

# **ИИ и наука о данных**

## **Загрузка и визуализация данных**

**Входной ассессмент для программ**

«Искусственный интеллект и предиктивная аналитика»

«Искусственный интеллект и бизнес-аналитика в реальном секторе  
экономики»

**Лекция 2**

# Искусственный интеллект

Искусственный интеллект (ИИ; англ. *artificial intelligence*, AI) – свойство искусственных интеллектуальных систем выполнять творческие функции, которые традиционно считаются прерогативой человека.

Искусственный интеллект – область знаний, рассматривающая разработку технологий, позволяющих вычислительным системам действовать таким образом, которое напоминает разумное поведение, в том числе, поведение человека.

# Сильный и слабый ИИ

Сильный (также используют термин универсальный) ИИ – искусственная система может приобрести способность мыслить и осознавать себя как отдельную личность (в частности, понимать собственные мысли), хотя и не обязательно, что их мыслительный процесс будет подобен человеческому.

Слабый (также используют термины прикладной или узкий) ИИ предназначен для решения какой-либо одной интеллектуальной задачи или их небольшого множества, которые не подразумевают наличия у компьютера подлинного сознания.

Требования к созданию сильного ИИ:

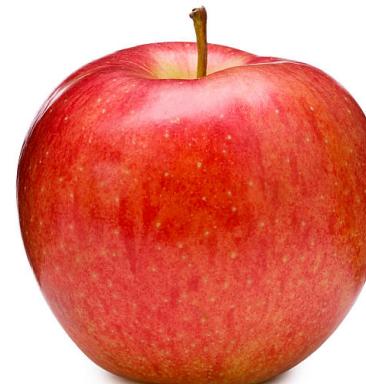
- принятие решений, использование стратегий, решение головоломок и действия в условиях неопределенности;
- представление знаний, включая общее представление о реальности;
- планирование;
- обучение;
- общение на естественном языке;
- сила воли;
- объединение всех этих способностей воедино для достижения общих целей.

# Проблемы создания ИИ

1. Отсутствие понимания устройства и принципов работы мозга и когнитивного аппарата человека.
2. Остаются нерешенными базовые проблемы построения сильного ИИ.
3. Человек и вычислительная система по-разному воспринимают окружающую действительность.

## Человек:

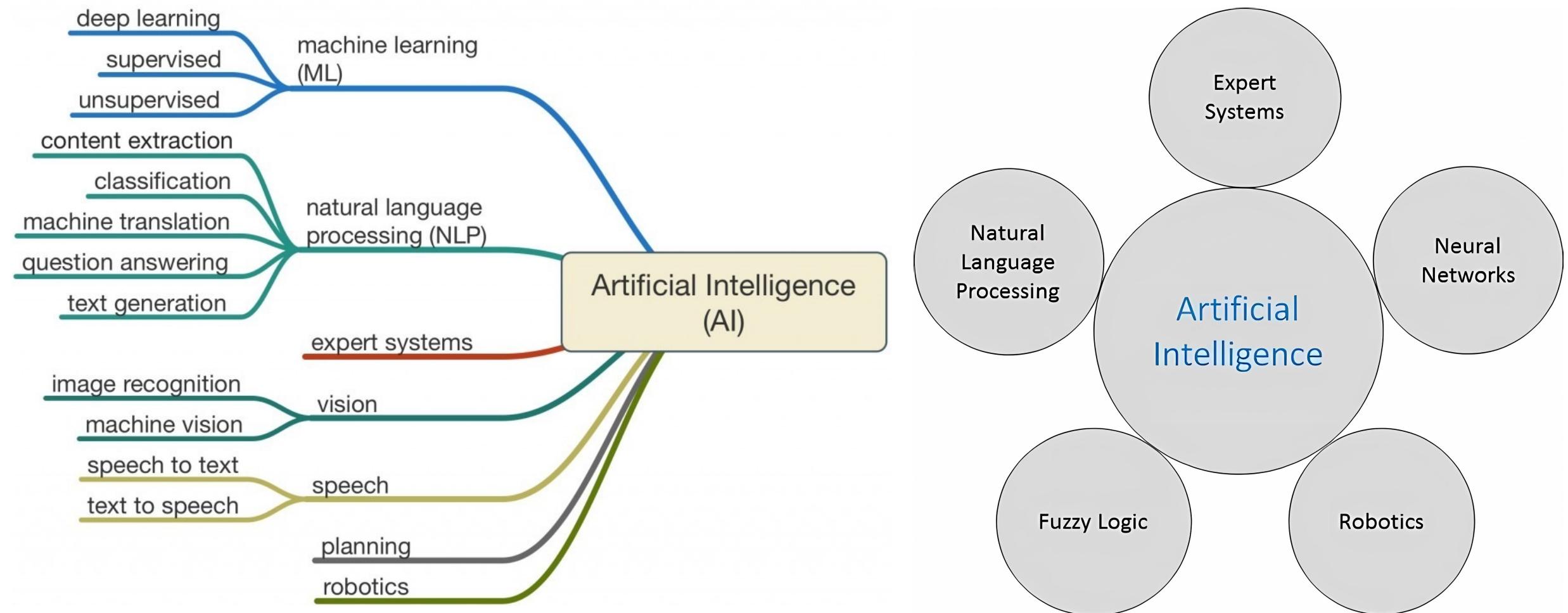
1. Зрительный образ
2. Звуки
3. Запах
4. Тактильные ощущения
5. Вкус
6. Память + опыт + восприятие мира



## Компьютер:

1. Биты (0 или 1): текст, изображение, звуки, 3D модель
2. Мат. модель + БД + БЗ

# Методы ИИ



# Наука о данных - 1

Наука о данных (англ. data science; иногда даталогия — datology) — раздел информатики, изучающий проблемы анализа, обработки и представления данных в цифровой форме.

