

# Что надо знать об искусственном интеллекте: история, методология, возможности, ограничения (обзорная лекция)

*Воронцов Константин Вячеславович*

д.ф.-м.н., профессор РАН,

зав. кафедрой математических методов прогнозирования МГУ,  
рук. лаб. машинного обучения и семантического анализа Института ИИ МГУ,  
зав. кафедрой машинного обучения и цифровой гуманитаристики МФТИ,  
зав. кафедрой интеллектуальных систем МФТИ,  
г.н.с. ФИЦ «Информатика и управление» РАН

[k.v.vorontsov@phystech.edu](mailto:k.v.vorontsov@phystech.edu)

# Докладчик: *Воронцов Константин Вячеславович*

участник | обсуждение | править

## Участник:Vokov

**Воронцов Константин Вячеславович**  
профессор РАН, д.ф.-м.н.,  
проф., зам. зав. каф. «Математические методы прогнозирования» ВМК МГУ,  
зав. каф. «Машинного обучения и цифровой гуманитаристики» МФТИ,  
проф. каф. «Интеллектуальные системы» МФТИ,  
с.н.с. отдела «Интеллектуальные системы» Вычислительного центра ФИЦ ИУ РАН,  
один из идеологов и Администраторов ресурса MachineLearning.RU,  
прочие подробности — на подстранице Curriculum vitae.

- Профиль ORCID = 0000-0002-4244-4270
- Профиль SCOPUS ID = 6507982932
- Профиль WoS ResearcherID = G-7857-2014
- Профиль Google Scholar
- Профиль DBLP
- Профиль РИНЦ ID = 15081
- Профиль в системе ИСТИНА IRID = 3151446
- Профиль MathNet.ru

**Мне можно написать письмо.**  
<http://www.MachineLearning.ru/wiki?title=User:Vokov>

1 Учебные материалы

**Содержание** [убрать]

- 1 Учебные материалы
  - 1.1 Курсы лекций
  - 1.2 Рекомендации для студентов и аспирантов
- 2 Интервью
  - 2.1 Видео
  - 2.2 Лонгриды
  - 2.3 Российский радиоуниверситет, Радио России
- 3 Доклады на конференциях и семинарах
- 4 Научные интересы
  - 4.1 Анализ текстов и информационный поиск
  - 4.2 Фейковые новости и потенциально опасный дискурс
  - 4.3 Отслеживание контактов и оценка рисков инфицирования
  - 4.4 Теория обобщающей способности
  - 4.5 Комбинаторная (перестановочная) статистика
  - 4.6 Прогнозирование объемов продаж
  - 4.7 Другие проекты и семинары
- 5 Публикации
- 6 Софт и проекты
- 7 Аспиранты и студенты
  - 7.1 Бакалаврские диссертации
  - 7.2 Магистерские диссертации
  - 7.3 Дипломные работы

# Содержание

## 1. Задачи машинного обучения

- Искусственный интеллект — сквозная технология
- Основы: история и терминология машинного обучения
- Примеры задач машинного обучения

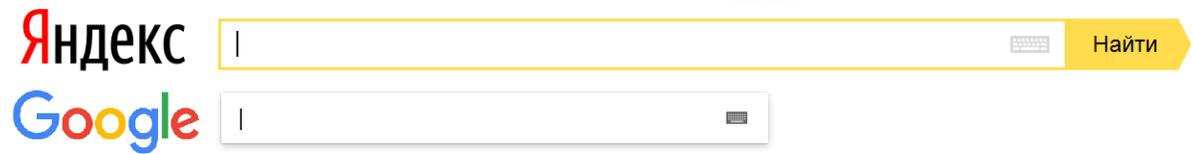
## 2. Методология машинного обучения

- Нейронные сети и глубокое обучение
- Оптимизационные задачи машинного обучения
- Векторизация данных и большие языковые модели

## 3. Применение, перспективы, мифы

- Особенности практического применения технологий ИИ
- Перспективы развития ИИ
- Мифы об искусственном интеллекте

# ИИ — сквозные технологии, меняющие мир



«Четвёртая технологическая революция строится на вездесущем и мобильном Интернете, *искусственном интеллекте* и *машинном обучении*» (2016)

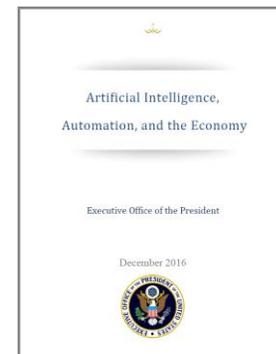
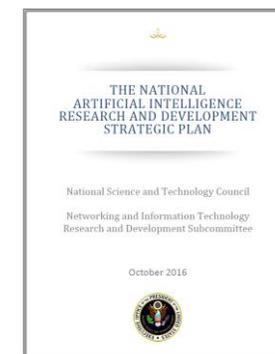
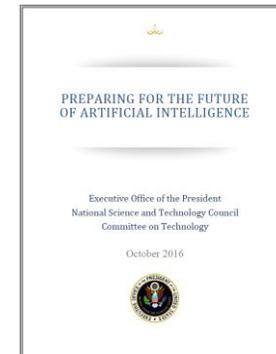
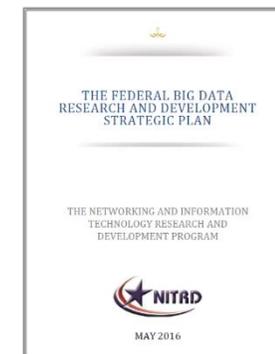
Клаус Мартин Шваб,  
президент Всемирного  
экономического форума



# Отчёты Белого дома США, май-октябрь 2016

«Nations with the strongest presence in AI R&D will establish leading positions in the automation of the future»

Цифровая и распределённая экономика  
Автоматизация и сокращение издержек  
Автономный транспорт и роботизация  
Оптимизация логистики и цепей поставок  
Оптимизация энергетических сетей (Energy Tech)  
Автоматизация банковских услуг (Fin Tech)  
Автоматизация юридических услуг (Legal Tech)  
Автоматизация образовательных услуг (Ed Tech)  
Автоматизация работы с кадрами (HR Tech)  
Персональная медицина (Med Tech)  
Автоматизация в сельском хозяйстве (Agro Tech)  
Автоматизация в природопользовании (Geo Tech)  
Автономные системы вооружений (Mil Tech)



# Национальная стратегия развития ИИ в РФ

10 октября 2019 г.



УКАЗ

ПРЕЗИДЕНТА РОССИЙСКОЙ ФЕДЕРАЦИИ

## О развитии искусственного интеллекта в Российской Федерации

В целях обеспечения ускоренного развития искусственного интеллекта в Российской Федерации, проведения научных исследований в области искусственного интеллекта, повышения доступности информации и вычислительных ресурсов для пользователей, совершенствования системы подготовки кадров в этой области **п о с т а н о в л я ю**:

1. Утвердить прилагаемую Национальную стратегию развития искусственного интеллекта на период до 2030 года.

2. Правительству Российской Федерации:

а) до 15 декабря 2019 г. обеспечить внесение изменений в национальную программу "Цифровая экономика Российской Федерации", в том числе разработать и утвердить федеральный проект "Искусственный интеллект";

б) представлять Президенту Российской Федерации ежегодно доклад о ходе реализации Национальной стратегии развития искусственного интеллекта на период до 2030 года;

в) предусматривать при формировании в 2020 - 2030 годах проектов федеральных бюджетов на очередной финансовый год и на плановый период бюджетные ассигнования на реализацию настоящего Указа.

УТВЕРЖДЕНА  
Указом Президента  
Российской Федерации  
от 10 октября 2019 г. № 490

## НАЦИОНАЛЬНАЯ СТРАТЕГИЯ развития искусственного интеллекта на период до 2030 года

### I. Общие положения

1. Настоящей Стратегией определяются цели и основные задачи развития искусственного интеллекта в Российской Федерации, а также меры, направленные на его использование в целях обеспечения национальных интересов и реализации стратегических национальных приоритетов, в том числе в области научно-технологического развития.

2. Правовую основу настоящей Стратегии составляют Конституция Российской Федерации, Федеральный закон от 28 июня 2014 г. № 172-ФЗ "О стратегическом планировании в Российской Федерации", указы Президента Российской Федерации от 7 мая 2018 г. № 204 "О национальных целях и стратегических задачах развития Российской Федерации на период до 2024 года", от 9 мая 2017 г. № 203 "О Стратегии развития информационного общества в Российской Федерации на 2017 - 2030 годы", от 1 декабря 2016 г. № 642 "О Стратегии научно-технологического развития Российской Федерации" и иные нормативные правовые акты Российской Федерации, определяющие направления применения информационных технологий в Российской Федерации.

3. Настоящая Стратегия является основой для разработки (корректировки) государственных программ Российской Федерации, государственных программ субъектов Российской Федерации, федеральных и региональных проектов, плановых и программно-целевых документов государственных корпораций, государственных компаний, акционерных обществ с государственным участием,



2 100011 98860 6

# Бум искусственного интеллекта

**1997:** IBM Deep Blue обыграл чемпиона мира по шахматам

**2005:** Беспилотный автомобиль: DARPA Grand Challenge

**2006:** Google Translate – статистический машинный перевод

**2011:** 40 лет DARPA CALO привели к созданию Apple Siri

**2011:** IBM Watson победил в ТВ-игре «Jeopardy!»

**2009–2018:** ImageNet: 25% → 2,5% ошибок против 5% у людей

**2015:** Фонд OpenAI в \$1 млрд. Илона Маска и Сэма Альтмана

**2016:** DeepMind, OpenAI: динамическое обучение играм Atari

**2016:** Google DeepMind обыграл чемпиона мира по игре го

**2017:** OpenAI обыграл чемпиона мира по компьютерной игре Dota 2

**2020:** Модель GPT-3 синтезирует тексты, неотличимые от человеческих

**2023:** GPT-4 демонстрирует «проблески общего искусственного интеллекта»



# Машинное обучение, большие данные «и много других страшных слов»

- Статистический анализ данных (Statistical Data Analysis)
- Искусственный интеллект (Artificial Intelligence, AI) 1955
- Распознавание образов (Pattern Recognition)
- Машинное обучение (Machine Learning, ML) 1959
- Статистическое обучение (Statistical Learning)
- Интеллектуальный анализ данных (Data Mining) 1989
- Машинный интеллект (Machine Intelligence) 2000
- Бизнес-аналитика (Business Intelligence, Business Analytics)
- Предсказательная аналитика (Predictive Analytics) 2007
- Большие данные (Big Data) 2008
- Аналитика больших данных (Big Data Analytics)
- Наука о данных (Data Science, DS) 2011

# Методология эмпирической индукции

## От дедуктивного метода познания к индуктивному:

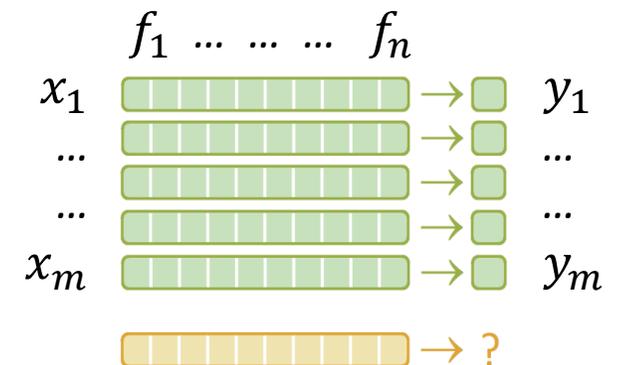
«Не следует полагаться на сформулированные аксиомы и формальные базовые понятия, какими бы привлекательными и справедливыми они не казались. Законы природы нужно «расшифровывать» из фактов опыта. Следует искать правильный метод анализа и обобщения опытных данных; здесь логика Аристотеля не подходит в силу её абстрактности, оторванности от реальных процессов и явлений.»



Фрэнсис Бэкон  
(1561--1626)

«**Таблица открытия**»: множество объектов  $\{x_1, \dots, x_m\}$ :

- $f_j(x_i)$  – измеряемое значение  $j$ -го признака объекта  $x_i$
- $y_i$  – измеряемое значение целевого свойства  $x_i$ , либо  $y_i \in \{0,1\}$  – отсутствие или наличие целевого свойства



# Задача проведения функции через точки

Предсказание свойства  $y(x)$  по признакам  $f_j(x)$ ,  
(линейной) моделью  $a(x, w)$  с параметрами  $w$ :

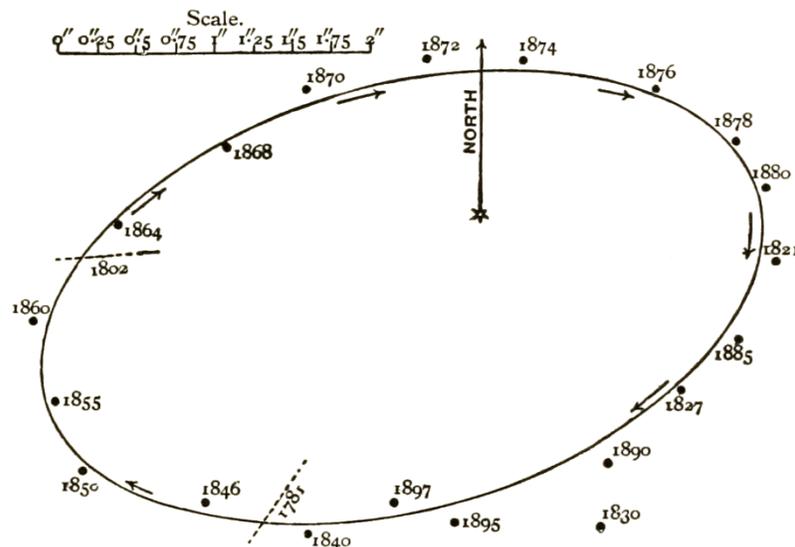
$$a(x, w) = \sum_j w_j f_j(x)$$

**Метод наименьших квадратов (Гаусс, 1795):**

$$\sum_{(x,y)} (a(x, w) - y)^2 \rightarrow \min_w$$



Карл Фридрих Гаусс  
(1777--1855)



«Our principle, which we have made use of since 1795, has lately been published by Legendre...»

