный анализ текстов естественного языка
5. Принятие решений
6. Когнитивные исследования
7. Машинное обучение
8. Инженерия знаний и онтологии
9. Нейросетевые технологии
10. Интеллектуальный анализ данных
11. Моделирование рассуждений

Пленарные доклады:

- Н.Г.Ярушкина. Возможности нечетких систем в задачах искусственного интеллекта
- В.Э.Карпов. Социальные сообщества роботов
- Г.С.Осипов. Активность в области искусственного интеллекта в России
-

Круглые столы:

- Современные тенденции в теории мягких вычислений
- Искусственный интеллект в цифровой экономике

ИИ, И-разум, НС

1. ИИ \neq НС

ИИ = {Логика, Нечеткие вычисления, НС, ГА,...}, **НС \subseteq ИИ**

2. ИИ \neq И-разум

Электронная библиотека ИФ РАН » Новая философская энциклопедия:

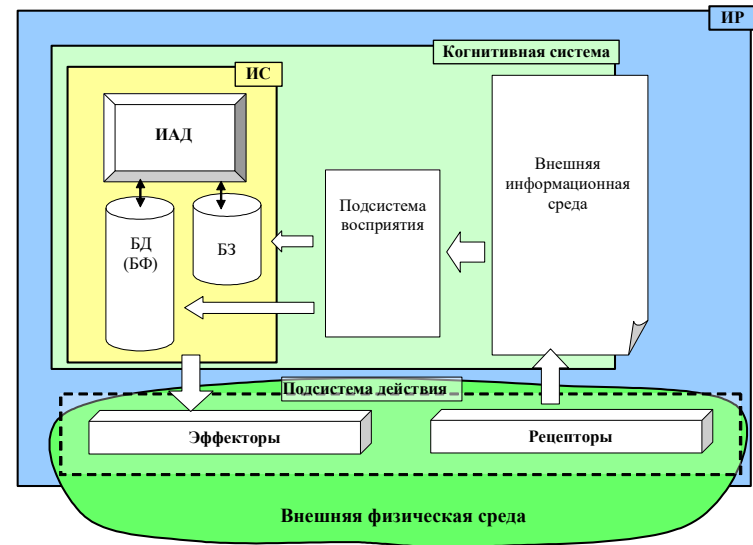
- И – способность к опосредованному, абстрактному познанию, включающая в себя такие функции, как сравнение, абстрагирование, образование понятий, суждение, умозаключение.
- И – это высшая познавательная способность, дающая принципы рассудку (Кант).
- РАЗУМ - полная нерасчлененность субъекта и объекта (Шеллинг). Основная человеческая способность – это **интеллектуальная** интуиция как непосредственное созерцание предмета в единстве противоположностей.

<https://iphlib.ru/greenstone3/library/collection/newphilenc/document/HASH523b26975f21f2447bc452>

Определения

Что такое ИС?

- Технологии ИИ
- Интеллектуальные системы



С определениями [И-]робота еще сложнее

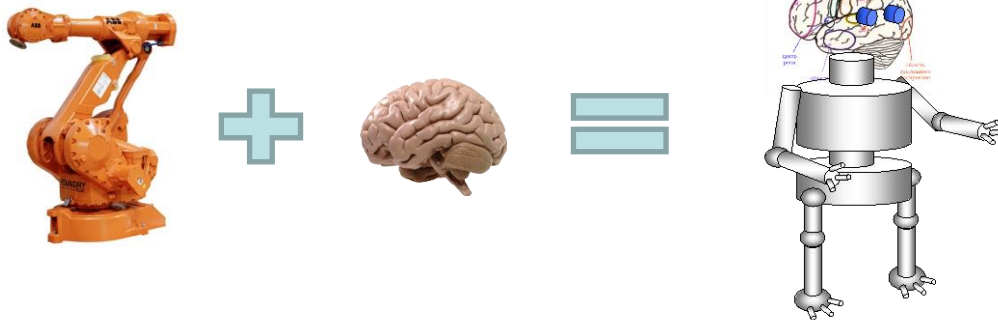
$ИР = ИС + \text{подсистема_восприятия} + \text{подсистема_действия}$

Кстати, о роботах

ISO 8373:2012 Robots and robotic devices:

- «2.6. robot: actuated mechanism programmable in two or more axes (4.3) with a degree of autonomy (2.2), moving within its environment, to perform intended tasks.
- 2.28. intelligent robot: robot (2.6) capable of performing tasks by sensing its environment and/or interacting with external sources and adapting its behavior».

