

Машинное обучение и интеллектуальные системы

Лекция 1. Определение ИИ. Направления исследований ИИ. История ИИ и современное состояние.

Структура курса

Лекции	ИДБ-16-10	ИДБ-16-09
04.09		
11.09		
	15.09	13.09
18.09		
	22.09	20.09
25.09		
	29.09	27.09
02.10		
	06.10	04.10
09.10		
		11.10
16.10		
		18.10
23.10		
		25.10
30.10		
	02.11	01.11
06.11		
	09.11	09.11
13.11		
	16.11	16.11
20.11		
	23.11	23.11

Структура курса

- Решение задач (Problem solving): поиск, эвристики, игры.
- Задачи оптимизации, генетические алгоритмы
- Логика, резолюция, логический вывод
- Представление знаний, планирование, экспертные системы
- Неопределенность, рассуждения с использованием вероятностей
- Машинное обучение. Введение.
- Машинное обучение и вероятностные модели.
- Глубокое обучение (Deep Learning).
- Различные применения и приложения. Компьютерное зрение, обработка естественных языков и др.

Структура курса

- Решение задач (Problem solving): поиск, эвристики, игры.
- Задачи оптимизации, генетические алгоритмы
- Логика, резолюция, логический вывод
- Представление знаний, планирование, экспертные системы
- Неопределенность, рассуждения с использованием вероятностей
- Машинное обучение. Введение.
- Машинное обучение и вероятностные модели.
- Глубокое обучение (Deep Learning).
- Различные применения и приложения. Компьютерное зрение, обработка естественных языков и др.

Определение ИИ

Определение ИИ



Алан Тьюринг (1912-1954)

- **Тест Тьюринга:** *«Человек взаимодействует с одним компьютером и одним человеком. На основании ответов на вопросы он должен определить, с кем он разговаривает: с человеком или компьютерной программой. Задача компьютерной программы — ввести человека в заблуждение, заставив сделать неверный выбор»* [Computing Machinery and Intelligence, 1950]
- Что требует:
 - Обработка естественных языков
 - Компьютерное представление знаний
 - Автоматические рассуждения
 - Машинное обучение
 - Дополнительно: компьютерное зрение
 - Дополнительно: робототехника

Определение ИИ

Человеческое мышление

"Новые усилия по созданию разумных машин"

"Автоматизация видов деятельности, которые мы связываем с человеческим мышлением, включая принятие решений, решение задач, обучение..."

Человеческое поведение

"Искусство создания машин, который выполняют функции, требующие разум при выполнении людьми"

"Изучение способов выполнения компьютерами задач, для которых в настоящий момент люди приспособлены лучше"

Рациональное мышление

"Изучение мыслительных способностей с помощью компьютерные моделей"

"Исследование вычислений, которые позволяют воспринимать мир, размышлять и действовать"

Рациональное поведение

"Компьютерный интеллект - это исследования по проектированию разумных агентов."

"Область ИИ ... занимается интеллектуальным поведением искусственных объектов"

Определение ИИ

- "Программы, демонстрирующие свойственные людям черты"
- "Построение целеустремлённых агентов"
- "Решение актуальных и предстоящих сложных вызовов"
- "Способ построения больших и сложных систем"
- "ИИ может делать то, что компьютеры сейчас не могут"