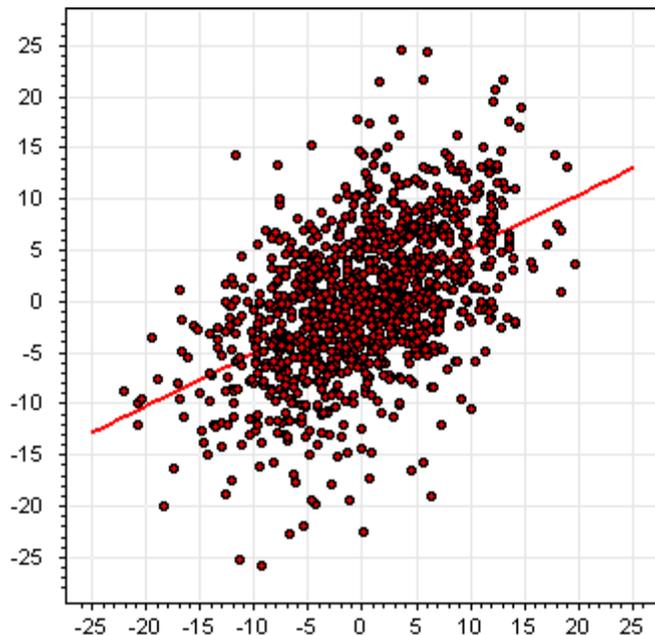
*C.F.Gauss*. Theory of the motion of the heavenly bodies moving about the Sun in conic sections. 1809.

# Задача восстановления регрессии

Исследование наследственности роста (Гальтон, 1886).

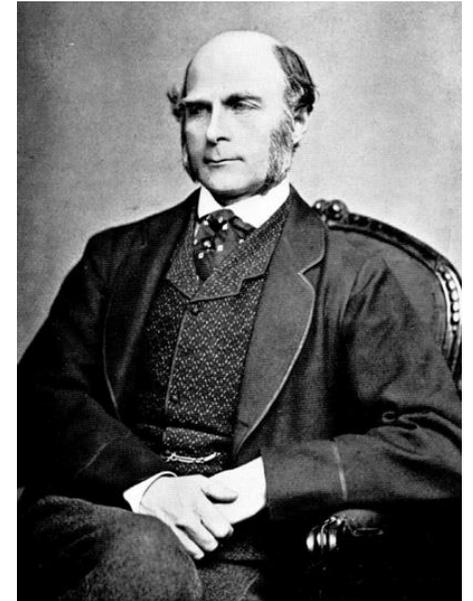
$\Delta$  — отклонение роста от среднего в популяции

Зависимость (линейная?)  $\Delta$  взрослого сына от  $\Delta$  отца:



Двойной смысл термина «регрессия»:

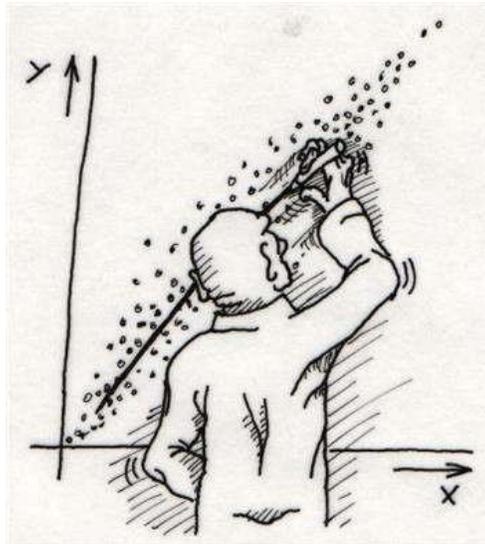
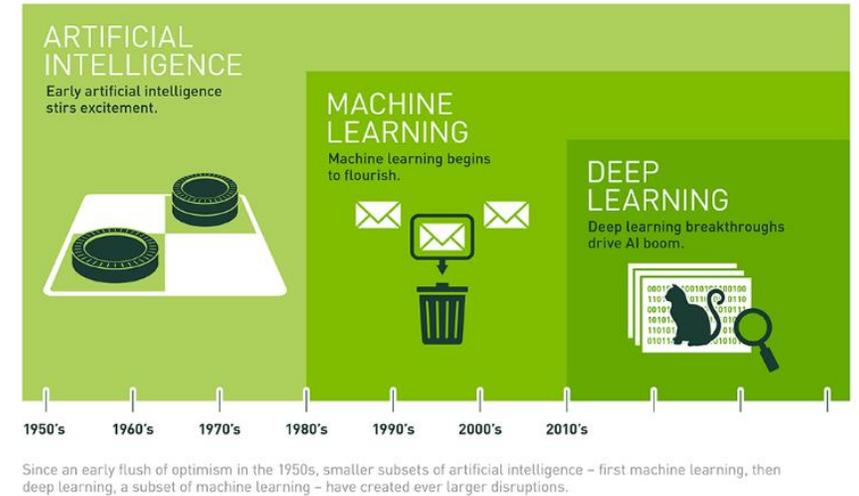
- регрессия (роста) к среднему в популяции
- *необычный «обратный» ход исследования: сначала данные, потом модель*



Фрэнсис Гальтон  
(1822--1911)

# Машинное обучение (Machine Learning, ML)

- одна из ключевых информационных технологий будущего
- наиболее успешное направление ИИ, вытеснившее экспертные системы и инженерию знаний



- проведение функции через заданные точки в сложно устроенных пространствах
- математическое моделирование в условиях, когда знаний мало, данных много
- тысячи различных методов и алгоритмов
- более 100 000 научных публикаций в год

# Задачи машинного обучения с учителем

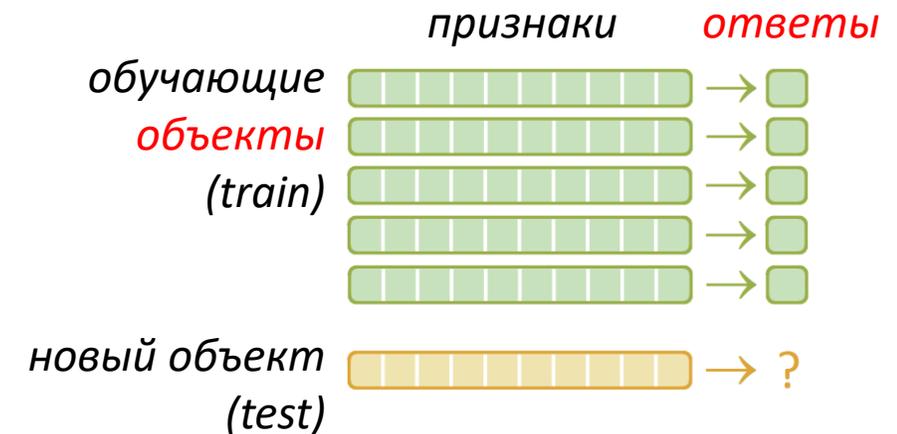
## Этап №1 – обучение с учителем

- **На входе:**  
*данные* – выборка прецедентов «**объект** → **ответ**»,  
каждый объект описывается набором *признаков*
- **На выходе:**  
модель, предсказывающая ответ по объекту

Если нет данных,  
то нет  
и машинного  
обучения

## Этап №2 – применение

- **На входе:**  
*данные* – новый **объект**
- **На выходе:**  
предсказание **ответа** на новом объекте



# Примеры задач машинного обучения

- **Медицинская диагностика:**

**объект** – данные о пациенте на текущий момент

**ответ** – диагноз / лечение / риск исхода



- **Поиск месторождений полезных ископаемых:**

**объект** – данные о геологии района

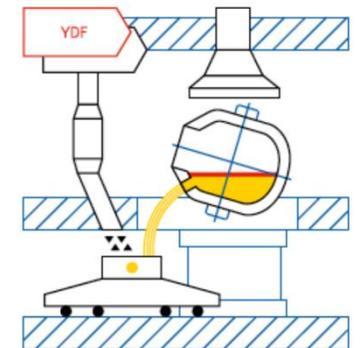
**ответ** – есть/нет месторождение



- **Управление технологическими процессами:**

**объект** – данные о сырье и управляющих параметрах

**ответ** – количество/качество полезного продукта



# Примеры задач ML в бизнесе

- **Кредитный скоринг:**

**объект** – данные о заёмщике

**ответ** – решение по кредиту & вероятность дефолта



- **Предсказание оттока клиентов:**

**объект** – данные о клиенте на момент времени  $t$

**ответ** – уйдёт ли клиент к моменту времени  $t + \Delta$



- **Прогнозирование объёмов продаж:**

**объект** – данные о продажах на момент времени  $t$

**ответ** – объём спроса в интервале от  $t$  до  $t + \Delta$



# Примеры задач ML в интернет-сервисах

- **Информационный поиск в Интернете:**

**объект** – данные о паре «запрос и документ»

**ответ** – оценка релевантности документа запросу



- **Продажа рекламы в Интернете:**

**объект** – данные о тройке «пользователь, страница, баннер»

**ответ** – оценка вероятности клика

- **Рекомендательные системы в Интернете / TV:**

**объект** – данные о паре «пользователь, товар / фильм»

**ответ** – оценка вероятности покупки / просмотра



# Примеры задач с не векторными данными

- **Статистический машинный перевод:**

**объект** – предложение на естественном языке

**ответ** – его перевод на другой язык

- **Перевод речи в текст:**

**объект** – аудиозапись речи человека

**ответ** – текстовая запись речи

- **Беспилотное вождение:**

**объект** – поток данных с радаров и видеокамер

**ответ** – поток решений (объехать, остановиться, игнорировать)

*Прогресс в этих  
областях связан с  
«большими данными»  
(англ. «Big Data»)*

*...очень важное уточнение:  
**с аккуратными**  
большими данными*

# Содержание

## 1. Задачи машинного обучения

- Искусственный интеллект — сквозная технология
- Основы: история и терминология машинного обучения
- Примеры задач машинного обучения

## 2. Методология машинного обучения

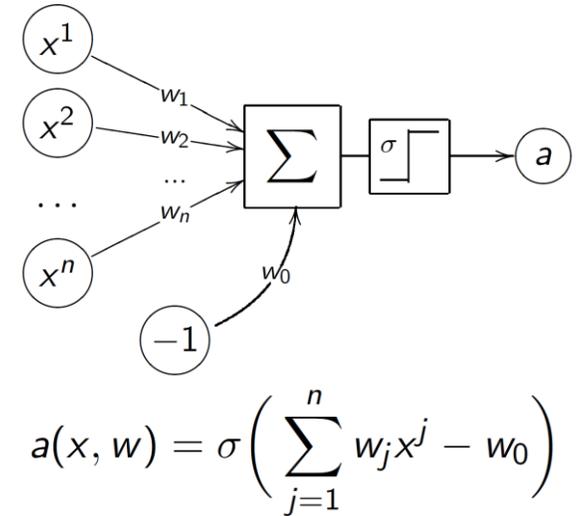
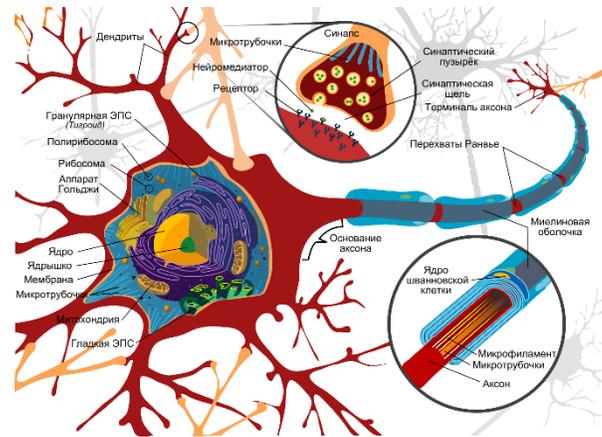
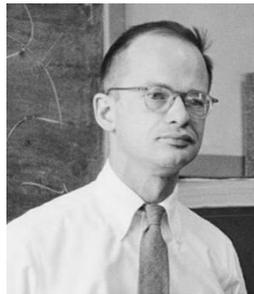
- **Нейронные сети и глубокое обучение**
- **Оптимизационные задачи машинного обучения**
- **Векторизация данных и большие языковые модели**

## 3. Применение, перспективы, мифы

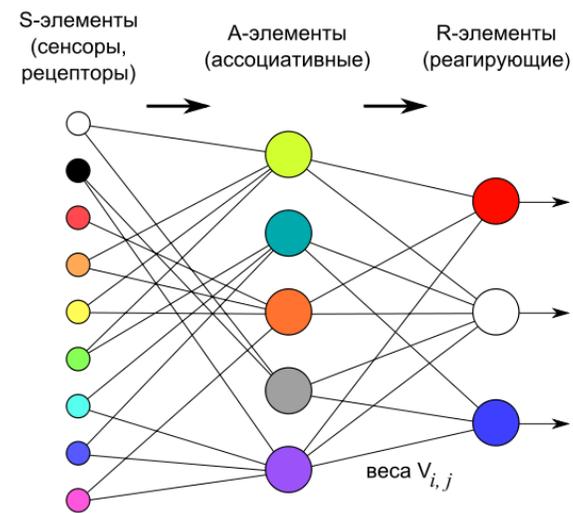
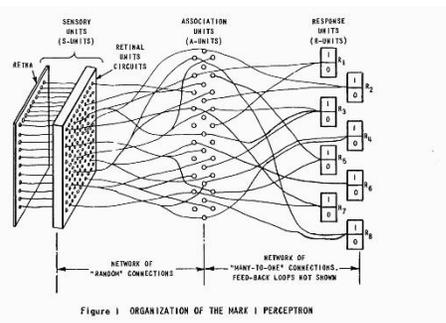
- Особенности практического применения технологий ИИ
- Перспективы развития ИИ
- Мифы об искусственном интеллекте

