



МИНИСТЕРСТВО НАУКИ
И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ



Московский
государственный
университет
имени М. В. Ломоносова



МОСКОВСКОЕ
УЧЕБНО-НАУЧНОЕ
ЦЕНТРОЕ



Центр
педагогического
мастерства



22-й
международный
фестиваль
науки и технологий

МЕЖДУНАРОДНЫЙ ФЕСТИВАЛЬ НАУКА +

КРАТКАЯ ИСТОРИЯ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

- Воронцов Константин Вячеславович
- д.ф.-м.н., профессор РАН
(ВМК МГУ, Институт ИИ МГУ)



«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,

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»

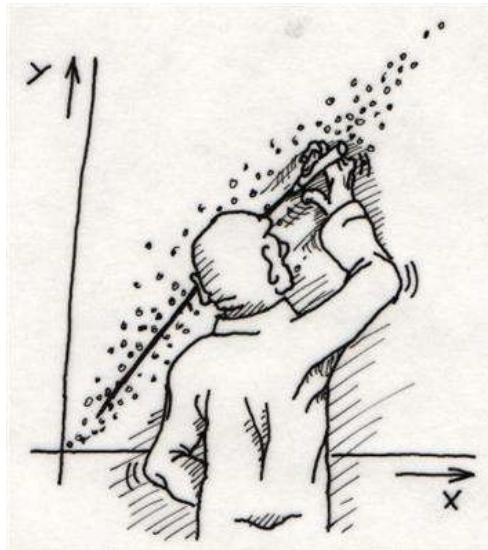
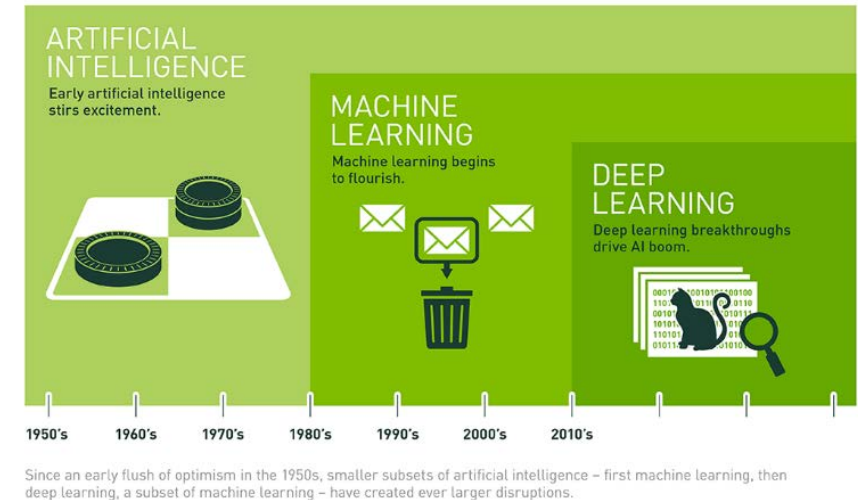


Машинное обучение, большие данные «и много других страшных слов»

- Статистический анализ данных (Statistical Data Analysis)
- Искусственный интеллект (Artificial Intelligence, AI) 1956
- 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

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

Методология эмпирической индукции

От дедуктивного метода познания к индуктивному:

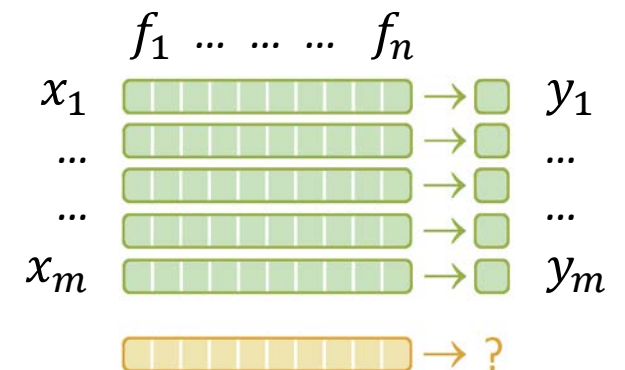
«Не следует полагаться на сформулированные аксиомы и формальные базовые понятия, какими бы привлекательными и справедливыми они не казались. **Законы природы нужно «расшифровывать» из фактов опыта.** Следует искать правильный метод анализа и обобщения опытных данных; здесь логика Аристотеля не подходит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