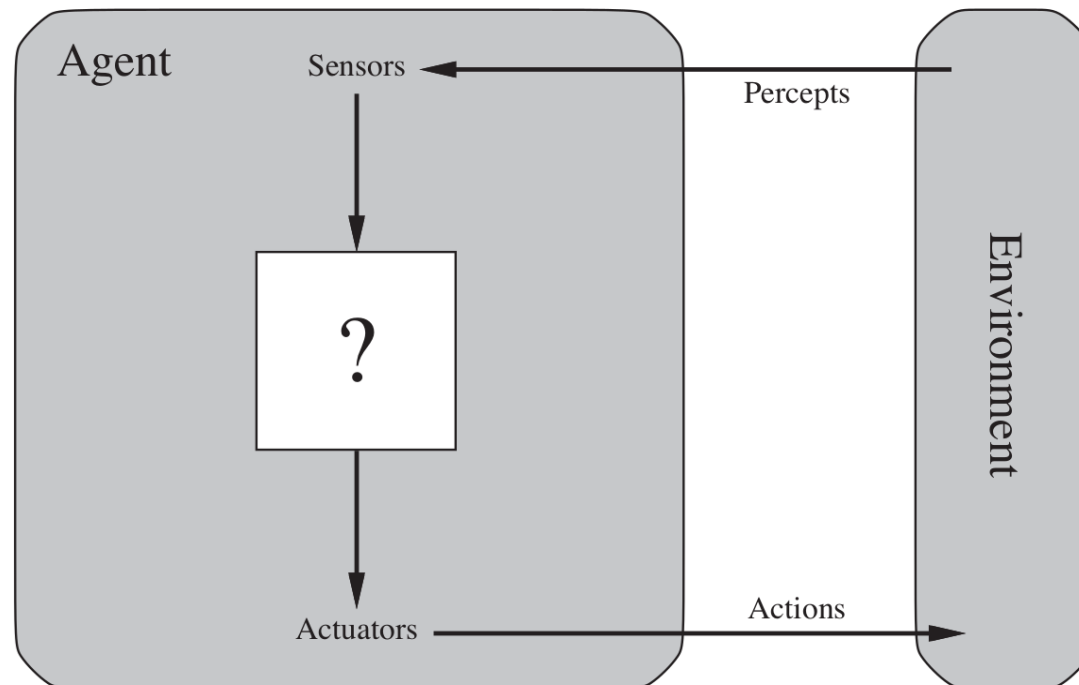
Определение ИИ

- ИИ - это обозначение, применяющееся к некоторым направлениям исследований и разработок.



Рациональные агенты

- Агент - любой объект способный *воспринимать* свою среду и *действовать* над ней.
- Для каждой возможной последовательности наблюдений *рациональный агент* выбирает действие, от которого ожидается максимизация его критерия производительности, опираясь на свидетельства, полученные через сенсоры и встроенные в него знания.
- Рациональный - не значит идеальный.



Проблемная среда

Агент	Критерии производительности	Среда	Исполняющие устройства	Сенсоры
Автопилот такси	Безопасность, скорость, соблюдение ПДД, комфортность, экономия топлива, прибыль	Дороги, города, другие автомобили, пешеходы, клиенты	Рулевое управление, тормоза, сигналы, дисплей	Камеры, сонар, дальномер, GPS, акселерометр, датчики
Поисковая система	Точность (процент страниц в выдаче, релевантных запросу), полнота (процент релевантных страниц, попавших в выдачу), прибыль, число пользователей, удобство использования, соблюдение законов и норм	Сеть Интернет, веб-страницы, гиперссылки, пользователи, общество, запросы на естественном языке	Отображение результатов в виде веб-страниц, отображений исправлений	Текстовый или голосовой ввод, Веб-краулеры.
Система распознавания лиц	Корректность распознавания (точность, полнота, процент распознанных)	Фотографии, картины, люди	Отображение выбранного человека или вероятностей.	Массивы из пикселей