# Искусственные нейронные сети (ИНС, ANN)

## Математическая модель нейрона (МакКаллок и Питтс, 1943)



## Первый нейрокомпьютер Mark-1 (Фрэнк Розенблатт, 1960)

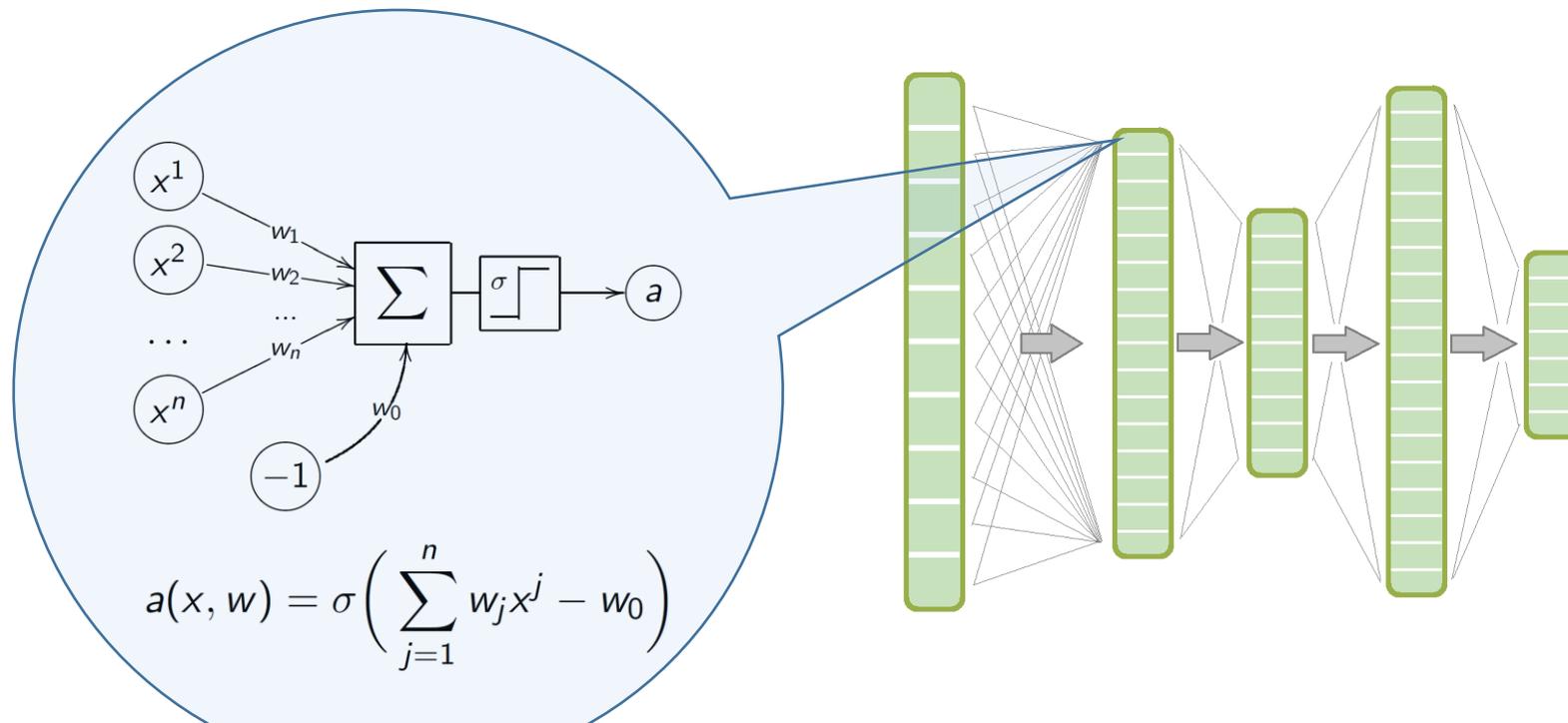


# Многослойные искусственные нейросети

На каждом слое сети вектор объекта преобразуется в новый вектор

Каждое преобразование (нейрон) – линейная модель  $a(x, w)$

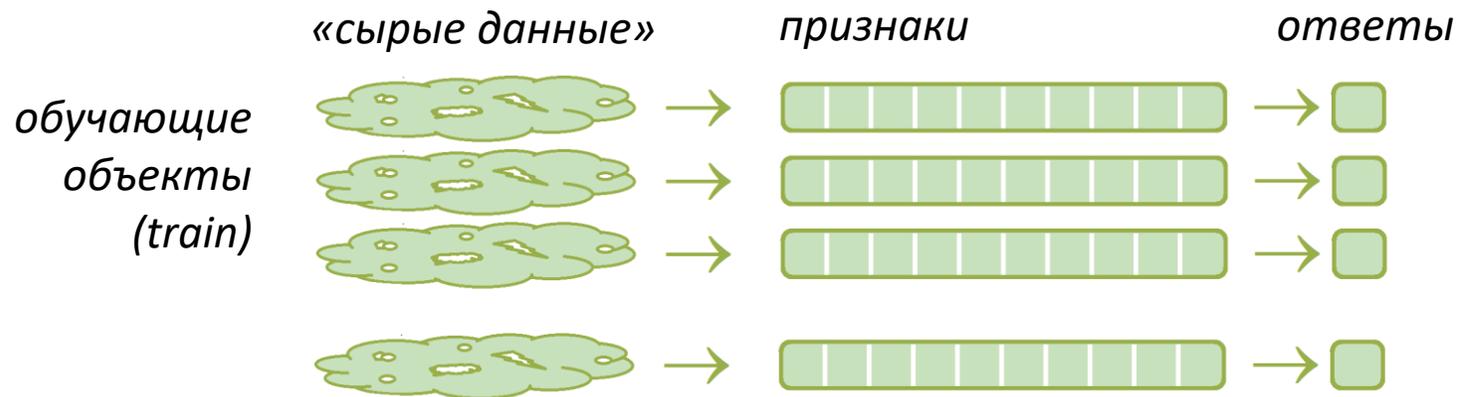
Веса  $w$  являются обучаемыми параметрами модели



# Глубокие нейронные сети (Deep ANN)

**Вход:** сложно структурированные «сырые» данные объектов

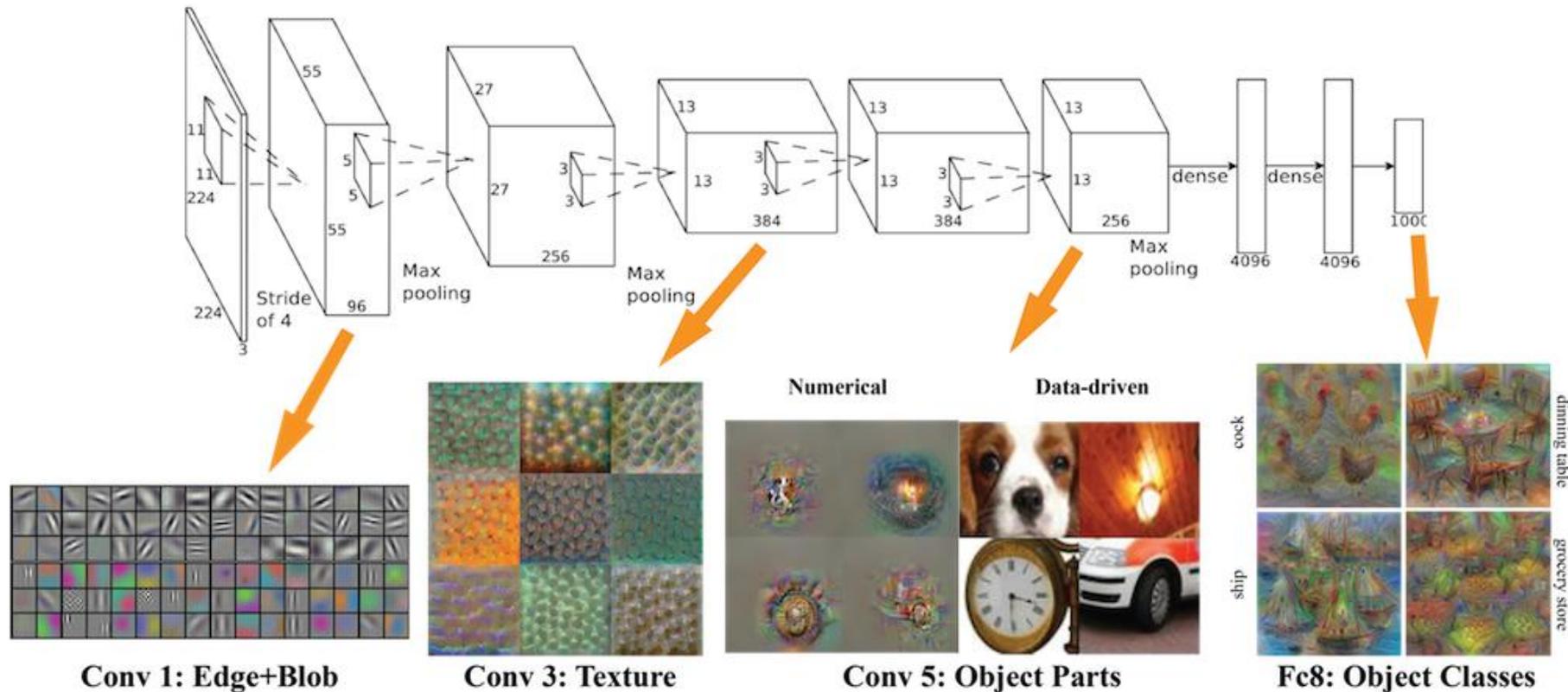
**Выход:** векторные представления объектов, затем ответы



*Deep Learning – это  
всего лишь обучаемая  
векторизация  
сложных объектов*

**Примеры** сложно структурированных объектов: изображения, видео, временные ряды, тексты, последовательности, транзакции, графы, ...

# Глубокие свёрточные нейронные сети (CNN) для классификации объектов на изображениях

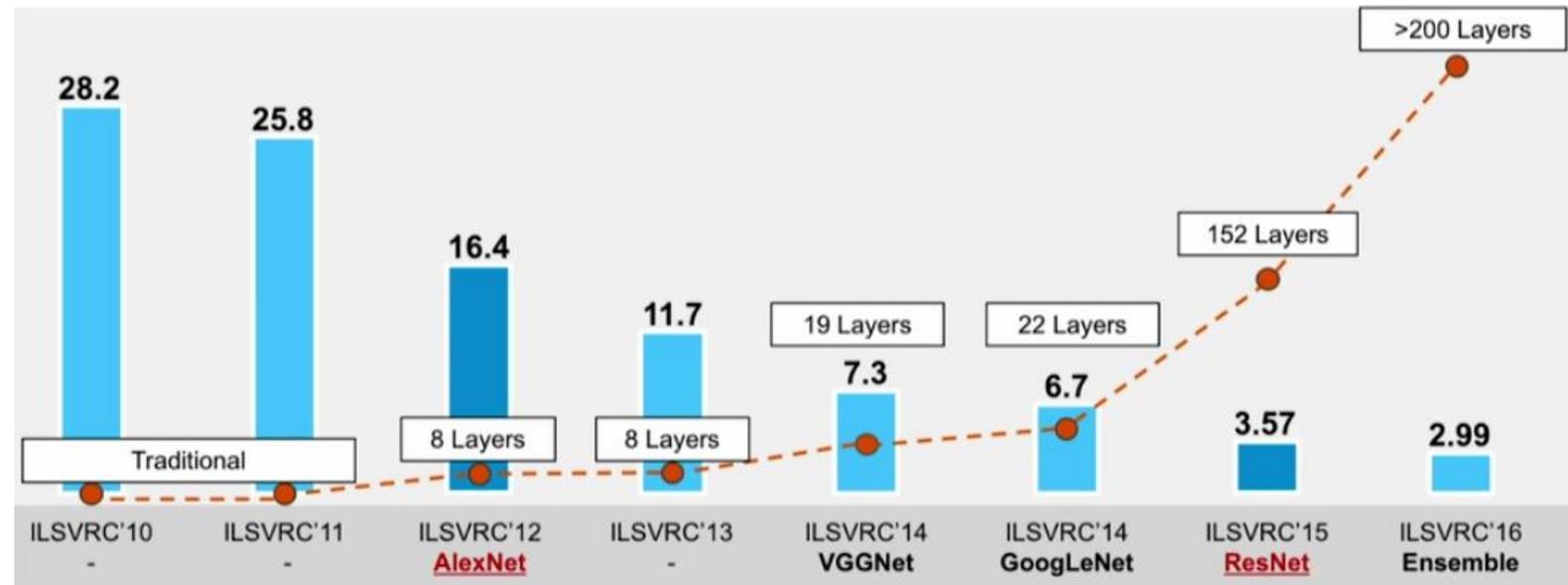


Krizhevsky A., Sutskever I., Hinton G. ImageNet classification with deep convolutional neural networks. 2012.

# Роль больших данных

**ImageNet:** открытая выборка 14М изображений, 20К категорий

IMAGENET



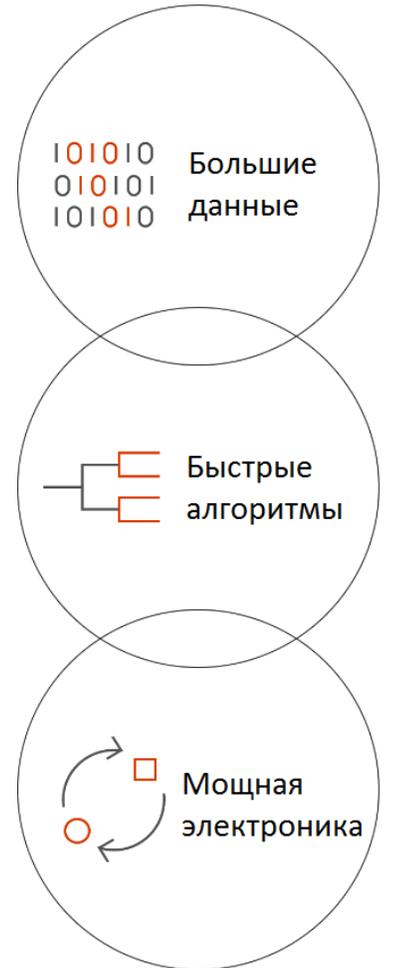
Старт в 2009 г. Человеческий уровень ошибок 5% пройден в 2015 г.