Задачи машинного обучения с учителем

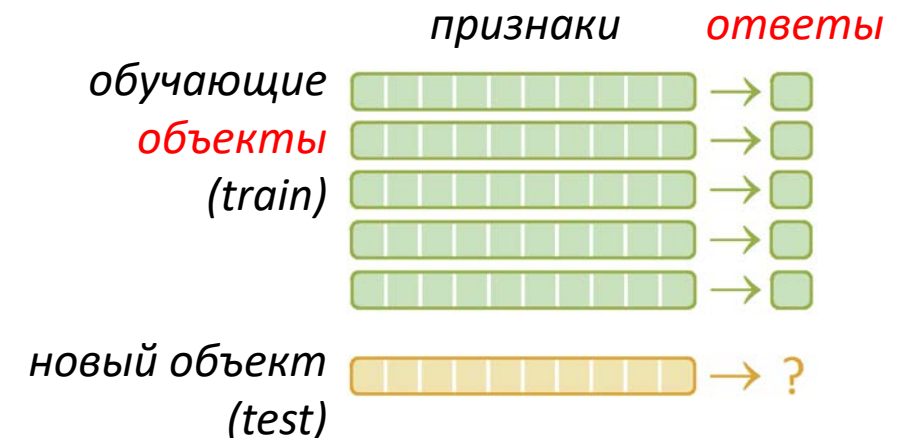
Этап №1 – обучение с учителем

- **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 (ML)

- **Медицинская диагностика:**

объект – данные о пациенте на текущий момент

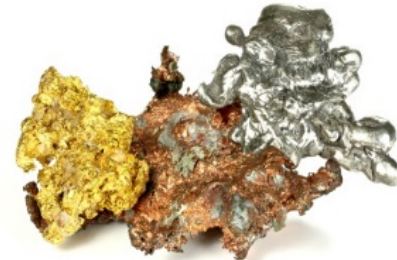
ответ – диагноз / лечение / риск исхода



- **Поиск месторождений полезных ископаемых:**

объект – данные о геологии района

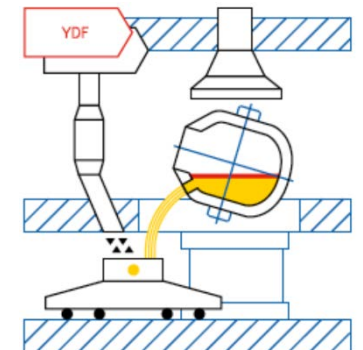
ответ – есть/нет месторождение



- **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 данными сложной структуры

- **Статистический машинный перевод:**

объект – предложение на естественном языке

ответ – его перевод на другой язык

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

- **Перевод речи в текст:**

объект – аудиозапись речи человека

ответ – текстовая запись речи

...очень важное уточнение:

***с аккуратными**
большими данными*

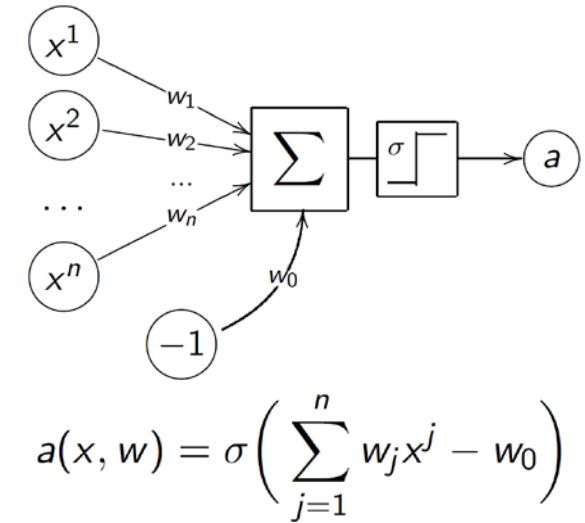
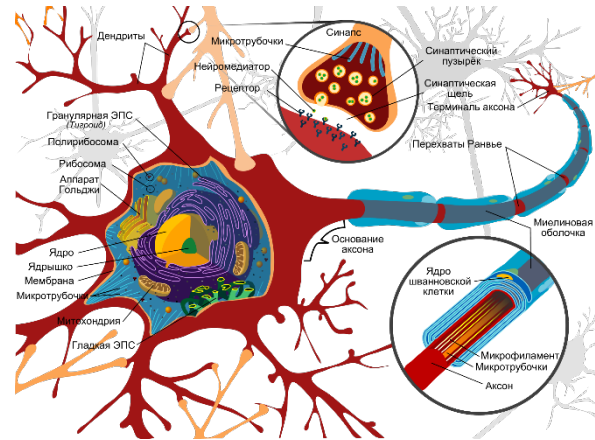
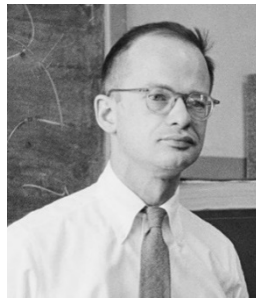
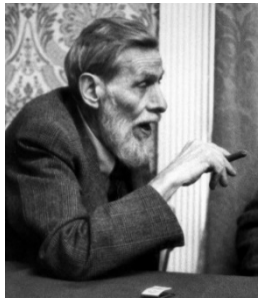
- **Беспилотное вождение:**

объект – поток данных с радаров и видеокамер

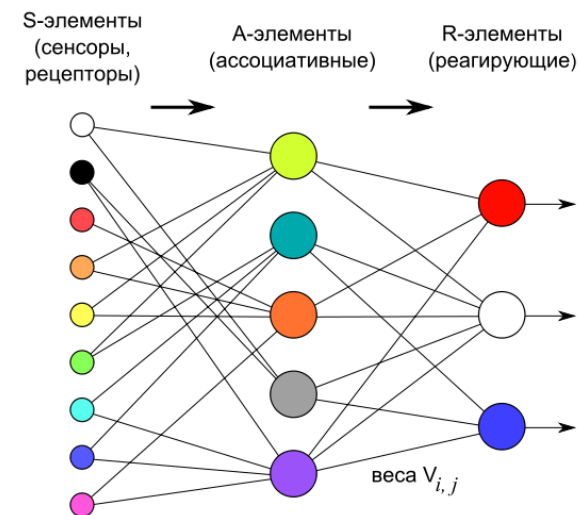
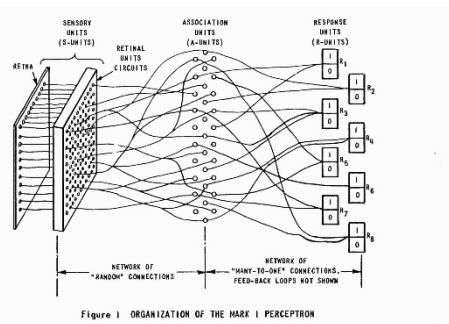
ответ – поток решений (объехать, остановиться, игнорировать, ...)

Искусственные нейронные сети (ИНС, ANN)

Математическая модель нейрона (МакКаллок и Питтс, 1943)



Первый нейрокомпьютер Mark-1 (Фрэнк Розенблатт, 1960)

