

## Введение в ИИ

ИИ можно рассматривать с точки зрения позиций:

- рациональность и подражание человеку
- мышление и поведение

В связи с чем можно выделить два подхода:

- человекоподобный интеллект
- рациональный интеллект

Тест Тьюринга 1950 — «Может ли машина мыслить?»

- обработка естественного языка (natural language processing – NLP) для успешного общения на человеческом языке
- представление знаний (knowledge representation) для хранения того, что он знает или слышит
- автоматизированное формирование рассуждений (automated reasoning) для ответа на вопросы и получения новых выводов
- машинное обучение для адаптации к новым обстоятельствам, обнаружения и экстраполяции закономерностей

Полный тест Тьюринга:

- компьютерное зрение и распознавание речи для восприятия мира
- робототехника для манипулирования объектами и передвижения

Все эти шесть дисциплин формируют большую часть систем ИИ. Но исследователи все реже обращаются к тесту Тьюринга.

Мы можем узнать о человеческом мышлении тремя способами:

- самоанализ — попытка уловить собственные мысли в процессе размышлений
- психологические эксперименты — наблюдение за человеком в действии
- визуализация мозга — наблюдение за мозгом в действии

Изучением искусственного интеллекта в купе с точными и повторяемыми моделями человеческого разума занимаются когнитивные науки.

Рациональное мышление:

- силлогизмы (Люди — смертны, Сократ — человек, Сократ — смертен)
- логика — точная нотация для высказываний об объектах мира и отношениях между ними
- теория вероятности — позволяет проводить строгие рассуждения с неопределенной информацией

Агент (от греч. agere — действовать) отличается от обычной программы:

- автономное функционирование
- восприятие окружающей среды
- длительное существование
- адаптация к изменениям
- создание и достижение целей

Рациональный агент — это агент, который действует так, чтобы достичь наилучшего результата или, при наличии неопределенности, наилучшего ожидаемого результата.

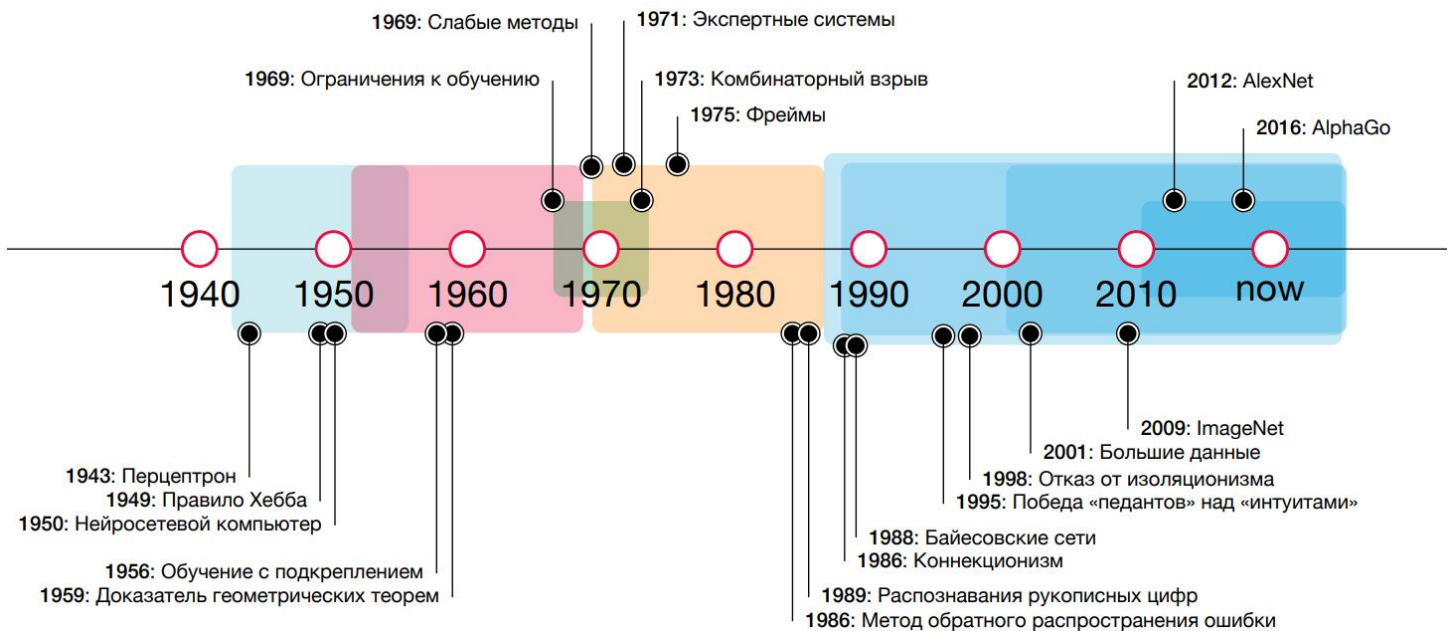
Рациональный агент хорошо работает, когда мы можем точно формализовать цель (например, шахматы). Но в реальной жизни сформулировать правильно цель может быть крайне сложной задачей.

Проблема достижения согласия между нашими истинными предпочтениями и целью, которую мы закладываем в машину, называется проблемой согласования ценностей.