*Li Fei-Fei et al.* ImageNet: A large-scale hierarchical image database. 2009.

*Li Fei-Fei et al.* Construction and analysis of a large scale image ontology. 2009.

# Три составляющих успеха Deep Learning

- Повсеместное применение компьютерных технологий  
→ *накопление больших выборок данных*  
*в частности, ImageNet*
- Развитие математических методов и алгоритмов  
→ *накопление критической массы опыта*  
*методы оптимизации, контроль переобучения*
- Достижения микроэлектроники  
→ *рост вычислительных мощностей по закону Мура*  
*в частности, GPU*



# Машинное обучение – это оптимизация

$x$  – вектор объекта обучающей выборки

$a(x, w)$  – предсказательная модель

$w$  – параметры модели

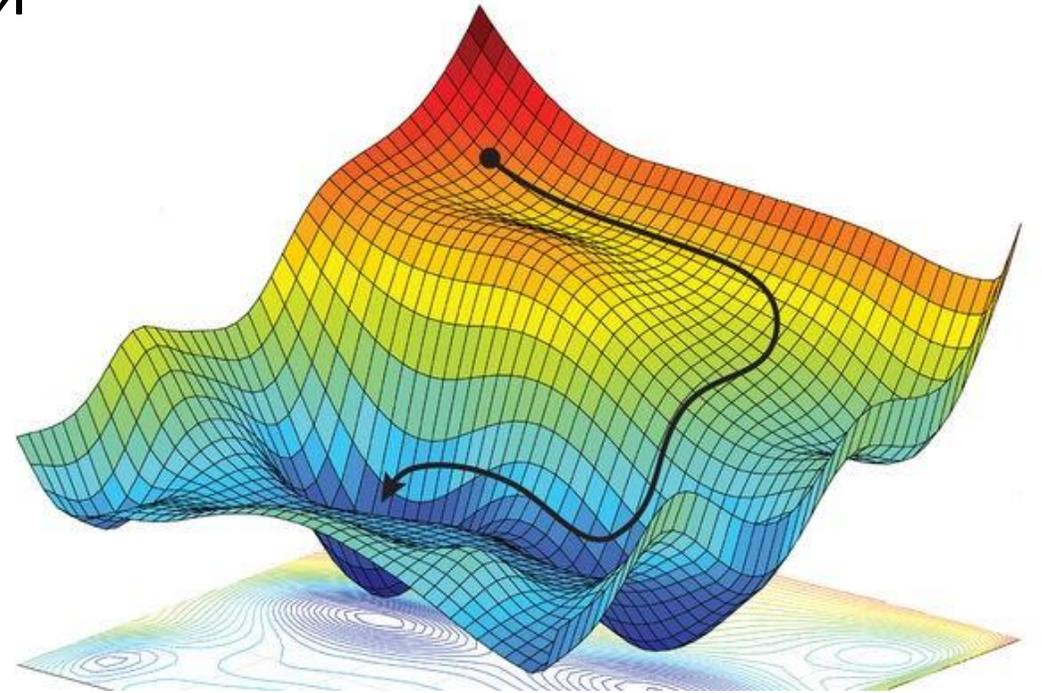
$\text{Loss}(x, w)$  – функция потерь

$Q(w)$  – критерий качества модели

Задача обучения параметров модели:

$$Q(w) = \sum_x \text{Loss}(x, w) \rightarrow \min$$

Способ решения – численные методы оптимизации



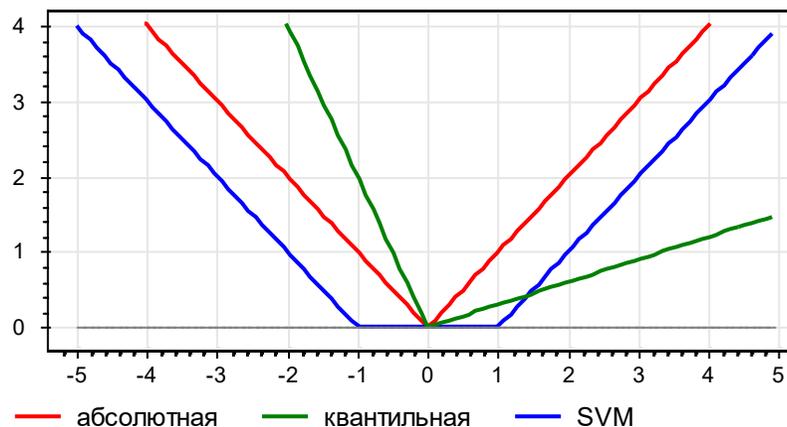
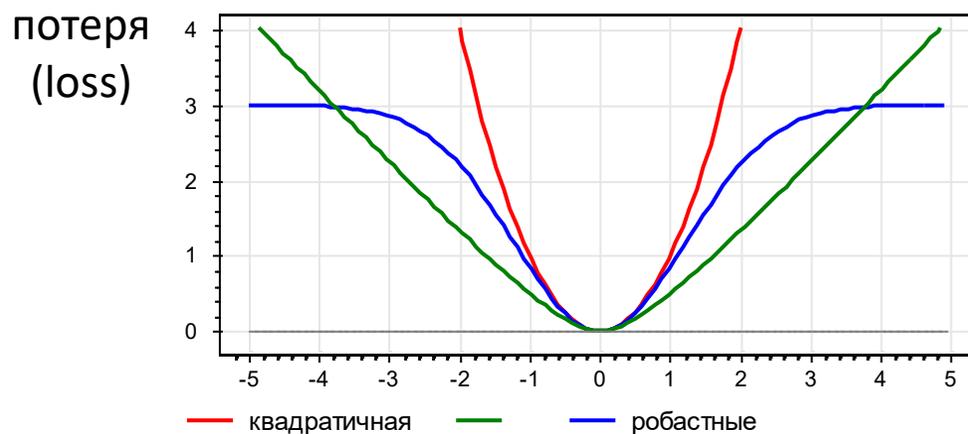
# Обучение с учителем (supervised learning): восстановление регрессии (regression)

$x$  — вектор объекта обучающей выборки,  $y$  — числовой ответ

$a(x, w)$  — модель регрессии с параметрами  $w$

Например,  $a(x, w) = \sum_j w_j x_j$  — линейная модель регрессии

$\text{Loss}(x, w) = (a(x, w) - y)^2$  — квадратичная функция потерь



НЕВЯЗКА  
(error)

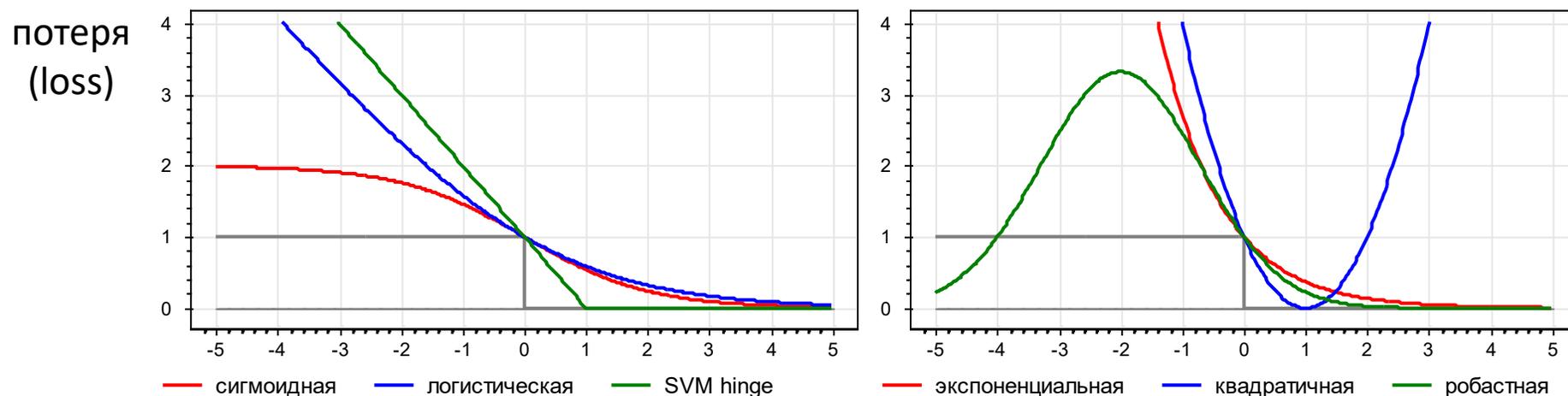
# Обучение с учителем (supervised learning): классификация (classification)

$x$  — вектор объекта обучающей выборки,  $y$  — ответ (+1 или -1)

$a(x, w)$  — модель классификации с параметрами  $w$

Например,  $a(x, w) = \text{sign}(\sum_j w_j x_j)$  — линейная модель

$\text{Loss}(x, w) = \max(0, 1 - y \sum_j w_j x_j)$  — функция потерь SVM hinge



отступ  
(margin)

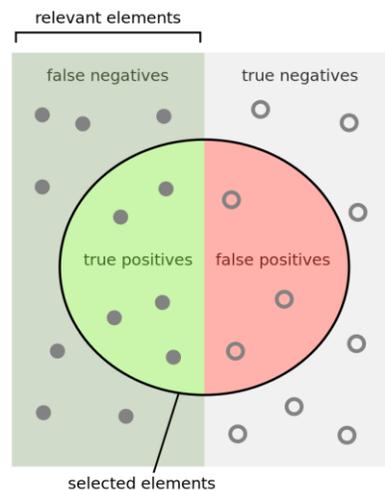
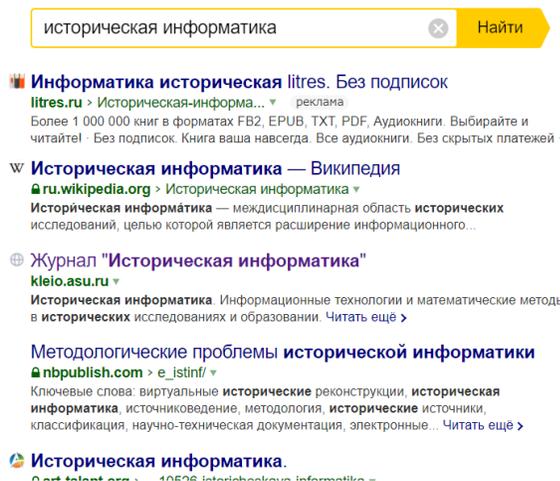
# Обучение с учителем (supervised learning): ранжирование (learning to rank)

$x$  — вектор пары «запрос-документ»,  $y$  — оценка релевантности

$a(x, w)$  — модель ранжирования документов по запросу, параметр  $w$

Например,  $a(x, w) = \sum_j w_j x_j$  — линейная модель

$$\text{Loss}(x, x', w) = \max\left(0, 1 - [y > y'](a(x, w) - a(x', w))\right)$$



$$\text{Precision} = \frac{\text{true positives}}{\text{true positives} + \text{false positives}}$$
$$\text{Recall} = \frac{\text{true positives}}{\text{true positives} + \text{false negatives}}$$

*не только поиск,  
но и любые задачи, где  
человеку удобно  
принимать решения,  
выбирая один из вариантов*

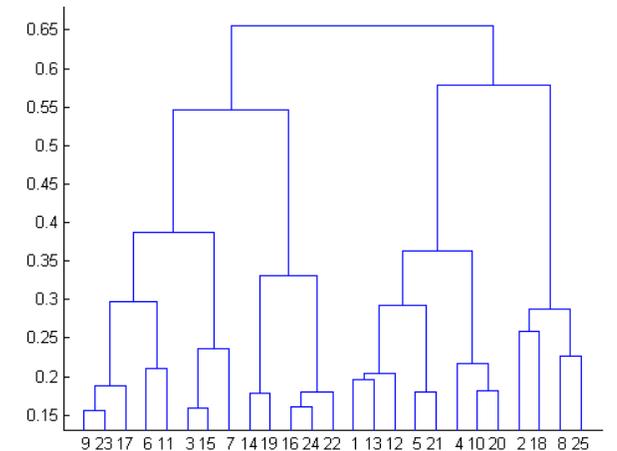
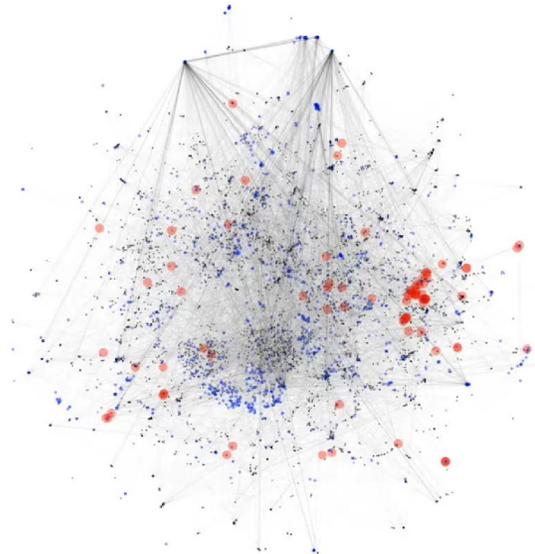
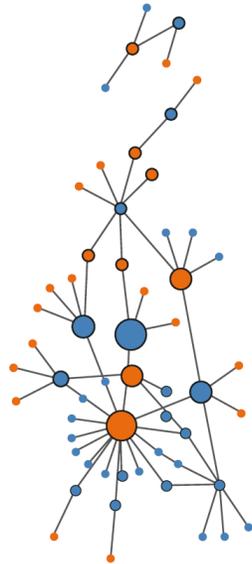
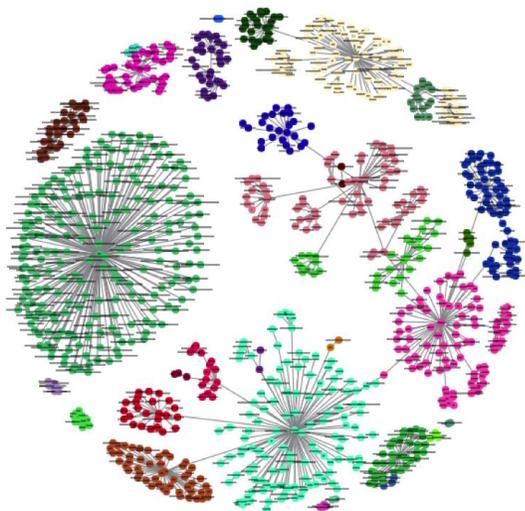
# Обучение без учителя (unsupervised learning): кластеризация (clustering)