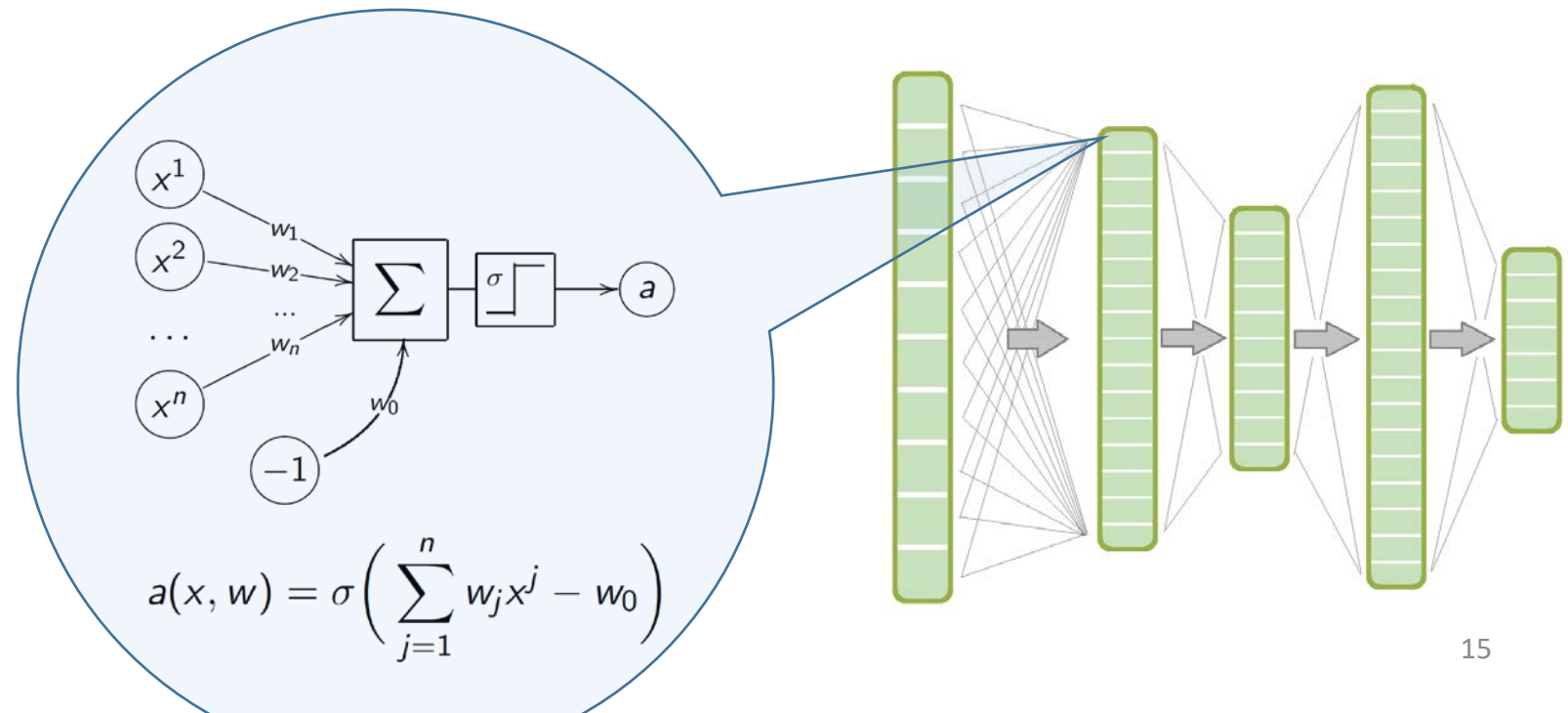
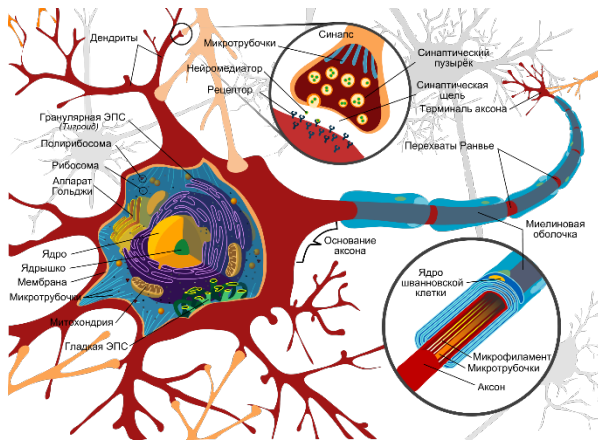