В итоге нам нужны агенты, которые будут доказательно полезны для людей.

Одним из кратких способов обобщить основные вехи истории ИИ является перечисление лауреатов премии Тьюринга:

- Марвин Мински (1969) и Джон Маккарти (1971) за определение основ области ИИ, основанной на представлении знаний и рассуждениях
- Эд Фейгенбаум и Радж Редди (1994) за разработку экспертных систем, которые кодируют человеческие знания для решения реальных проблем
- Джудеа Перл (2011) за разработку методов вероятностных рассуждений, которые принципиально работают с неопределенностью
- Йошуа Бенгио, Джеффри Хинтон и Янн ЛеКун (2019) за то, что «глубокое обучение» (многослойные нейронные сети) стало важной частью современных вычислительных систем



## Зарождение ИИ

- Уорреном МакКаллохом и Уолтером Питтсом (1943) была предложена модель искусственного нейрона — перцептрон
- Дональд Хебб (1949) предложил правило изменения силы связи нейронов, что легло в основу обучения нейронных сетей
- Марвин Мински и Дин Эдмондс (1950) построили первый нейросетевой компьютер
- Алан Тьюринг (1950) в своей статье представил тест Тьюринга, машинное обучение, генетические алгоритмы и обучение с подкреплением

## Большие ожидания

- Ньюэлл и Саймон (1976) сформулировали знаменитую гипотезу о системе физических символов, которая гласит, что «система физических символов обладает необходимыми и достаточными средствами для общего разумного действия»

- Герберт Геллернтер (1959) создал программу Geometry Theorem Prover, которая позволяла доказывать теоремы, которые многие студенты-математики сочли бы довольно сложными
- Артур Самуэль (1956) написал программу играющую в шахматы на любительском уровне, первая реализация обучения с подкреплением
- Джон Маккарти (1958) представил язык Lisp и Advice Taker, гипотетическую программу, которая воплощает общие знания о мире и может использовать их для выработки планов действий

#### Возвращение к реальности

- Неспособность справиться с «комбинаторным взрывом» была одним из основных критических замечаний в адрес ИИ, содержащихся в докладе Лайтхилла (Lighthill, 1973), который лег в основу решения британского правительства прекратить поддержку исследований ИИ во всех университетах, кроме двух.
- В книге Мински и Пейперта «Перцептроны» (1969) было доказано, что перцептрон с двумя входами не может быть обучен распознавать, когда два его входа различны.

#### Экспертные системы

Механизмы поиска общего назначения, пытающийся соединить элементарные шаги рассуждений для нахождения полных решений, были названы «слабыми методами»

- Программа DENDRAL (Buchanan et al., 1969) была ранним переходом к частным узкоспециализированным системам, которая была создана для решения проблемы вывода молекулярной структуры из информации, предоставляемой масс-спектрометром
- Система MYCIN для диагностики инфекций крови, имела около 450 правил, которые были получены путем опроса большого количества экспертов. Для расчета неопределенности в диагнозе был введен «фактор неуверенности», который, как казалось (в то время), хорошо согласуется с тем, как врачи оценивают влияние доказательств на диагноз.

#### Возвращение к нейронным сетям

- В середине 1980-х годов по крайней мере четыре различные группы заново изобрели алгоритм обучения методом обратного распространения ошибки (back-propagation), впервые разработанный в начале 1960-х годов.
- Коннекционизм предполагает, что мыслительные явления могут быть описаны сетями из взаимосвязанных простых элементов

#### Перемены

- Хрупкость экспертных систем привела к новому, более научному подходу, включающему вероятность, а не булеву логику, машинное обучение, а не ручное кодирование, и экспериментальные результаты, а не философские утверждения.
- Отказ от изоляционизма
- Джудеи Перла (1988) разработал байесовские сети и дал строгий и эффективный формализм для представления неопределенных знаний, а также практические алгоритмы для вероятностных рассуждений
- Рич Саттон (1988) связал обучение с подкреплением с теорией марковских процессов принятия решений, разработанной в области исследования операций

#### Большие данные

- Значительный рост вычислительных мощностей и создание Всемирной паутины способствовали созданию очень больших массивов данных - явление, иногда называемое большими данными

- Банко и Брилл (2001) утверждают, что улучшение производительности, полученное при увеличении размера набора данных на два или три порядка, перевешивает любое улучшение, которое может быть получено при настройке алгоритма
- Доступность десятков миллионов изображений в базе данных ImageNet (Deng et al., 2009) вызвала революцию в области компьютерного зрения.
- Победа системы Watson компании IBM над чемпионами в игре-викторине "Jeopardy!" в 2011 году, что оказало большое влияние на восприятие ИИ обществом

### Глубокое обучение

Термин «глубокое обучение» относится к машинному обучению с использованием нескольких слоев простых, настраиваемых вычислительных элементов.

- Использование глубокой сети для представления функции оценки способствовало победам ALPHAGO над ведущими игроками в го (Silver et al., 2016, 2017, 2018).

### Риски и преимущества

- автоматизированное оружие,
- тотальная слежка,
- необъективное принятие решений,
- рынок труда,
- критическая важность безопасности в некоторых приложениях,
- кибер-безопасность,
- проблема гориллы,
- проблема царя Мидаса.