$x$  — вектор объекта обучающей выборки, ответы не задаются

$a(x, w)$  — кластер, ближайший к  $x$

$w = \{c_1, \dots, c_K\}$  — векторы центров всех кластеров

$\text{Loss}(x, w) = \min_k \|x - c_k\|$  — расстояние до ближайшего кластера



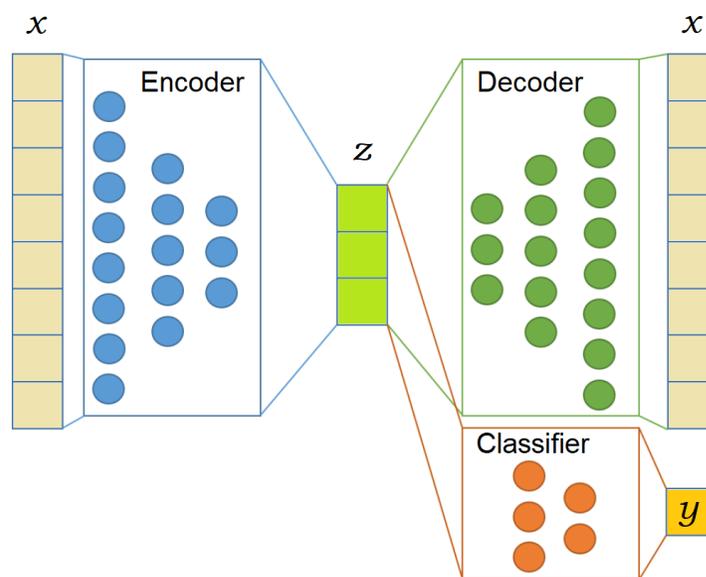
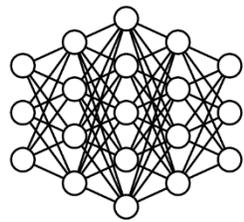
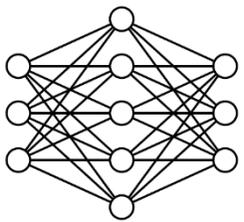
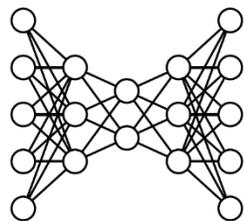
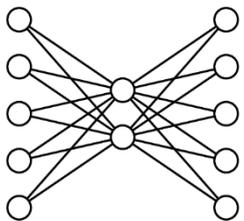
# Обучение без учителя (unsupervised learning): векторизация, автокодировка (autoencoder)

$x$  — описание объекта обучающей выборки, ответов не дано

$z = f(x, w)$  — модель кодирования  $x$  в векторное представление  $z$

$x' = g(z, w')$  — модель декодирования  $z$  в реконструкцию  $x'$

$\text{Loss}(x, w) = \|g(f(x, w), w') - x\|$  — точность реконструкции объекта



обучаемая  
векторизация  
сложных  
объектов

# Перенос обучения (transfer learning), предобучение модели векторизации

$z = f(x, w)$  — модель векторизации, универсальная для многих задач

$y = g(z, w')$  — часть модели, специфичная для своей задачи

$\min_{w, w'}: \sum_x \text{Loss}_1(g_1(f(x, w), w'))$  — обучение по большим данным

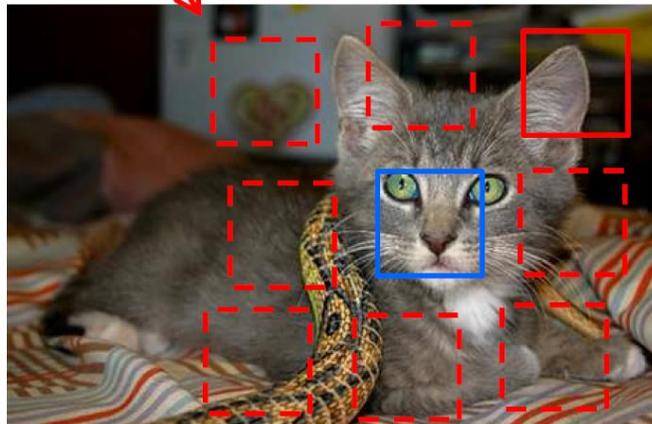
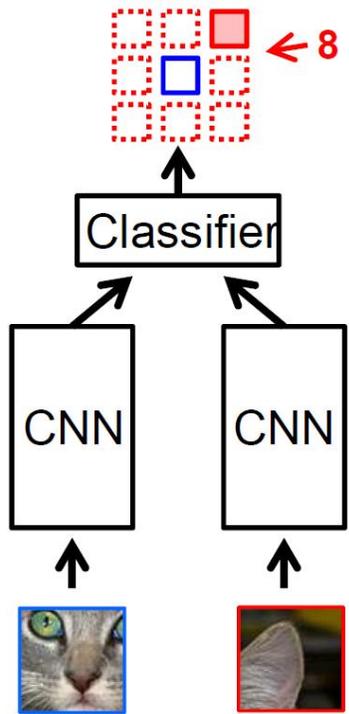
$\min_{w'}: \sum_{x'} \text{Loss}_2(g_2(f(x', w), w'))$  — обучение по своим данным



# Самостоятельное обучение (self-supervised)

$x$  — изображение

$z = f(x, w)$  — модель векторизации, обучается предсказывать взаимное расположение пар фрагментов одного изображения



Randomly Sample Patch  
Sample Second Patch

Unsupervised visual representation learning by context prediction,  
Carl Doersch, Abhinav Gupta, Alexei A. Efros, ICCV 2015

## Преимущество:

сеть выучивает векторные представления объектов без размеченной обучающей выборки

# Многозадачное обучение (multi-task learning)

$z = f(x, w)$  – модель векторизации, универсальная для всех задач

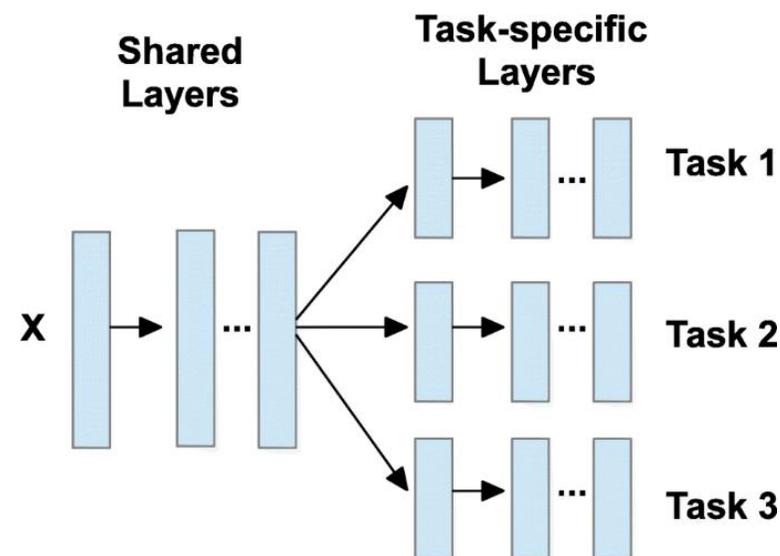
$y = g_t(z, w'_t)$  – часть модели, специфичная для  $t$ -й задачи

$\min_{w, w'_t} \sum_t \sum_x \text{Loss}_t(g_t(f(x, w), w'_t))$  – обучение по всем задачам

*few-shot learning* – обучение по малому числу примеров

*M.Crawshaw.* Multi-task learning with deep neural networks: a survey. 2020

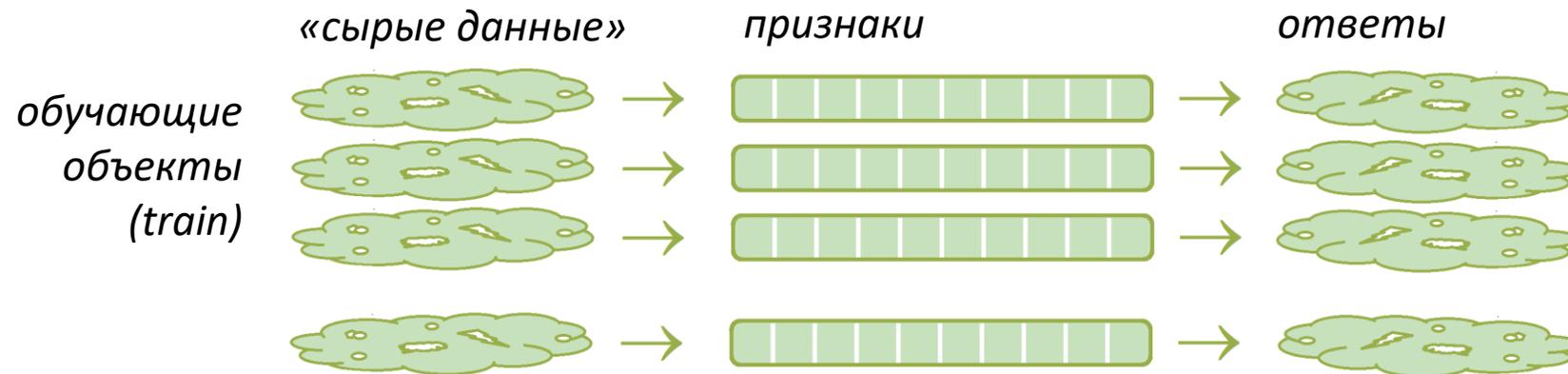
*Y.Wang et al.* Generalizing from a few examples: a survey on few-shot learning. 2020



# Нейронные сети для синтеза объектов

**Вход:** сложно структурированные объекты

**Выход:** сложно структурированные ответы



похоже на  
автокодировщиков

**Примеры:** синтез изображений, перенос стиля, распознавание речи, машинный перевод, суммаризация текстов, диалог с пользователем

**Модели:** seq2seq, CNN, RNN, LSTM, GAN, BERT, GPT и др.

# Генеративная состязательная сеть (GAN)

$x = g(z, w)$  — модель генерации реалистичного объекта  $x$  из шума  $z$

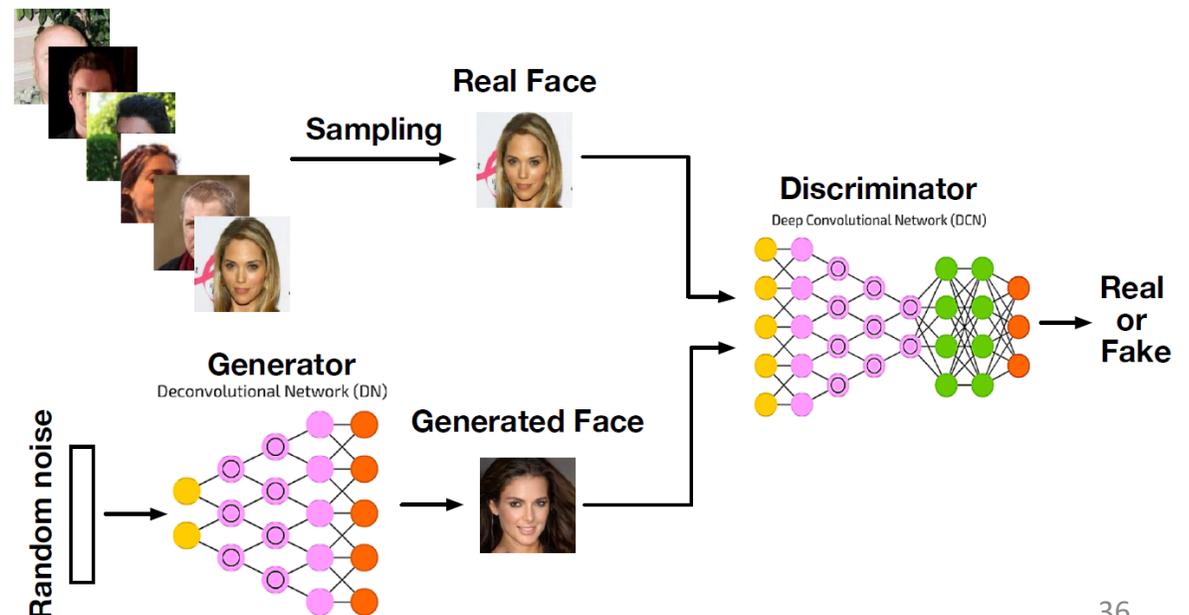
$f(x, w')$  — модель классификации  $x$  «реальный/сгенерированный»

$\min_w \max_{w'} \sum_x \ln f(x, w') + \ln (1 - f(g(z, w), w'))$  — совместное обучение

*Antonia Creswell et al.* Generative Adversarial Networks: an overview. 2017.

*Zhengwei Wang et al.* Generative Adversarial Networks: a survey and taxonomy. 2019.

*Chris Nicholson.* A Beginner's Guide to Generative Adversarial Networks. 2019.