Модели искусственных нейронных сетей

На каждом слое сети вектор объекта преобразуется в новый вектор

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

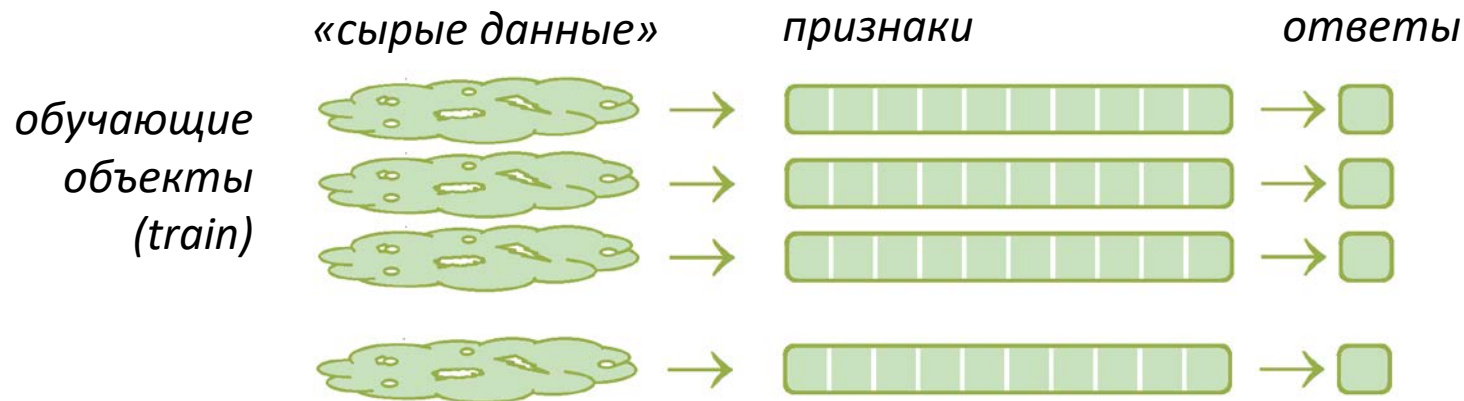
Веса w являются обучаемыми параметрами модели



Глубокие нейронные сети (Deep ANN)

Вход: сложно структурированные «сырые» данные объектов

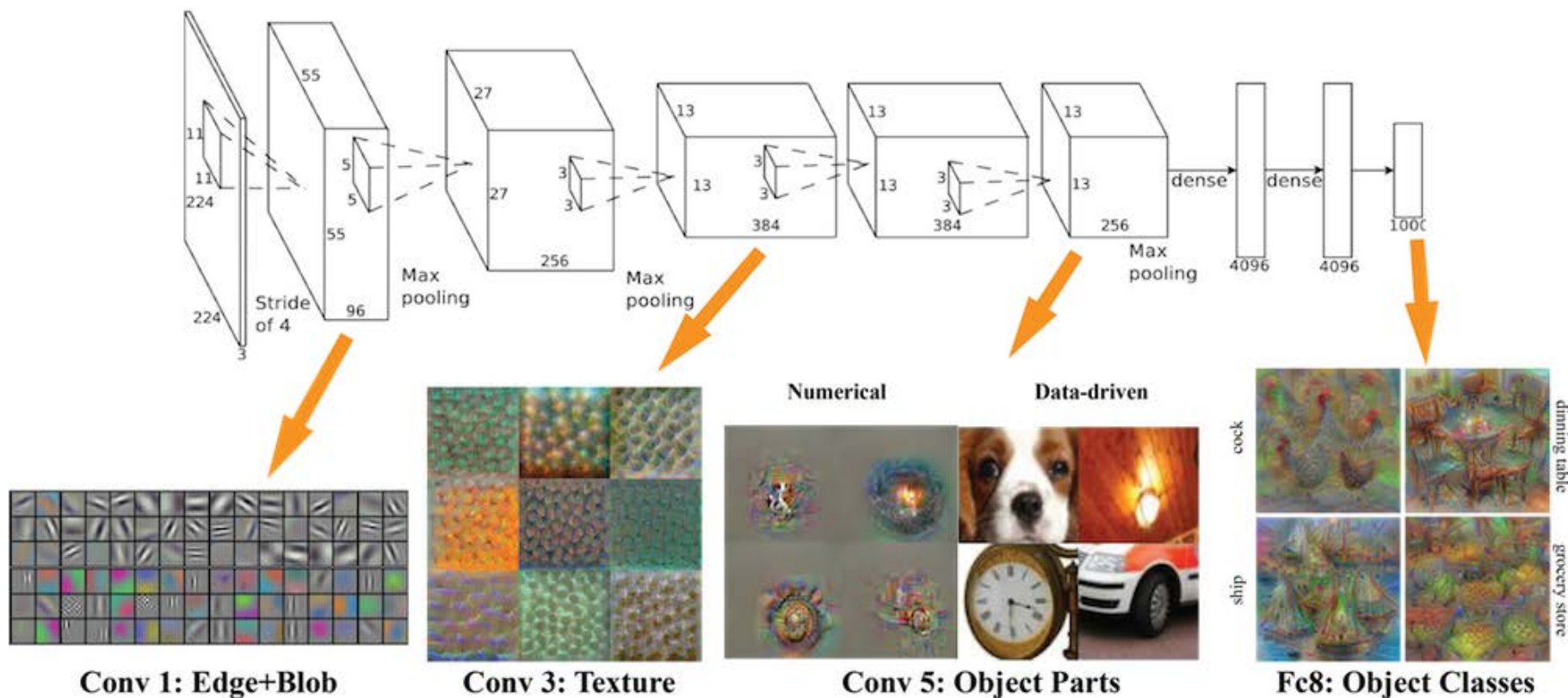
Выход: векторные представления объектов, затем ответы