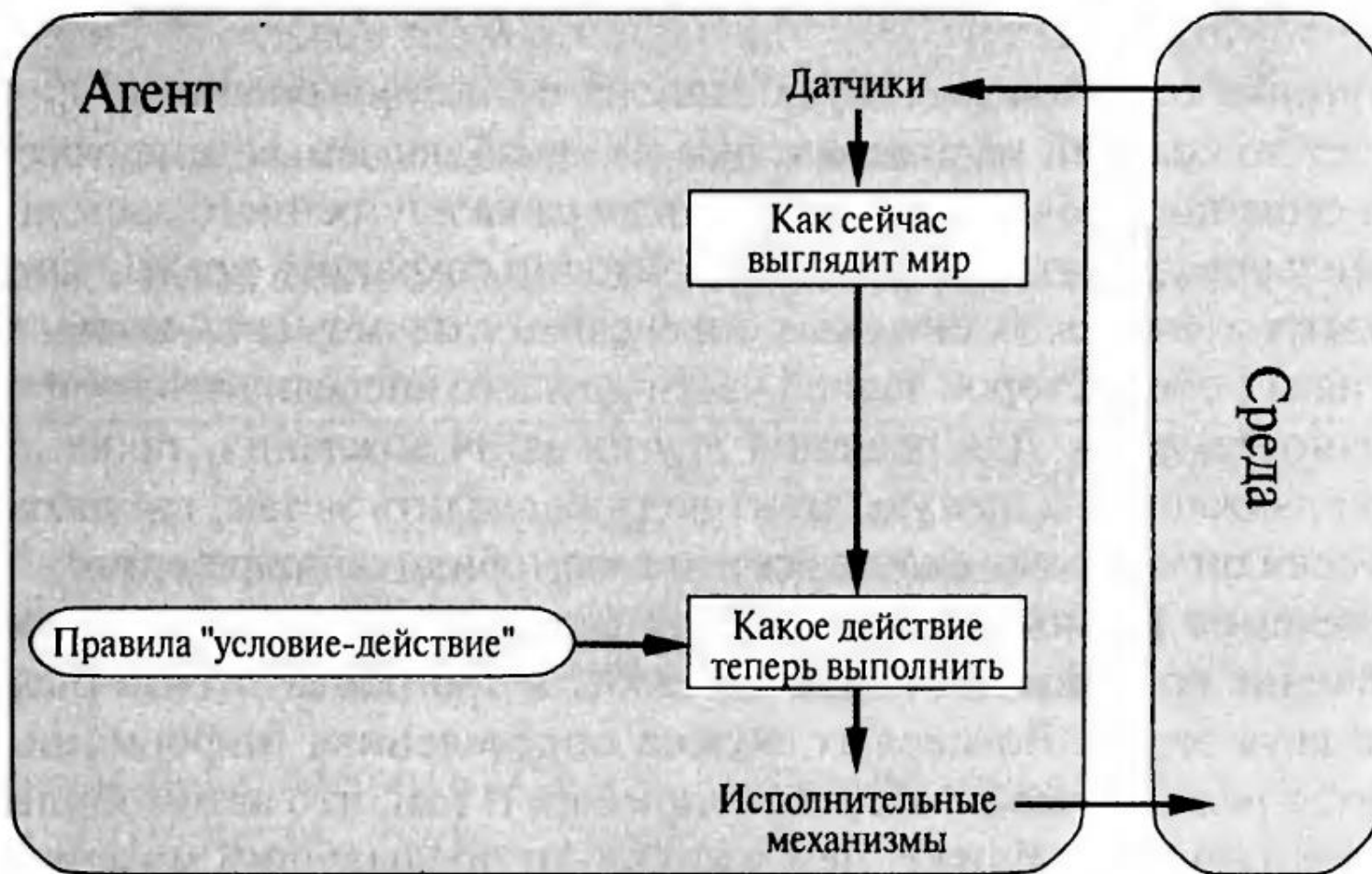
Виды сред

- Полностью или частично наблюдаемые
- Одноагентные и мультиагентные (соревновательные, кооперативные)
- Детерминированные и стохастические
- Эпизодические и последовательные (зависит ли следующий эпизод от действий агента в предыдущем).
- Статические и динамические (меняется ли среда, пока агент выбирает действие)
- Дискретные и непрерывные
- Известные и неизвестные (знает ли агент последствия своих действий или хотя бы их вероятности).