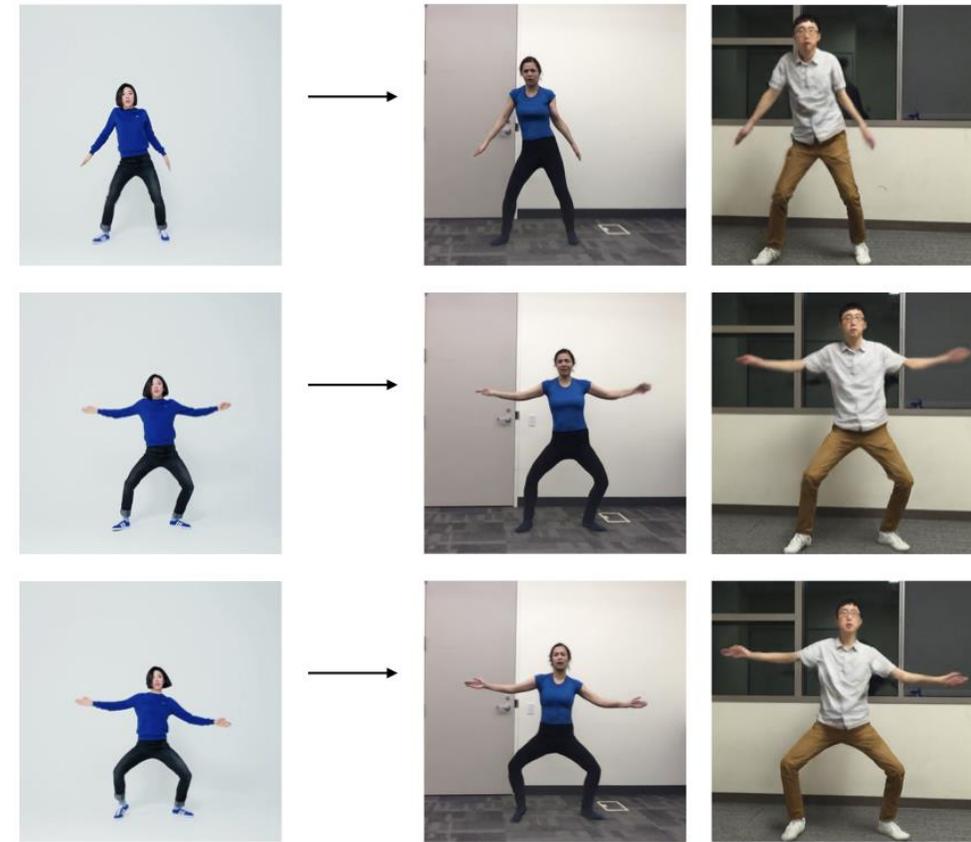
# Синтез изображений и видео



(d) input image

(e) output 3d face

(f) textured 3d face



Source Subject

Target Subject 1

Target Subject 2

# Эволюция подходов в обработке текстов

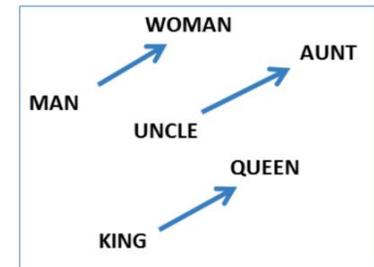
## Декомпозиция задач по уровням «пирамиды NLP»

- морфологический анализ, лемматизация, опечатки, ...
- синтаксический анализ, выделение терминов, NER, ...
- семантический анализ, выделение фактов, тем, ...



## Модели векторизации слов (эмбедингов)

- модели дистрибутивной семантики: word2vec [Mikolov, 2013], FastText [Bojanowski, 2016], ...
- тематические модели LDA [Blei, 2003], ARTM [2014], ...



## Нейросетевые модели контекстной векторизации

- рекуррентные нейронные сети: LSTM, GRU, ...
- «end-to-end» модели внимания и трансформеры: машинный перевод [2017], BERT [2018], GPT-4 [2023], ...

$$\text{softmax} \left( \frac{\begin{matrix} Q \\ \text{grid} \end{matrix} \times \begin{matrix} K^T \\ \text{grid} \end{matrix}}{\sqrt{d}} \right) \begin{matrix} V \\ \text{grid} \end{matrix}$$

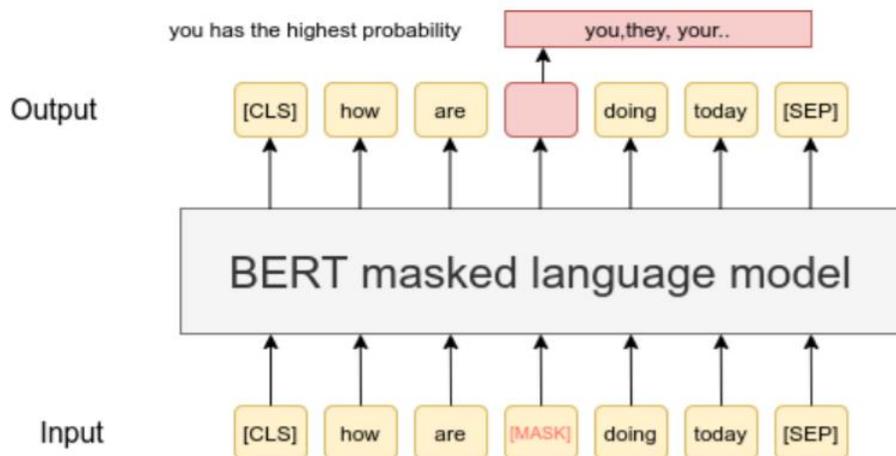
# Обучение контекстной векторизации слов

$x_i$  — слово на  $i$ -й позиции в коллекции текстовых документов

$z_i = f(x_i, C_i, w)$  — модель векторизации слова  $x_i$  по контексту  $C_i$

$p(x|i, z, w')$  — вероятностная модель предсказания слова по вектору  $z$

$\text{Loss}(x_i, w) = -\ln p(x_i|i, f(x_i, C_i, w), w')$  — потеря от предсказания слова на  $i$ -й позиции по его контексту (Masked Language Model)

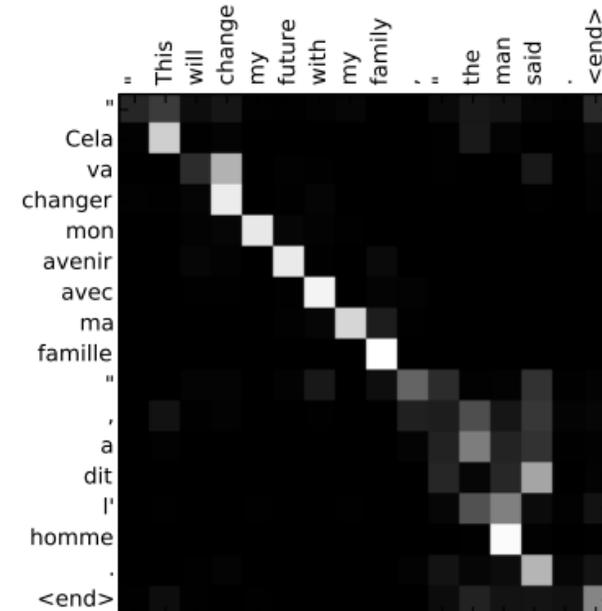
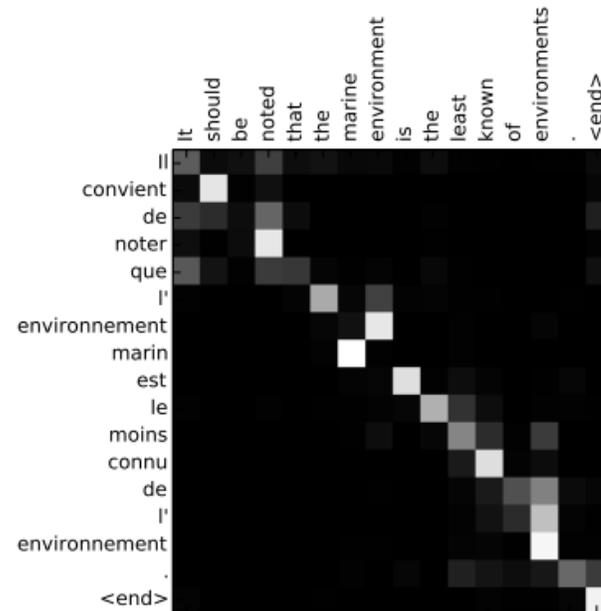
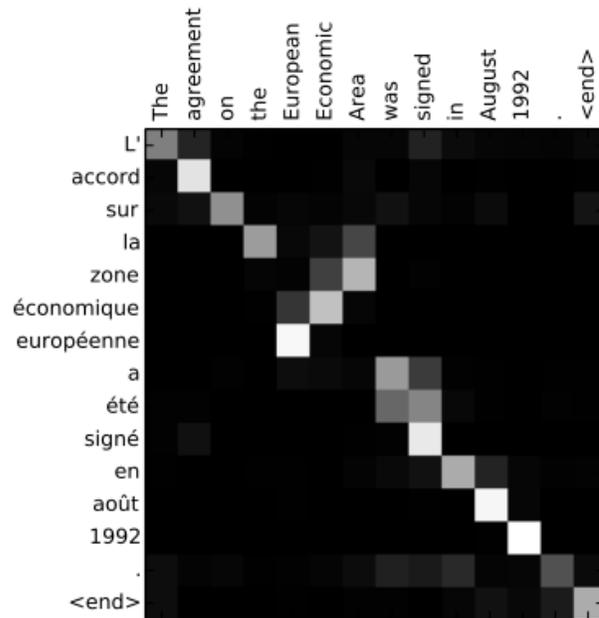


*Vaswani et al. (Google) Attention is all you need. 2017.*

*Jacob Devlin et al. (Google AI Language)*

*BERT: pre-training of deep bidirectional transformers for language understanding. 2019.*

# Модели внимания: машинный перевод



**Интерпретация** моделей внимания: *матрица семантического сходства*  $A[t,i]$  показывает, на какие слова  $x[i]$  входного текста модель обращает внимание, когда генерирует слово перевода  $y[t]$

# Модели внимания: аннотирование изображений



A woman is throwing a frisbee in a park.



A dog is standing on a hardwood floor.



A stop sign is on a road with a mountain in the background.



A little girl sitting on a bed with a teddy bear.



A group of people sitting on a boat in the water.

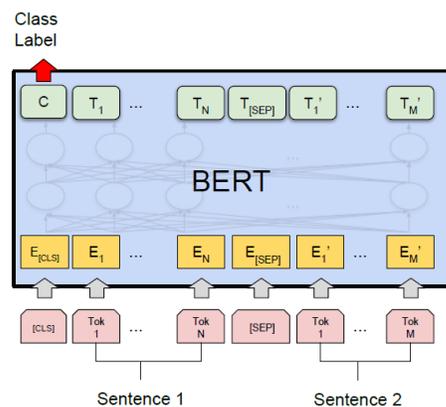


A giraffe standing in a forest with trees in the background.

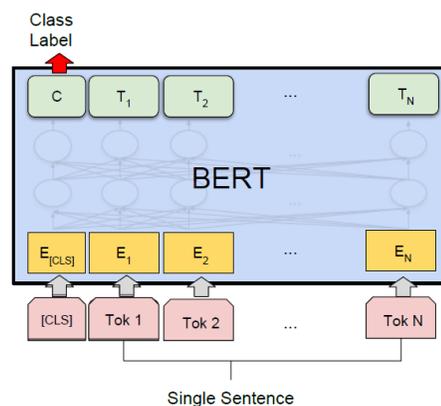
**Интерпретация:** на какие области модель обращает внимание, генерируя подчёркнутое слово в описании изображения

# Трансформеры: большие языковые модели

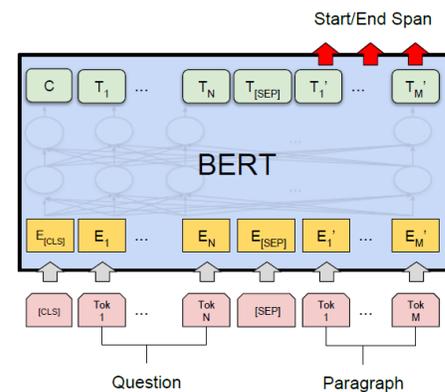
- обучаются векторизовать и предсказывать слова по контексту
- обучаются по терабайтам текстов, «они видели в языке всё»
- мультиязычны: обучаются на десятках языков
- мультизадачны: для каждой новой задачи NLP/NLU достаточно предобученной модели или дообучения на небольшой выборке



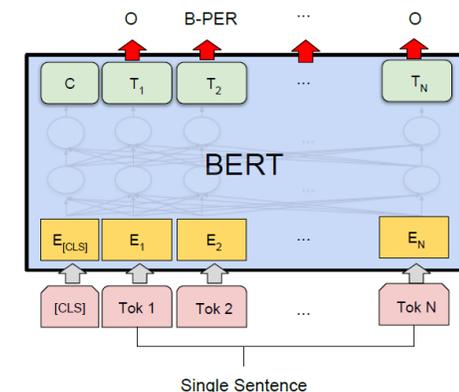
(a) Sentence Pair Classification Tasks:  
MNLI, QQP, QNLI, STS-B, MRPC,  
RTE, SWAG