*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

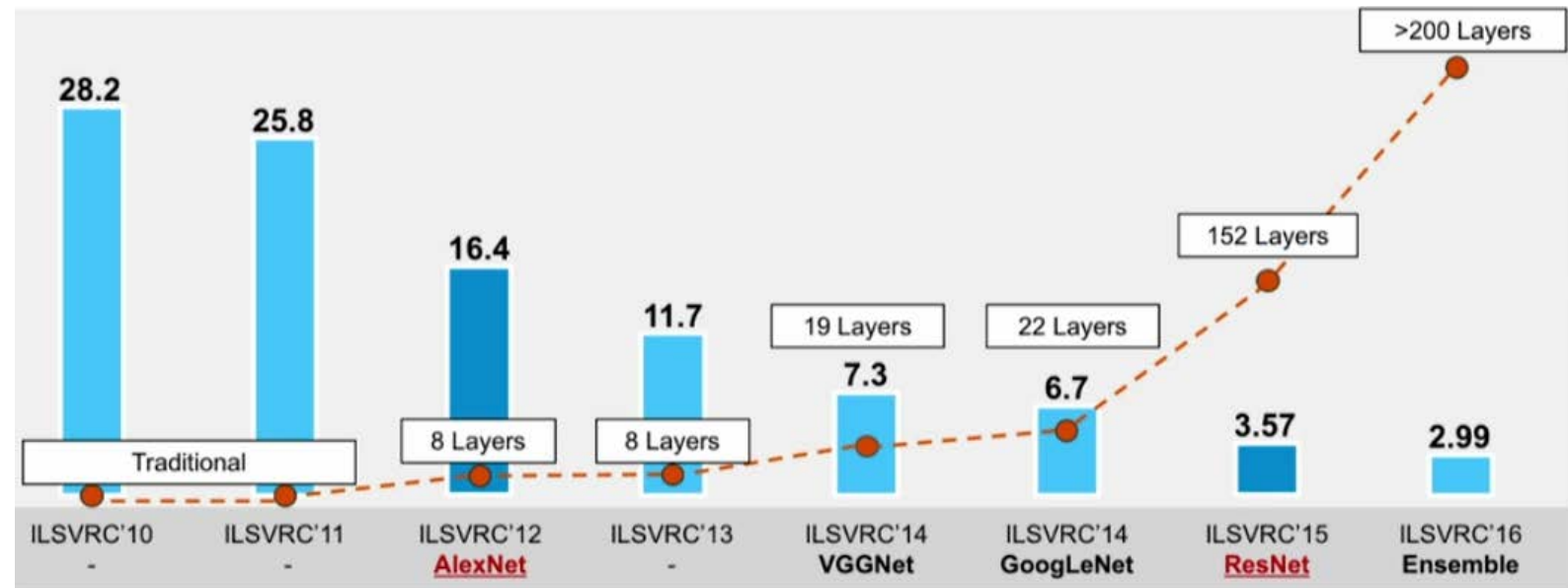


Krizhevsky A., Sutskever I., Hinton G. ImageNet classification with deep convolutional neural networks. 2012.

Роль больших данных

ImageNet: открытая выборка 14М изображений, 20К категорий

IMAGENET



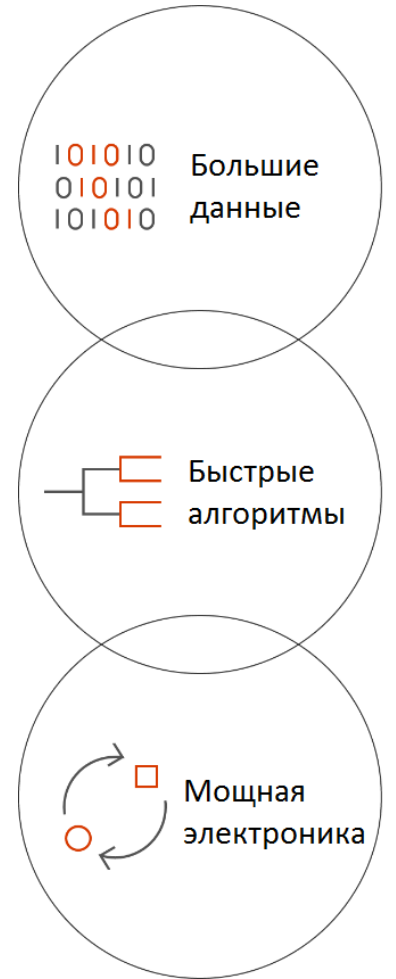
Старт в 2009 г. Человеческий уровень ошибок 5% пройден в 2015 г.