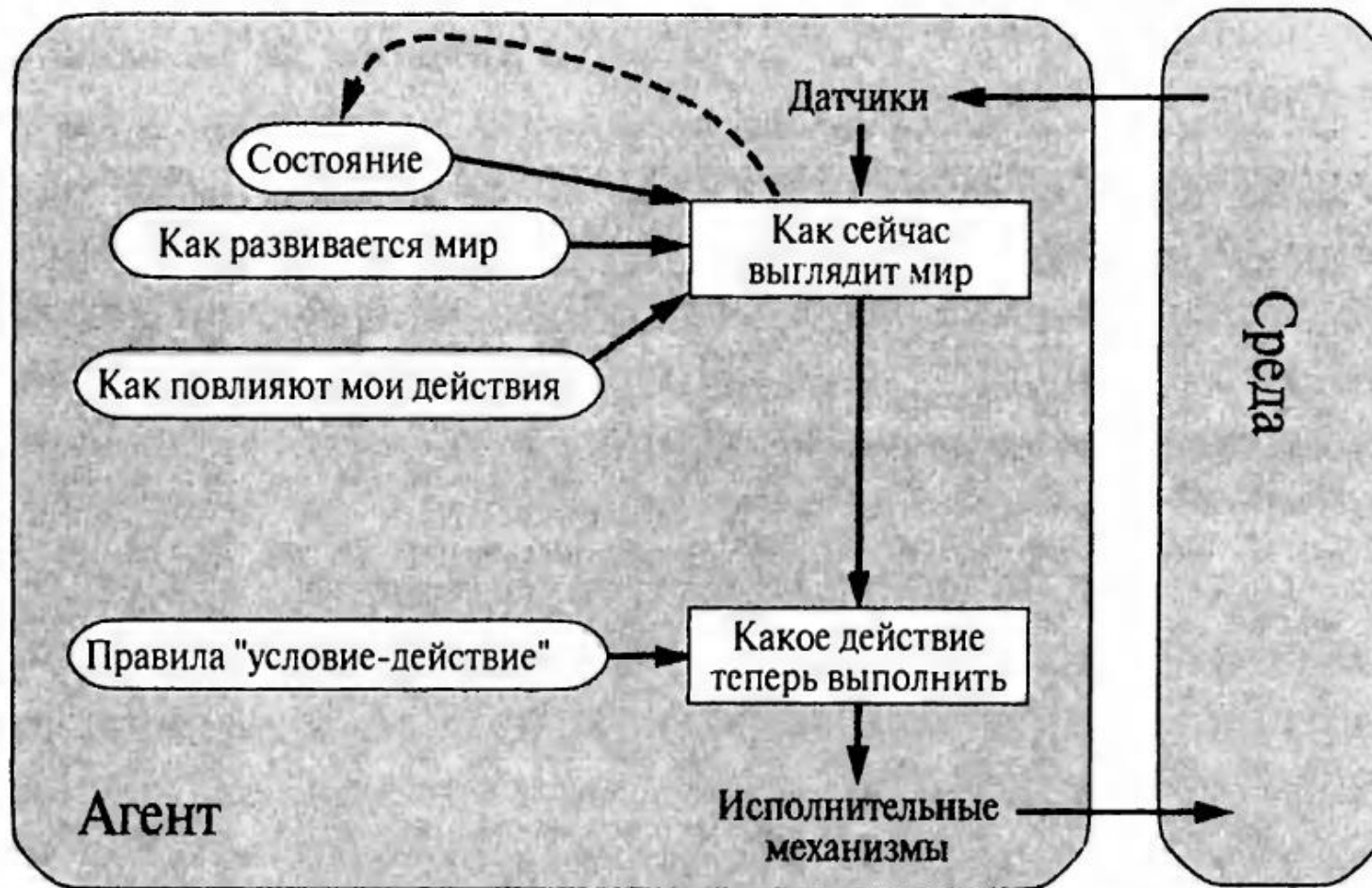
Структура агентов

- Агент = архитектура + программа
- Архитектура - вычисляющее устройство с сенсорами и исполняющими устройствами.
- Программа - выполняет функцию агента, т.е. отображает последовательности наблюдений на действия