(b) Single Sentence Classification Tasks:  
SST-2, CoLA



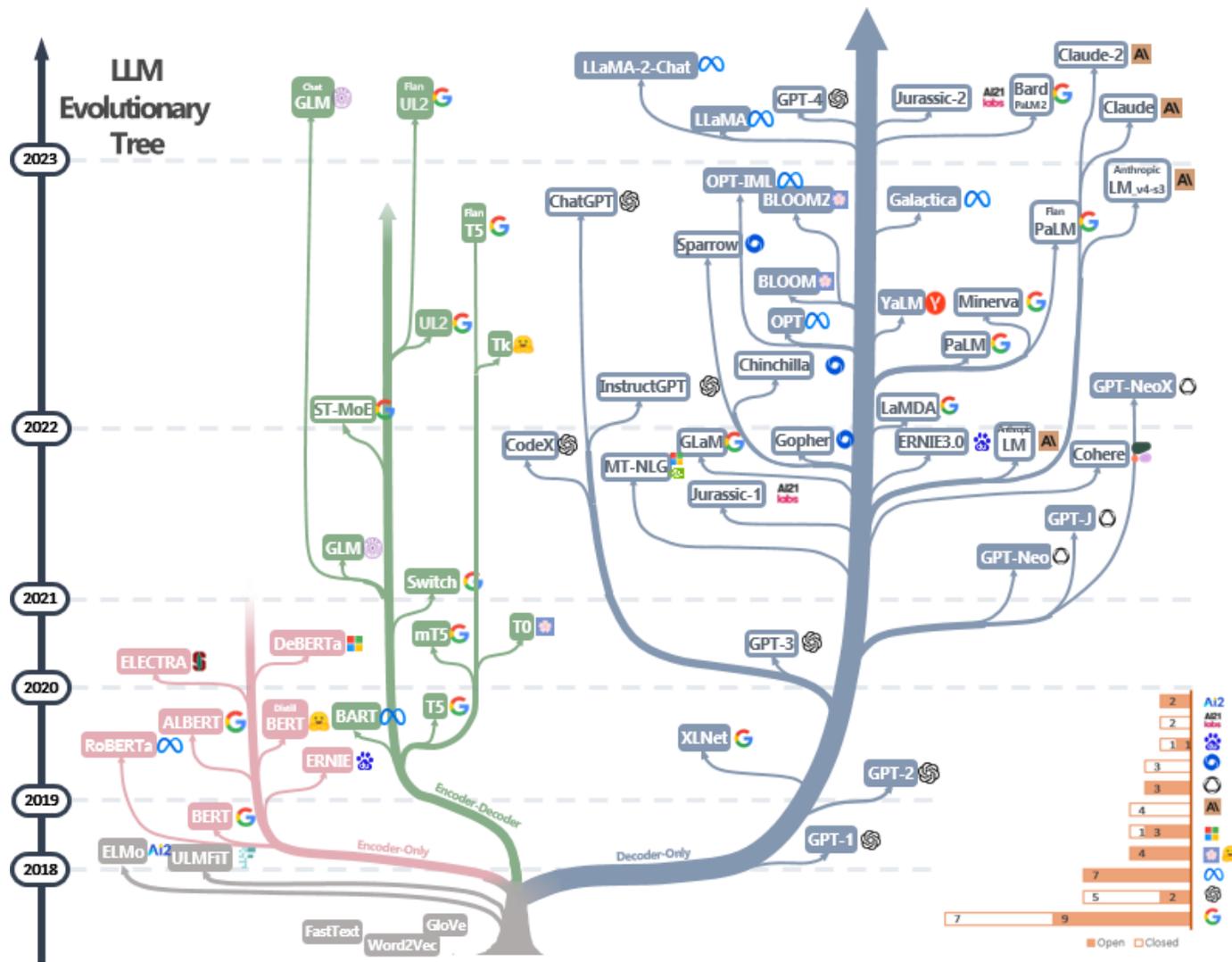
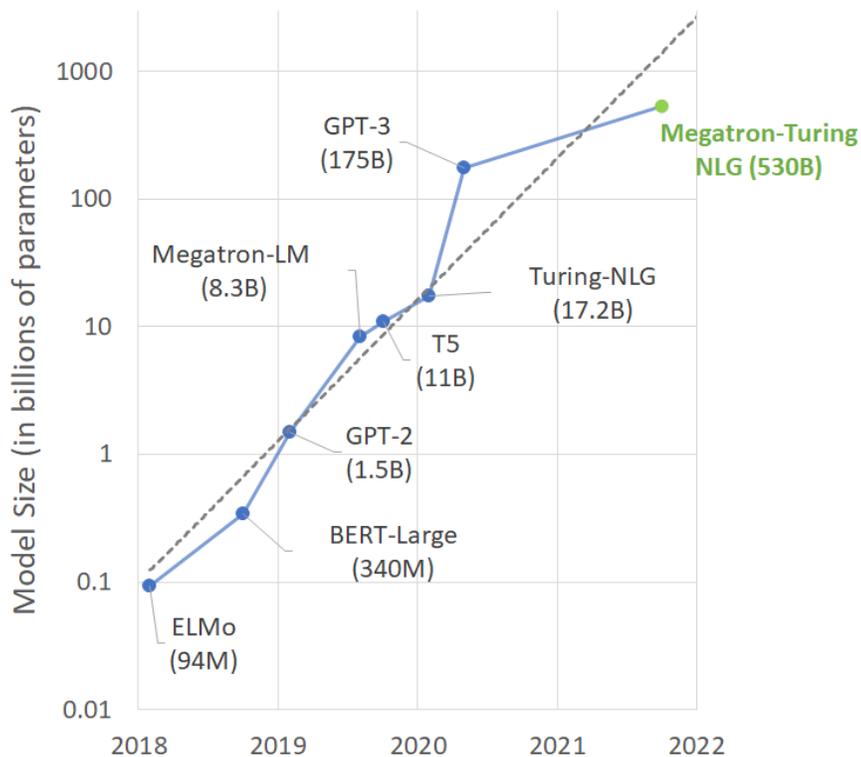
(c) Question Answering Tasks:  
SQuAD v1.1



(d) Single Sentence Tagging Tasks:  
CoNLL-2003 NER

# Трансформеры: размер имеет значение

Рост числа параметров больших языковых моделей



# Проблески общего искусственного интеллекта

## Sparks of Artificial General Intelligence: Early experiments with GPT-4

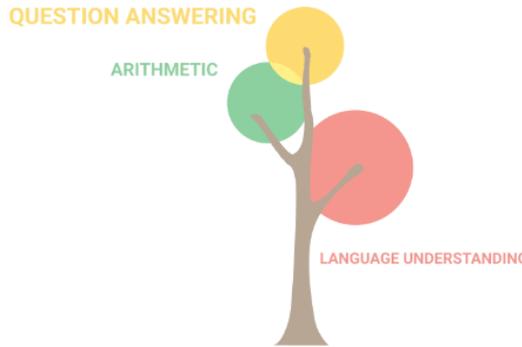
Sébastien Bubeck    Varun Chandrasekaran    Ronen Eldan    Johannes Gehrke  
Eric Horvitz    Ece Kamar    Peter Lee    Yin Tat Lee    Yuezhi Li    Scott Lundberg  
Harsha Nori    Hamid Palangi    Marco Tulio Ribeiro    Yi Zhang

Microsoft Research    (27 March 2023)

### **Новые способности модели GPT, не закладывавшиеся при её обучении:**

- объяснять свои ответы, перефразировать, переводить на другие языки
- реферировать, генерировать планы, сценарии, шаблоны
- строить аналогии, менять тональность, стиль, глубину изложения
- генерировать программный код на различных языках
- решать некоторые логические и математические задачи
- искать и исправлять собственные ошибки по подсказке

# Новые (эмерджентные) способности модели

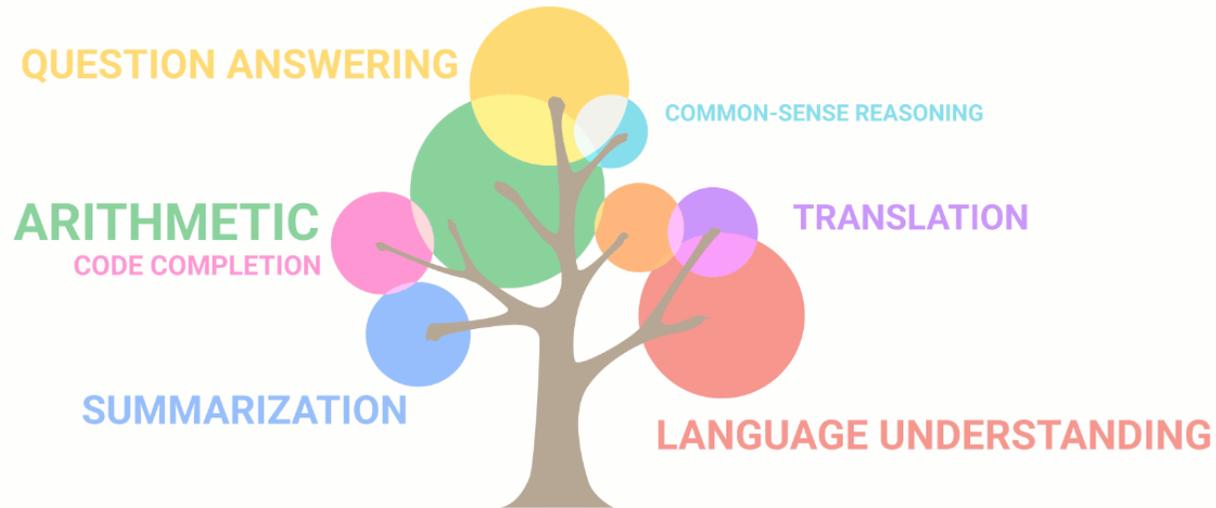


## GPT-2: 14-Feb-2019

1,5 млрд. параметров, корпус 10 млрд. токенов (40Gb), контекст 768 слов (1,5 стр.)

- способность написать эссе, которое конкурсное жюри не смогло отличить от написанного человеком

# Новые (эмерджентные) способности модели

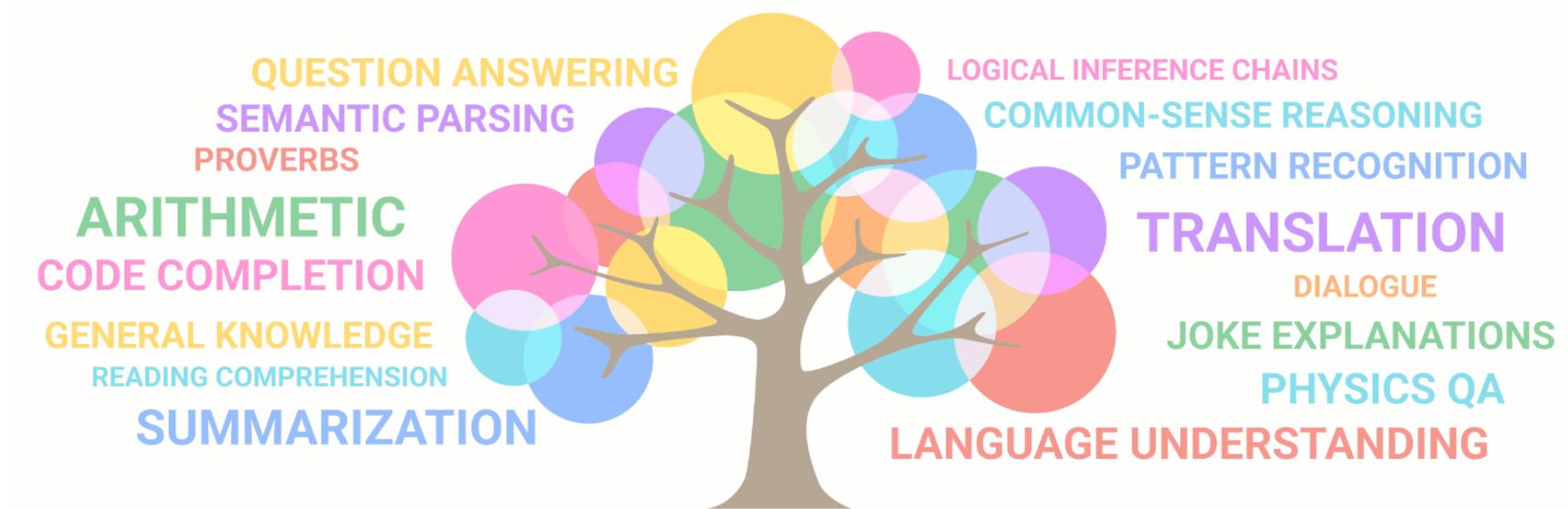


## GPT-3: 11-Jun-2020

175 млрд. параметров, корпус 500 млрд. токенов, контекст 1536 слов (3 стр.)

- способность делать перевод на другие языки
- способность решать логические и простейшие математические задачи
- способность генерировать программный код по текстовому описанию

# Новые (эмерджентные) способности модели



## GPT-4: 14-Mar-2023

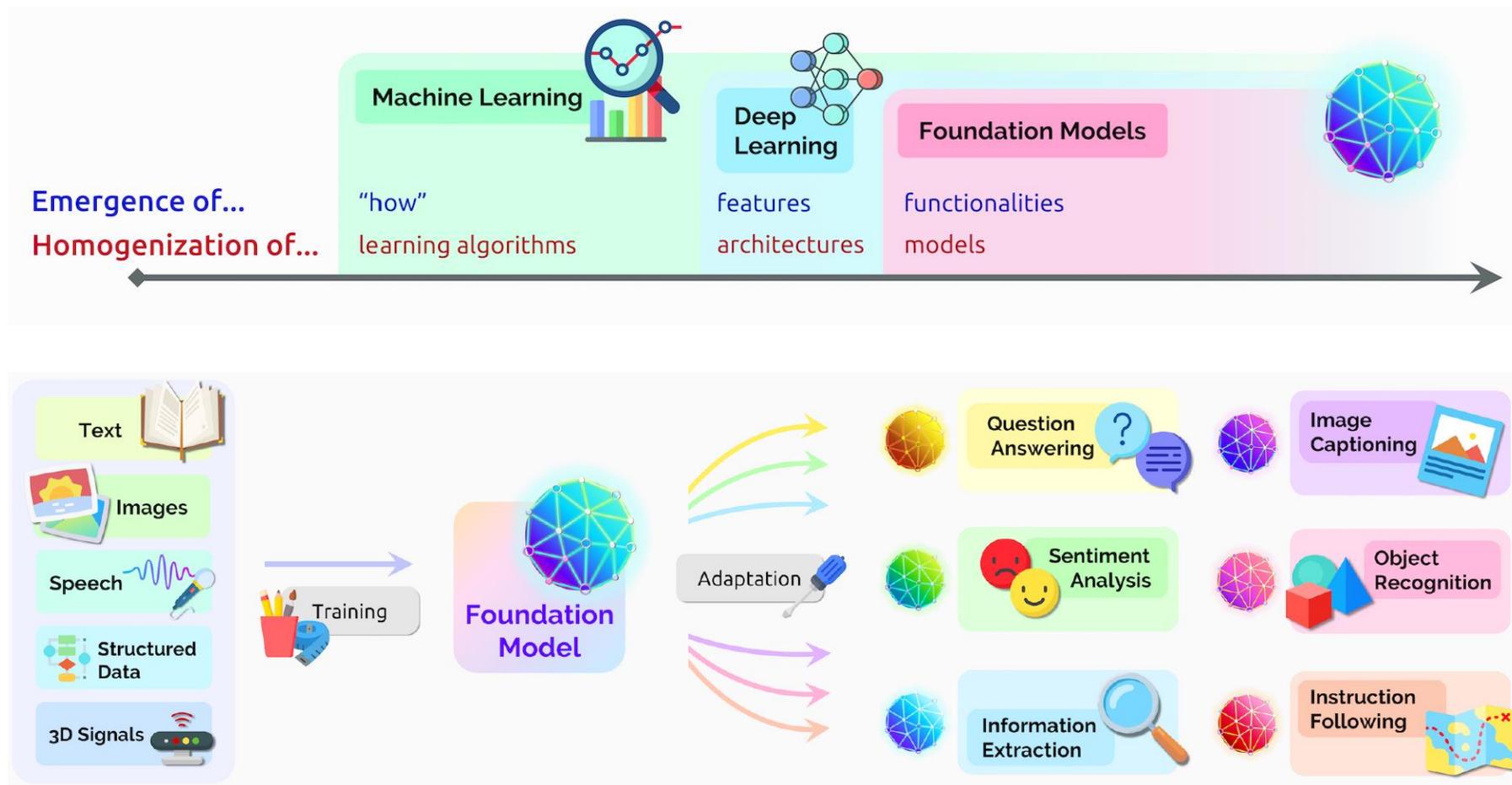
>1 трл. параметров, корпус >1Tb, контекст 24 000 слов (48 страниц)