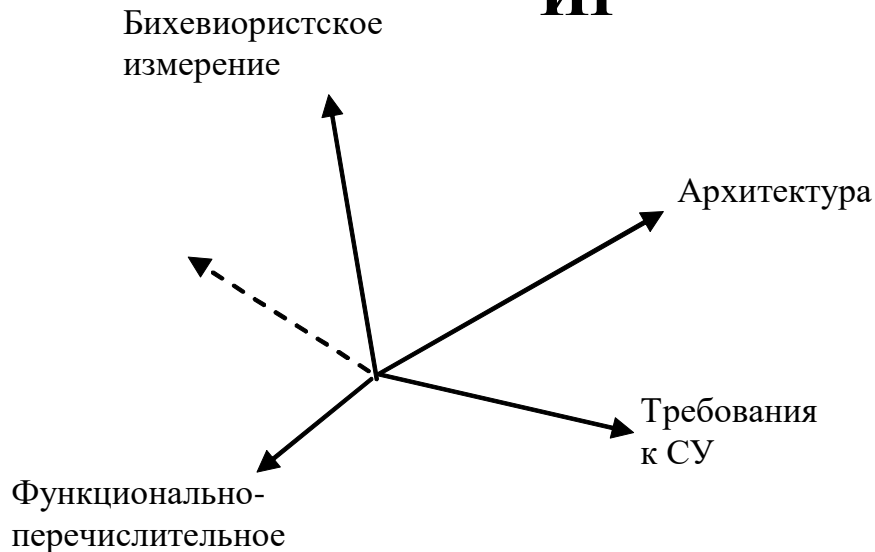
Робот – это машина. (1) Автономная, (2) антропоморфная, (3) универсальная.



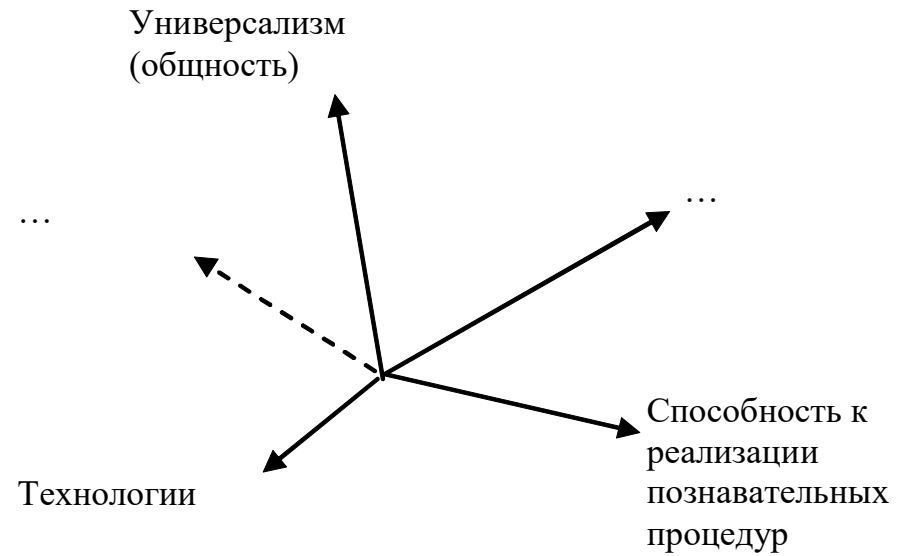
Многомерное определение

- **Замечание.** Основная цель любого определения, как и любой классификации, - дать возможность различать между собой объекты, соотнося их к тому или иному классу.
- Пример ИС – интеллектуальный робот
- ИС / ИР следует рассматривать как некий **многомерный** объект, с разных точек зрения, с позиций разных координатных осей.