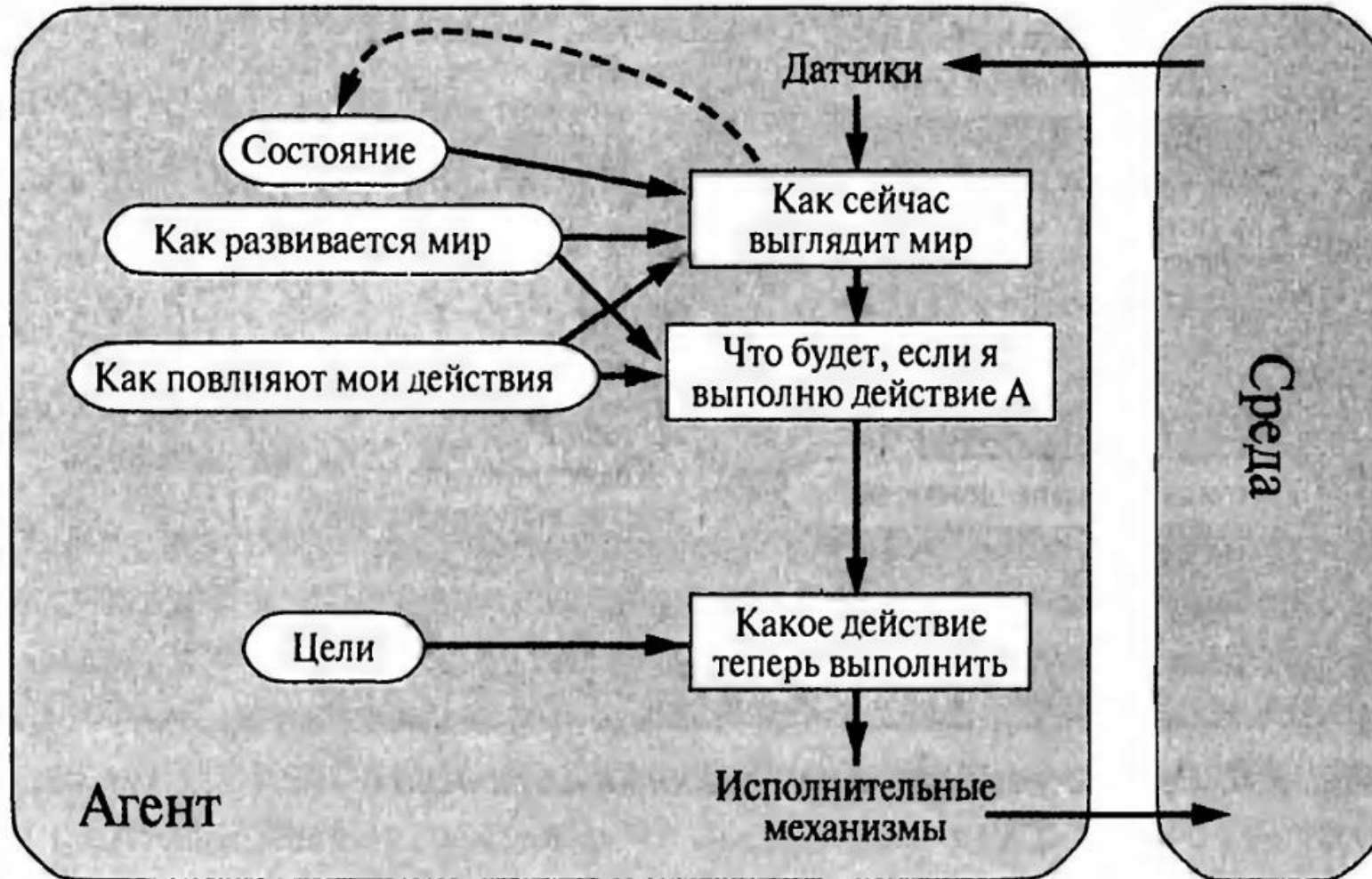
Простой рефлексный агент



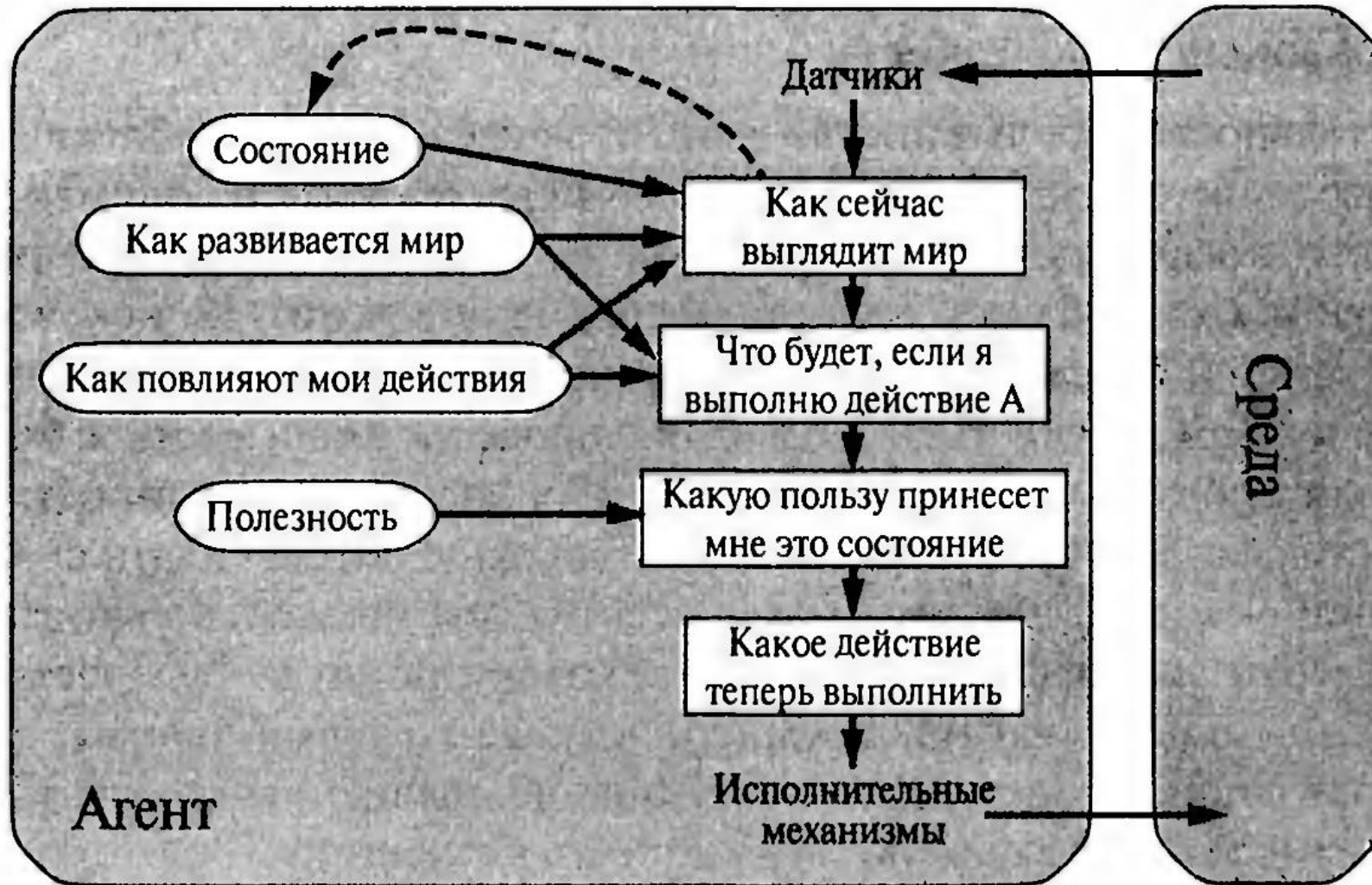
Рефлексный агент с моделью мира



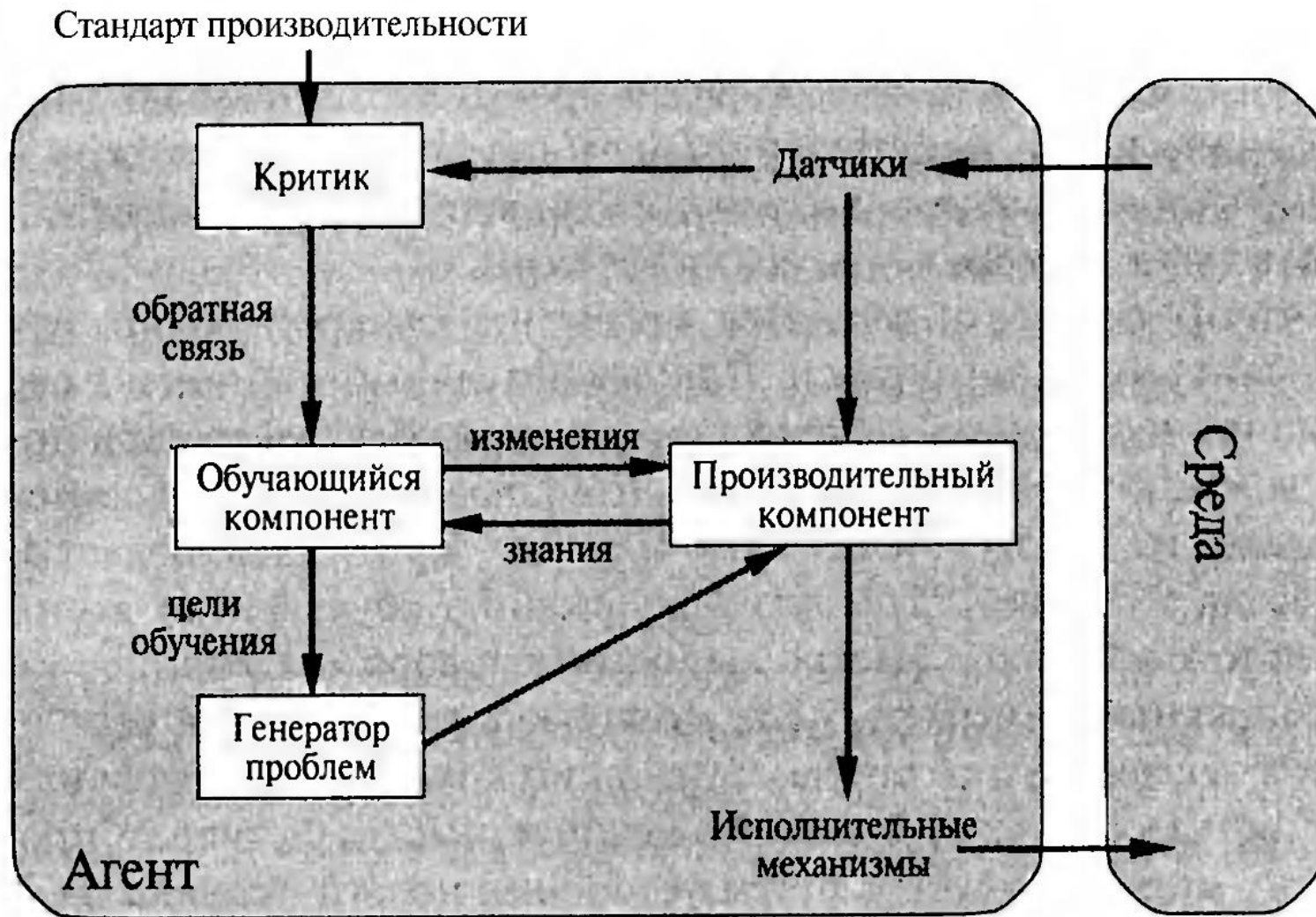
Агент с целью и моделью



Агент, основанный на полезности



Обучающиеся агенты



История AI

- 1950 - статья Алана Тьюринга Computing Machinery and Intelligence.
- 1956 - встреча в Дартмутском колледже (рождение термина AI, Allen Newell, Herbert Simon, Marvin Minsky, John McCarthy, Claude Shannon)
- 1965 - алгоритм резолюции (Робинсон)
- 1969-1980 - развитие методов представления знаний, экспертных систем.
- 1980 – 1988 бум экспертных систем и классического, символьного AI.
- 1988 - ? AI Winter. Коллапс экспертных систем, Lisp машин и многих других направлений и разработок.
- 1986 - алгоритм обратного распространения ошибки, открывавшийся до этого несколько раз, наконец получил должное внимание. Многослойные нейронные сети стали тренируемыми.
- 198*-н.в. AI становится более научно строгим, адаптирует вероятностные и статистические методы. Расцвет машинного обучения.
- 200* - н.в. большие и сверхбольшие наборы данных
- 201* - н.в. революция Deep Learning. Прорывы в зрении, машинном переводе и др.

Цитаты из истории AI