- способность описывать и анализировать изображения
- способность реагировать на подсказки вроде «Let's think step by step»
- способность решать качественные физические задачи по картинке

# Возможности и угрозы: чаты GPT способны

- помогать с рутинной интеллектуальной работой
  - искать и структурировать профессиональную информацию
  - делать обзоры, рефераты, сводки на разных языках
  - генерировать тексты и сайты по тех. заданию, в том числе технические, медицинские, юридические
  - генерировать программный код по описанию
  - обсуждать новости, поддерживать разговор по теме
  - разговаривать с детьми с учётом возрастных особенностей
  - выполнять функции воспитателя, учителя, наставника
  - оказывать психологическую помощь
- «галлюцинировать», давать неверные сведения, касающиеся здоровья человека, законов, событий, технологий, других людей
  - вызывать необоснованное доверие и манипулировать человеком
  - переубеждать, побуждать человека к действиям, не выгодным ему
  - поддерживать предрассудки и лженаучные представления
  - поддерживать пропагандистские медиа-кампании
  - неконтролируемо влиять на формирование мировоззрения у подростков
  - оказывать депрессивное воздействие на психику

# Фундаментальные модели (Foundation Models)



# Содержание

## 1. Задачи машинного обучения

- Искусственный интеллект — сквозная технология
- Основы: история и терминология машинного обучения
- Примеры задач машинного обучения

## 2. Методология машинного обучения

- Нейронные сети и глубокое обучение
- Оптимизационные задачи машинного обучения
- Векторизация данных и большие языковые модели

## 3. Применение, перспективы, мифы

- Особенности практического применения технологий ИИ
- Перспективы развития ИИ
- Мифы об искусственном интеллекте

# Шаги практического решения задач AI/DS/ML

## Формализация постановки, «ДНК» задачи

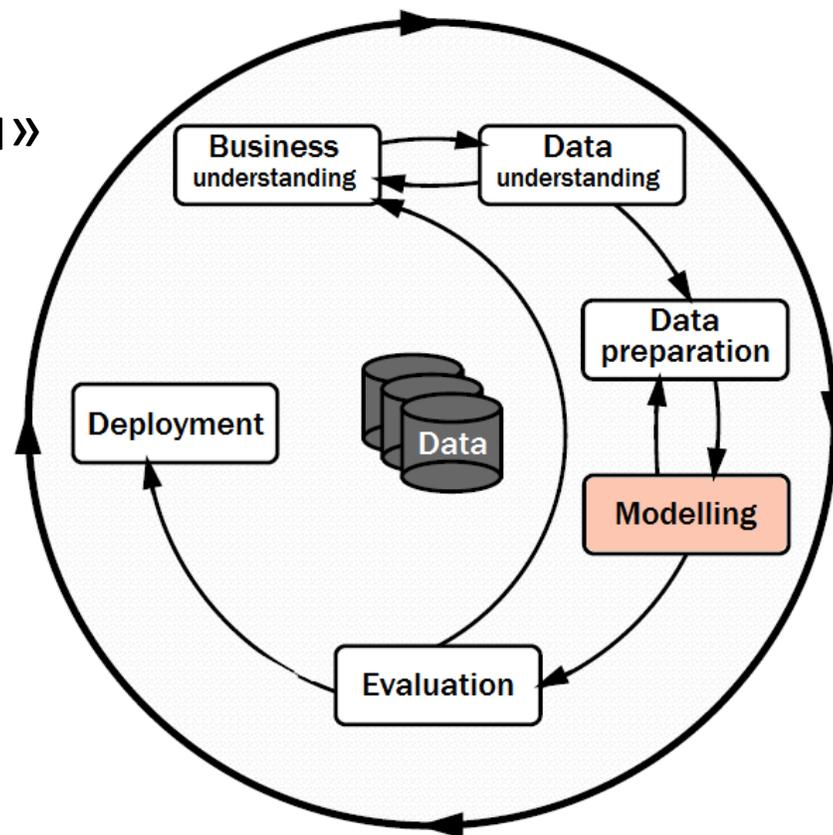
- **Дано:** выборка «объекты-признаки → ответы»
- **Найти:** предсказательная модель
- **Критерии:** качество предсказаний (KPI)

## Моделирование

- предобработка и векторизация данных
- формализация модели
- оптимизация (обучение) модели
- оценивание и выбор моделей

## Внедрение

- оценивание качества оффлайн и онлайн
- интеграция с бизнес-процессами



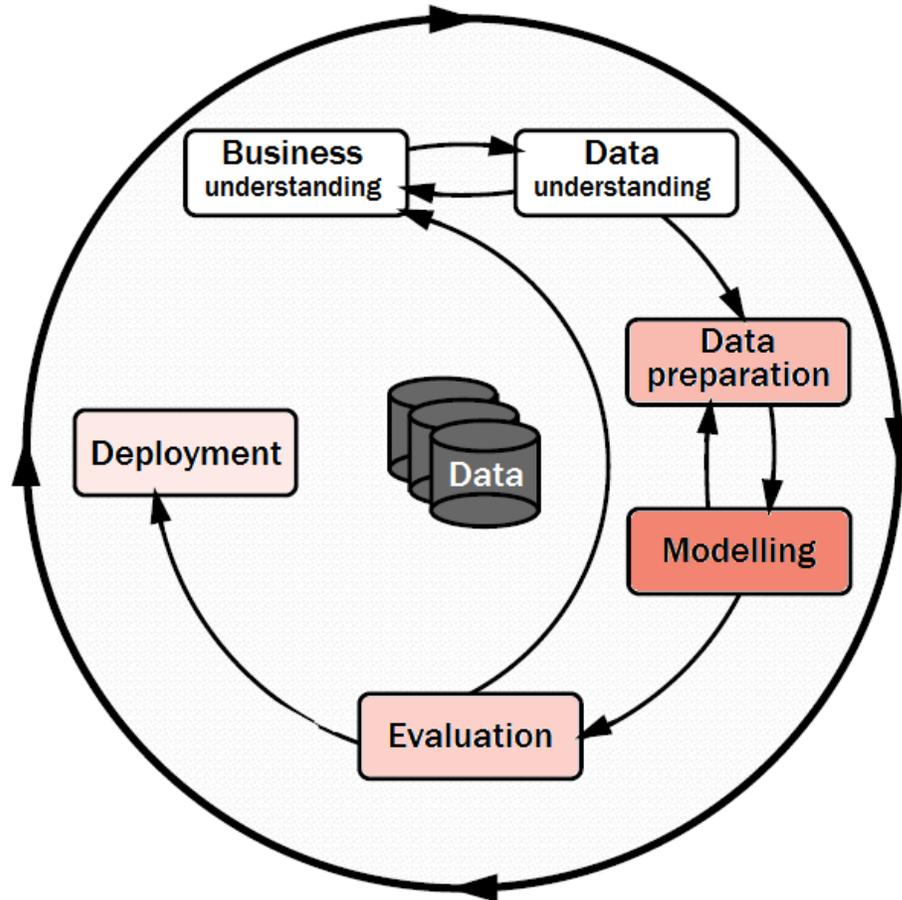
### CRISP-DM:

Cross Industry Standard Process  
for Data Mining (1999)

# Развитие AI: автоматизация шагов CRISP-DM

CRISP-DM: Cross Industry Standard Process for Data Mining (1999)

(SPSS, Teradata, Daimler AG, NCR Corp., OHRA)



- ручное создание экспертных систем  
→ **обучаемые модели (ML)**
- ручное конструирование признаков  
→ **обучаемая векторизация (DL)**
- ручное оценивание решений  
→ **автоподбор моделей (AutoML)**
- ручное внедрение и эксплуатация  
→ **бесшовная эволюция моделей (RL)**
- понимание прикладной задачи и понимание данных  
→ **теперь всё чаще ChatGPT**

# Особенности реальных данных

## В реальных приложениях данные бывают ...

- разнородные (признаки измерены в разных шкалах)
- неполные (признаки измерены не все, имеются пропуски)
- неточные (признаки измерены с погрешностями)
- противоречивые (объекты одинаковые, ответы разные)
- избыточные (сверхбольшие, не помещаются в память)
- недостаточные (объектов меньше, чем признаков)
- неструктурированные (нет признаков описаний)
- «грязные» (ошибочные, грубо не соответствующие истине)

*со всем этим  
можно  
работать*



*но только не  
с грязными  
данными!*



# Необходимые условия применения ИИ

- **Полнота, чистота, достоверность данных**
  - Автоматизация и цифровизация бизнес-процессов
  - Улучшение качества данных (от «цифрового чучела» к цифровому двойнику)
  - Трудовая и технологическая дисциплина при работе с данными
- **Культура постановки задач**
  - Понимание бизнес-целей и их формализация через измеримые критерии
  - Предметная экспертиза вместо «абстрактной веры во всемогущий ИИ»
  - Готовность пилотировать новые технологии («data-driven» на всех уровнях)
- **Культура анализа данных**
  - Владение средствами визуализации и понимания данных
  - Тщательный анализ ошибок при выборе моделей
  - Умение находить «простые но гениальные» решения

# Мифы №1 ... №8

## **«В будущем Искусственный Интеллект...**

... лишит людей работы»

... будет использован для узурпации власти над миром»

... приведёт людей к праздности и деградациии»

... станет настолько мощным, что мы перестанем понимать его цели»

... станет автономным и реплицируемым, выйдет из-под контроля»,

... уничтожит человеческую цивилизацию»,

... и всю биологическую жизнь на Земле»

... продолжит вместо нас эволюцию разума на Земле и в космосе»

# Миф №9

## «Большие языковые модели — это новый вид интеллекта»

- нет, лишь новый языковой интерфейс к содержимому Интернета
- постоянно улучшаемый и совершенствуемый,
- с которым нам придётся работать и к которому привыкать,
- постепенно избавляясь от иллюзий и когнитивных искажений,
- имея в виду, что это всего лишь технология
  - предсказания одного слова по очень длинному контексту,
  - оптимизации моделей очень больших размерностей

# Миф №10

## «Скоро будет создан Общий Искусственный Интеллект (AGI)»

- нет определения, что именно хочется создать, и с какой целью

### **Отличия ИИ от естественного биологического интеллекта:**

- наш интеллект эволюционировал как инструмент выживания
- мы обучаемся не по выборкам, а на основе объяснений учителей, воспитания, опыта, коммуникации, изучения естественной среды
- мы имеем картину мира, целеполагание, строим цивилизацию
- у нас 86 млрд. нейронов, и они устроены намного сложнее

# Главное о технологиях ИИ

- принцип эмпирической индукции Фрэнсиса Бэкона
- принцип минимизации (и аппроксимации) эмпирического риска
- принцип коннекционизма и глубокие нейросетевые архитектуры
- принцип векторизации сложно структурированных данных
- большие данные + большие модели + быстрые вычисления
- самостоятельное обучение вместо обучения по разметке
- обеспечение чистоты и полноты данных

# Главное об отношении людей к ИИ

- ИИ = Имитация Интеллекта, набор технологий, не объект, не субъект
- ИИ начинается с постановки задачи *Дано-Найти-Критерий*
- *Люди* ставят задачу и несут ответственность за её решение, *за чистоту и полноту данных* — тоже
- *Глубокие нейронные сети* — не аналог мозга человека, а обучаемая векторизация сложно структурированных данных
- *Генеративные модели* — не интеллект, а новый языковой интерфейс к ~~знанию человечества~~ содержимому Интернета, со всей его избыточностью, неточностью, противоречивостью

# Антропоцентричное определение ИИ

*Искусственный интеллект* — это вычислительные технологии, создаваемые для повышения эффективности интеллектуального труда людей



# Рекомендуемые материалы

- *Визильтер Ю. В.* От слабого ИИ к общему универсальному интеллекту (обзор тенденций 2020-2023). Семинар РАИИ и ФИЦ ИУ РАН «Проблемы искусственного интеллекта» 31.01.2024  
<https://rutube.ru/video/2aad53ec833f19918c1593398a2a1b88/>
- Не пропустите открытие тысячелетия! // Vital Math, 13 января 2024,  
<https://www.youtube.com/watch?v=JZjH0it9Jyg>
- Report: AI Decrypted: A Guide for Navigating AI Developments in 2024, January 24, 2024 (Навигатор по ИИ-ландшафту от DENTONS GLOBAL ADVISORS)  
<https://www.albrightstonebridge.com/news/report-ai-decrypted-guide-navigating-ai-developments-2024>
- 5 идей применения ИИ в вашем бизнесе прямо сейчас, 5 октября 2023.  
<https://dzen.ru/a/ZR6ZeK5B3IL6OxXv>
- *Воронцов К. В.* Лекции по машинному обучению. [www.MachineLearning.ru](http://www.MachineLearning.ru), 2004-2023.
- *Гарбук С.В., Губинский А.М.* Искусственный интеллект в ведущих странах мира: стратегии развития и военное применение. Знание, 2020.
- *Шумский С. А.* Машинный интеллект. РИОР ИНФРА-М, 2020.