ИР



ИС



ИИ повсюду

Путин: «Лидер по созданию искусственного интеллекта станет властелином мира», сент. 2017

Армия

Образование

Гуманитарии

- Юристы (авторское право, законотворчество для ИИ-систем)
- юридической деятельности)
- Художники (SciArt)
- Социальные психологи

Инженерное сообщество

- Стандартизация IEEE
- LegalTech (информационно-технологическое обслуживание



Робот Борис, профессиональный форуме "Проектория", Ярославль, 2018

ИИ в РФ

Указ Президента РФ от 10 октября 2019 г. № 490 "О развитии искусственного интеллекта в Российской Федерации", 10 октября 2019

30. Фундаментальные научные исследования должны быть направлены на создание принципиально новых научных результатов, в том числе на создание универсального (сильного) искусственного интеллекта, и решение иных задач, предусмотренных настоящей Стратегией, включая реализацию следующих приоритетов:

- а) алгоритмическая имитация биологических систем принятия решений, в том числе распределенных коллективных систем, таких как пчелиный рой или муравейник;
- б) автономное самообучение и развитие адаптивности алгоритмов к новым задачам;
- в) автономная декомпозиция сложных задач, поиск и синтез решений.

<https://www.garant.ru/products/ipo/prime/doc/72738946/>

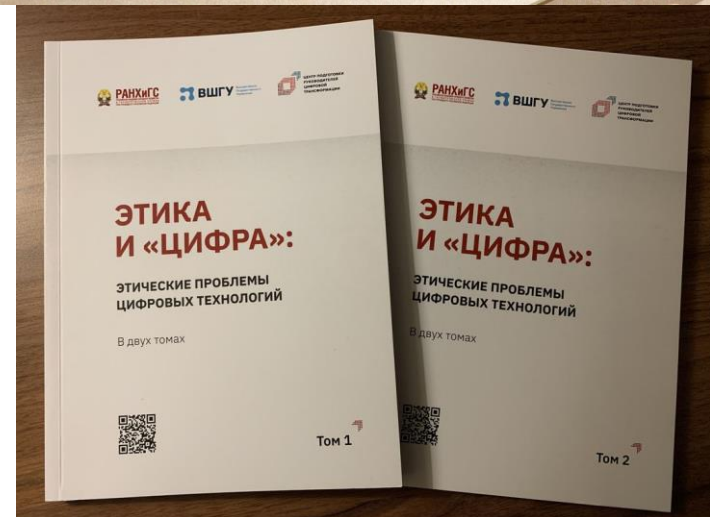
Федеральный проект «ИИ». Разработан в целях реализации Национальной стратегии развития ИИ на период до 2020 г., утвержденной указом президента РФ от 10 окт. 2019 г. №490

Гайдаровский форум - 2020

15 января 2020, XI
Гайдаровский форум.

Тематическая дискуссия
«Этические проблемы
применения цифровых
технологий: как искать
решение?»

Доклад «Этика и "цифра":
этические проблемы цифровых
технологий»



Из материалов докладов для юристов

- Основной источник - СМИ

СМИ. «Робот-юрист с искусственным интеллектом»

- Американская юридическая фирма Baker&Hostetler объявила о приеме на работу созданного IBM робота AI Ross, который вместе с другими 50-ю сотрудниками будет заниматься делами, связанными с банкротствами.
- AI Ross первый в мире робот-адвокат с ИИ на базе известного **когнитивного компьютера Watson** – был разработан с целью понимать естественный язык, выдвигать гипотезы, если задан вопрос, проводить исследование и давать развернутый ответ (с ссылками и цитатами). ...



<https://neuronus.com/news-tech/1076-robot-yurist-s-iskusstvennym-intellektom-nezavisimoe-i-ob-ektivnoe-pravosudie-sovsem-ryadom.html>

СМИ. «В МВД задумались о замене юристов искусственным интеллектом»

МВД России ведет активную работу по разработке специального программного обеспечения, которое в автоматическом режиме будет искать пробелы и противоречия в законах, фактически выполняя функции юристов. Открытый конкурс на выполнение соответствующей научно-исследовательской работы размещен на портале госзакупок.

<https://lenta.ru/news/2018/07/04/mvd/>

На самом деле. С сайта госзакупок

Наименование НИР – «Формирование требований на создание модуля **управления знаниями ...** и использования результатов **семантического анализа нормативных правовых актов**».

- поиск пробелов и противоречий в НПА;
- определение способов классификации НПА;
- поиск аналогов НПА в БЗ; ...
- поиск аналогов коррупциогенных факторов в БЗ;
- проверка соответствия терминов, используемых в проекте НПА, терминам, используемым в действующем законодательстве.