- "Уже сейчас существуют машины, которые думают, обучаются и создают. Более того, их способность к этому будет быстро увеличиваться, пока в недалёком будущем диапазон задач, с которыми они могут справиться будет одинаков с таковым у людей" Herbert Simon, 1957
- "Когда мы доберёмся до отличной игры в шахматы, мы доберёмся до сущности интеллектуального поведения человека" Разные люди, 195* - не позднее 1997.
- "Основываясь на последних разработках и скорости прогресса, я полагаю, что в ближайшие 10 лет:
 1. Компьютер станет чемпионом мира по шахматам
 2. Компьютер обнаружит и докажет новую важную математическую теорему.
 3. Компьютеры будут писать музыку, высоко оцениваемую критиками
 4. Большинство теорий в психологии примут форму компьютерных программ или качественных утверждений о компьютерных программах." Herbert Simon, Allen Newell, 1958

Цитаты из истории AI

- "Область AI показывает повторяющийся паттерн: ранние, драматические успехи, основанные на простейших задачах и низкокачественная работа на сложных, останавливающийся прогресс, разочарование, пессимизм" Hubert Dreyfus, 1965
- "Они говорят, что прогресс есть и определяют его весьма осторожно "движение в сторону цели". Такое определение позволяет первому человеку, залезшему на дерево, заявить, что он прогрессирует на пути к луне" Hubert Dreyfus, 1965

Цитаты из истории AI

Какие изменения в AI были самыми непредсказуемыми для вас?

"Когда я начал в 1980, AI означал следующее: магистрант записывает множество фактов через декларативный язык и мучается с этими фактами и механизмами вывода до получения хороших результатов на тщательно отобранных нескольких примерах, после чего пишет об этом статью. Несмотря на то, что я получил PhD на этом, я видел три проблемы:

- Факты записывались очень долго.
- Не было хороших способов справляться с противоречиями и размытостью
- Процесс не был научен, не было никаких гарантий успеха за пределами отобранных примеров

В настоящее время, область отвечает на эти проблемы следующим образом:

- Мы полагаемся на машинное обучение вместо крови, пота и слёз магистрантов.
- Мы используем вероятностные рассуждения вместо Булевой логики
- От нас требуется научная строгость. Мы используем понятия обучающих и тестовых выборок для сравнения разных систем на стандартных задачах"

Питер Норвиг, 2016, Quora.

AI-задачи

- Решение проблем (комбинаторная оптимизация, эвристические методы, игры)
- Экспертиза по определённым областям
- Сложные системы планирования
- Компьютерное зрение
- Распознавание речи
- Понимание естественного языка

ИИ-полная задача - задача, решение которой может быть эквивалентным решению центральной задачи AI - созданию артефактов, настолько же разумных, как и люди (сильный AI)