Наука о данных объединяет:

1. Методы работы с большими данными.
2. Методы параллельной и распределенной обработки данных.
3. Статистические методы анализа данных.
4. Интеллектуальный анализ данных.
5. Методы проектирования и разработки баз данных.

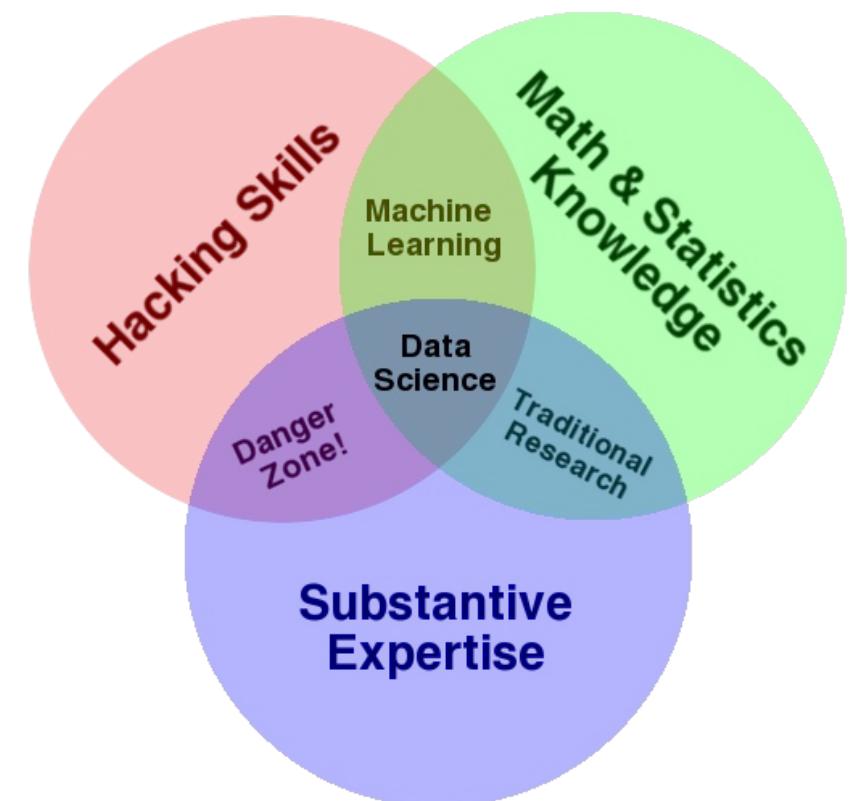
Наука о данных – практическая межотраслевая сфера деятельности.

# Наука о данных - 2

Основная цель применения методов науки о данных – обнаружение закономерностей в данных, извлечение знаний из данных в обобщённой форме.

Основные навыки, которые необходимы для специалиста в области анализа данных:

- знание предметной области + опыт и знания;
- владение информационными технологиями;
- владение методами математической статистики.



# Задача анализа данных

Первоначально задача анализа данных ставится следующим образом:

- имеется достаточно крупная база данных;
- предполагается, что в базе данных находятся некие «скрытые знания».

Свойства «скрытых знаний»:

- ранее неизвестные — то есть такие знания, которые должны быть новыми (а не подтверждающими какие-то ранее полученные сведения);
- нетривиальные — то есть такие, которые нельзя просто так увидеть (при непосредственном визуальном анализе данных или при вычислении простых статистических характеристик);
- практически полезные — то есть такие знания, которые представляют ценность для исследователя или потребителя;
- доступные для интерпретации — то есть такие знания, которые легко представить в наглядной для пользователя форме и легко объяснить в терминах предметной области.

# Этапы анализа данных

1. Постановка задачи анализа.
2. Сбор данных.
3. Подготовка данных (фильтрация, дополнение, кодирование).
4. Выбор модели (алгоритма анализа данных).
5. Подбор параметров модели и алгоритма обучения.
6. Обучение модели (автоматический поиск остальных параметров модели).
7. Анализ качества обучения, если анализ неудовлетворительный – переход на п. 5 или п. 4.
8. Анализ выявленных закономерностей, если анализ неудовлетворительный – переход на п. 1, 4 или 5.

# Требования к данным

1. Данные должны соответствовать решаемой задаче.
2. Методы анализа данных могут быть применены как для работы с большими данными, так и для обработки сравнительно малых объемов данных.  
В качестве критерия достаточного количества данных рассматривается как область исследования, так и применяемый алгоритм анализа.
3. Данные должны быть представлены в удобной для анализа форме: матрицы и/или векторы чисел.