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

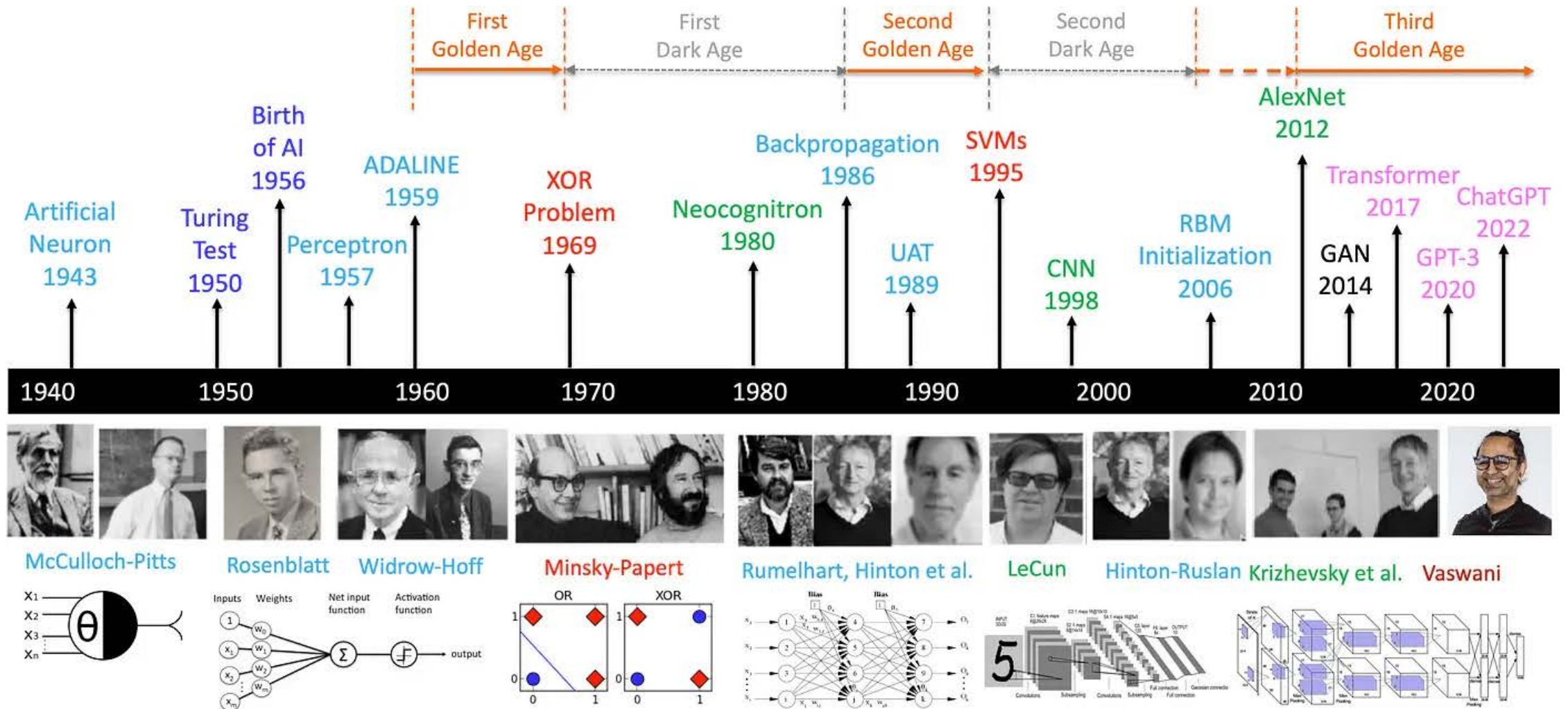
Li Fei-Fei et al. Construction and analysis of a large scale image ontology. 2009.

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 для больших размерностей
- Достижения микроэлектроники
→ *рост вычислительных мощностей по закону Мура*
в частности, GPU



Зимы искусственных нейросетей (AI winters)