Победитель - ООО «Проектный сервис» (ООО «ПРО»)

<http://zakupki.gov.ru/epz/order/notice/ok44/view/common-info.html?regNumber=0373100088718000021>

СМИ. ИИ ПОБЕДИЛ СОТНЮ ЛУЧШИХ ЛОНДОНСКИХ ЮРИСТОВ

- 2017 г. Соревнования между ИИ – программой Case Crunch Alpha и командой, состоящей из 100 юристов ведущих фирм Лондона. Предсказание решения омбудсмена по финансам по 100 делам по страхованию платежей. В результате было сделано 775 предположений. Победил компьютер, прогнозы которого сбылись на **86,6%**, результат юристов – на 66,3%.
- Программа была создана стартапом CaseCrunch студентов-юристов из Кембриджа, которые начали с разработки чат-бота, отвечающего на простые юридические вопросы.

<http://legalinsight.ru/ii-pobedil-sotnyu-luchshih-londonskih-yuristov/>



СМИ. Искусственный интеллект лучше юристов анализирует документы

- ИИ превзошел юристов по скорости и эффективности оценки документов.
- В ходе эксперимента сотрудники Goldman Sachs, Cisco, а также других компаний должны были оценить риски в пяти разных соглашениях, обнаружив в них слабые места.
- Люди смогли выполнить это задание с точностью в 94%, а нейросети — в 100%.
- Представители стартапа LawGeex, которые **разработали ИИ**, отметили, что ИИ никогда раньше не сталкивался с анализируемыми в ходе тестирования документами.



<https://iz.ru/714327/2018-02-28/iskusstvennyi-intellekt-luchshe-iuristov-analiziruet-dokumenty>

СМИ. ИИ работает вместо юристов

- Сбербанк запустил новый сервис, позволяющий в несколько кликов создавать юридические документы на базе встроенной **интеллектуальной системы выбора условий документа и многовариативных шаблонов**. Похожую услугу предложили для российских клиентов и специалисты французской компании Wonder.Legal. Они применили возможности ИИ для создания юридических документов.
- Сервис "Конструктор документов" доступен для подключения в интернет-банке "Сбербанк Бизнес Онлайн". Интеграция в систему электронного документооборота E-invoicing. "Это позволяет клиентам подписывать создаваемые документы электронной подписью и отправлять по системе E-invoicing своему партнеру для подписания в электронном виде", - пояснили в пресс-службе банка.



<https://www.comnews.ru/content/111669/2018-02-05/iskusstvennyy-intellekt-rabotaet-vmesto-yuristov>

Этика и стандартизация

Этические вопросы в науке

- Опасности, «вызовы» современного высокотехнологического мира
 - Этика нанотехнологий
 - Этические проблемы генных технологий
 - Этика больших данных ...
- Декларации этических принципов научной деятельности:
 - оценка социальных и гуманитарных рисков научно-технических разработок и тех благ, которые они могут принести человеку и обществу.
- Этика робототехники (UNESCO, 2017. REPORT OF COMEST ON ROBOTICS ETHICS)
- Этические проблемы ИИ

Этические проблемы в ИИ

Опасность «думающих машин»

- Turing, A., 1950. Computing machinery and Intellicence. Машинные ошибки
- Wiener, N., 1960. Some Moral and Technical Consequences of Automation. Опасность для человека и этические проблемы
- Bostrom, N., Yudkowsky, E., 2011. The Ethics of Artificial Intelligence. Негативные сценарии возможного развития ИИ
- **IEEE, 2016. Ethically Aligned Design.** Основные ближнесрочные угрозы, связанные с внедрением автономных систем на базе ИИ
- UNESCO, 2017. REPORT OF COMEST ON ROBOTICS ETHICS

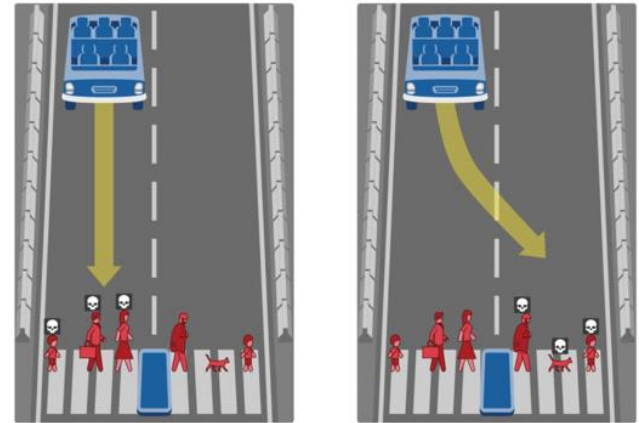
О чем идет речь

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.