Два основных подхода к AI

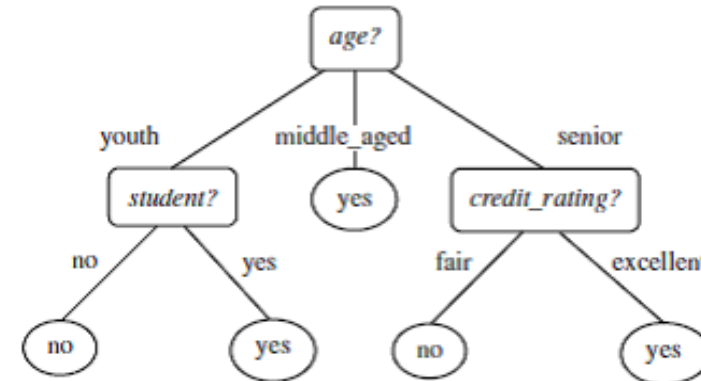
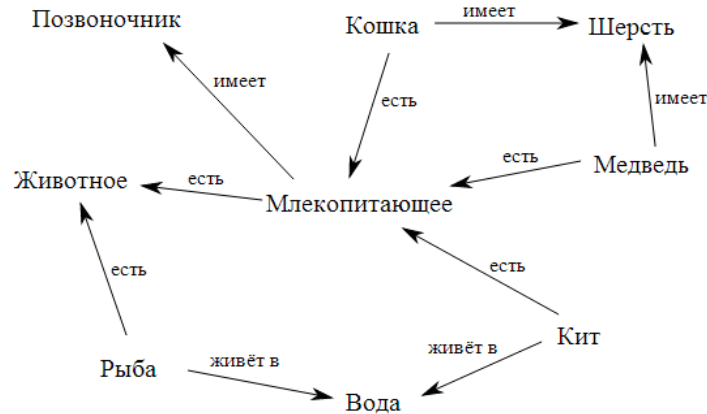
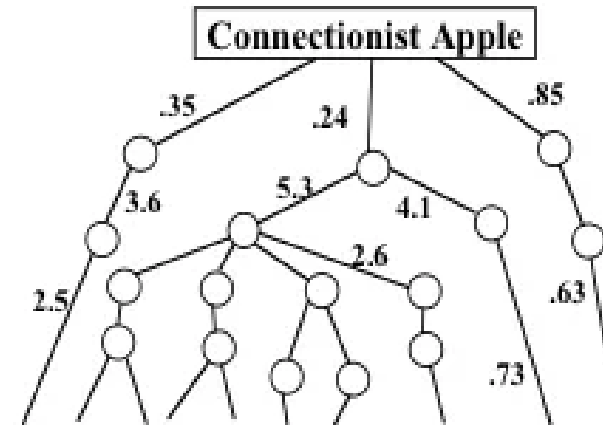
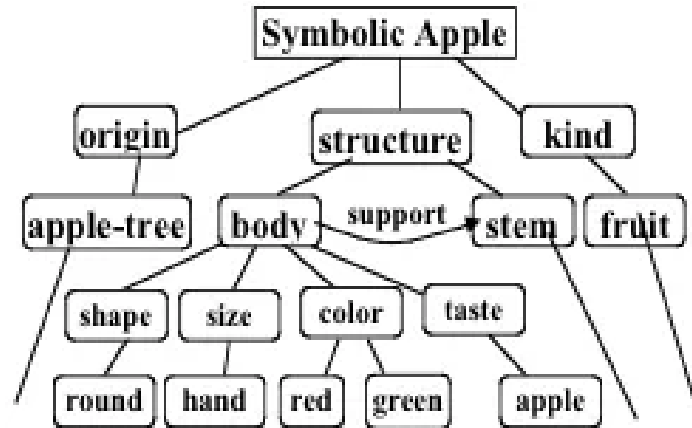
Символьный AI (GOFAI)

- Представления мира и проблемы в виде набора абстрактных и понятных для человека символов
- Явные правила манипуляции символами
- Интерпретируемость, способность объяснить свои решения человеку. Отвечает на вопросы "как?" и "почему?".
- Акцент на логике, дедуктивных рассуждениях

Несимвольный AI

- Минимальные и неявные представления мира (выделенные признаки экземпляра задачи, изображения в виде пикселей, преобразования над ними, связи между искусственными нейронами)
- Модель преобразований, тренируемая экземплярами задачи (вход-выход), неявные правила
- Меньшая интерпретируемость. Отвечает на вопросы "как?"
- Акцент на интуиции и индуктивных рассуждениях

Два основных подхода к AI



$$\forall x \ x \in Persons \Rightarrow [\forall y \ HasMother(x,y) \Rightarrow y \in FemalePersons]$$

Современный AI (достижения)

- Deep Blue - компьютер от IBM, победивший Гарри Каспарова в 1997 году.
- IBM Watson - система ответов на вопросы на естественном языке от IBM. Победитель в Jeopardy ("Своя игра"), 2011. Использует статистические методы и достает ответы из сети.
- AlexNet - свёрточная нейронная сеть, совершившая прорыв в классификации изображений (2012).
- Активно тестируются и задействуются автомобили с автопилотом. Пример: Uber.



AI в повседневной жизни

- Спам-фильтры, новостные агрегаторы
- Обнаружение подозрительных транзакций
- Голосовой поиск, управление, распознавание речи
- Поисковые системы (ранжирование, ответы на вопросы, синонимичные запросы)



Список рекомендуемой литературы

- Russell S., Norvig P. Artificial Intelligence: a Modern Approach 3rd ed., 2010