Moral Machine: 40 млн. решений (примеров для обучения)

Новая постановка задачи

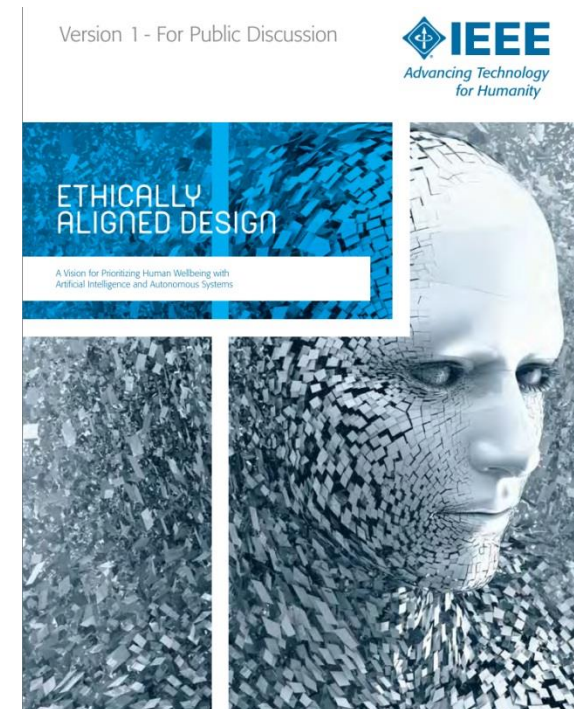
Стандартизация =>

Ответ на вопрос: *насколько система ИИ соответствует 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 ИИС решения **соответствуют этическим нормам**, т.е. насколько она «этична».

Перечень проблем

Предмет исследования этически обусловленного проектирования - ИИС, совершающие выбор того или иного **значимого действия** или решения.

При этом совершение выбора осуществляется на основе некоторых **эвристик**, основанных на этических императивах.

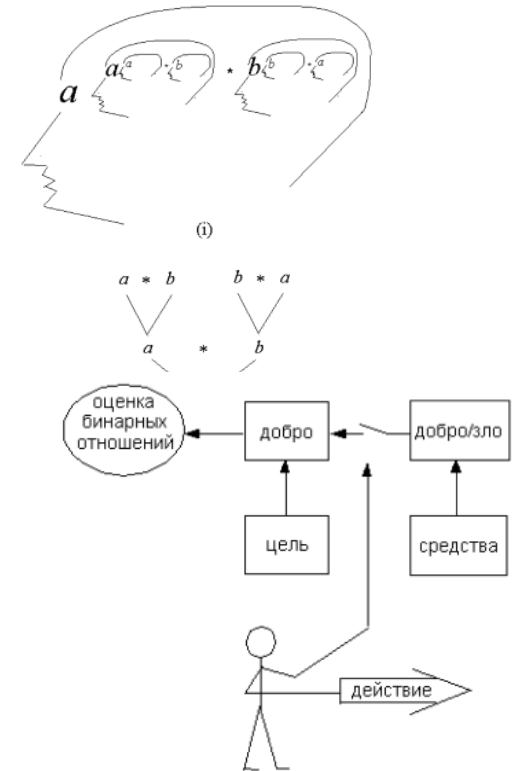
1. Формализация. Вопросы конструктивных определений и онтологий.
2. Математический аппарат. Модели и методы.
3. Этическая верификации. Определение степени соответствия ИИС этическим нормам.

Формализация этических понятий

Формальная этика: нахождение формальных критериев «правильности» поведения (алгебра поступков).

- Gensler, H.J. 1996. Formal Ethics. London, New York.
- Лефевр В. **Алгебра совести** / Пер.с англ. М.: Когито-Центр, 2003. 426 с

-
- Фоминых И.Б. Эмоции как аппарат оценок поведения интеллектуальных систем. // КИИ-2006. -М.: Физматлит, 2006, Т.2.
 - Фоминых И.Б. Классификация эмоций: информационный подход //ИМиМВ в ИИ. -М.: Физматлит, 2007, Т.2.



Математический аппарат. Модели и методы

- Неклассические логики
(В.К. Финн, О.П. Кузнецов, В.Б. Тарасов,
В.Н. Вагин, А.С. Карпенко, ...)
- Вероятностные модели и нечеткая логика
- Методы вербального анализа решений
(О.И. Ларичев)
- ...

Базовый набор инструментальных средств, позволяющих решить задачу формализации этических норм, на настоящий момент имеется.

Верификация

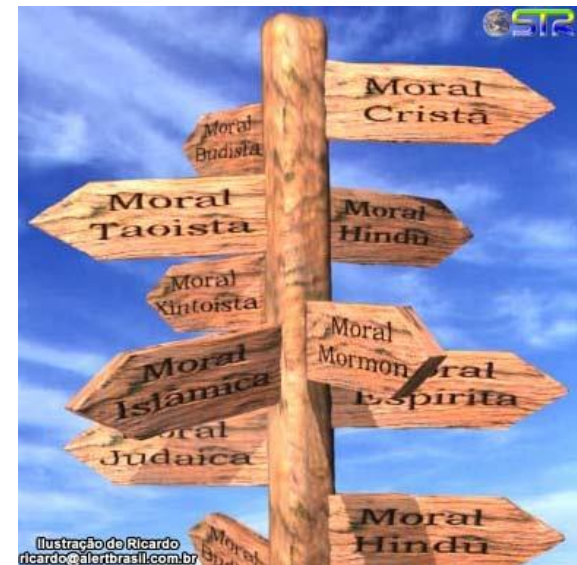
ИИС осуществляют принятие решений, основанные на этических императивах и нормах = эвристики.

Верификация этичности ИИС

- Проблема объяснительной компоненты
- Противоречивость и многозначность этических норм
- Неэтичный вывод из последовательности этичных рассуждений
- Относительность весов правил
- Проблема контекста
- Черные ящики

Вывод:

- Тесты
- Возврат к тезису Тьюринга



Выводы

1. Предмет исследования этически обусловленного проектирования - ИИС, совершающие **выбор** того или иного значимого действия или решения. Совершение выбора осуществляется на основе некоторых **эвристик**, основанных на этических императивах.
2. Принципиально важным являются вопросы конструктивных определений и **онтологий**. Эти определения должны иметь многомерный характер и позволять рассматривать сущности с различных сторон, систем оценок и проч. Задача онтологий заключается во взаимном увязывании и согласовании этических и технических понятийных систем.
3. В настоящее время **существует математический аппарат**, способный в той или иной мере реализовать формализм, необходимый для этически обусловленного проектирования.
4. С технической точки зрения основной проблемой является проблема **этической верификации**. Эта верификация заключается в комплексе **тестов**, способных определить "степень этичности" интеллектуальной системы. Иного способа определения этой степени, кроме наблюдений за реакциями и поведением исследуемой ИИС, не существует.

Оргструктуры

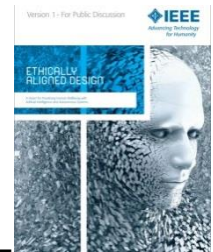
Российская рабочая группа IEEE <http://ecai.raai.org/>

Цели:

- представлять мнения и предложения российских ученых, участвующих в работах над документами IEEE: Ethically Aligned Design;
- информирование российских ученых о результатах деятельности рабочей группы IEEE по созданию документа IEEE: Ethically Aligned Design;
- привлечение российских ученых к исследованиям по тематике «Этика систем ИИ»

Состав рабочей группы:

- Готовцев П.М., руководитель, координатор
- Карпов В.Э., Ройзензон Г.В., Овсянникова Е.Е., Огородов Д.В., Нестик Т.А., Платонова М.В.



Научный совет РАН по методологии искусственного интеллекта и когнитивных исследований.

Секция «Этические проблемы искусственного интеллекта» на базе Института философии РАН, философского факультета МГУ и РАИИ