Машинное обучение – это оптимизация

Дано: $\{x_1, \dots, x_m\}$ – обучающая выборка объектов

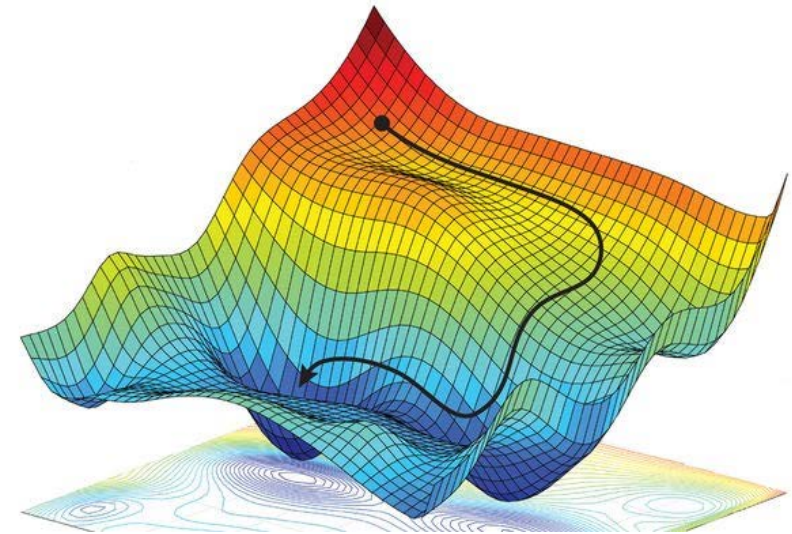
Найти: совокупность параметров w
предсказательной модели $a(x, w)$

Критерий: минимум суммы потерь

$$Q(w) = \sum_{i=1}^m \text{Loss}(x_i, w) \rightarrow \min$$

где $\text{Loss}(x_i, w)$ – потеря от предсказания $a(x_i, w)$ на объекте x_i

Способ решения – численные методы оптимизации



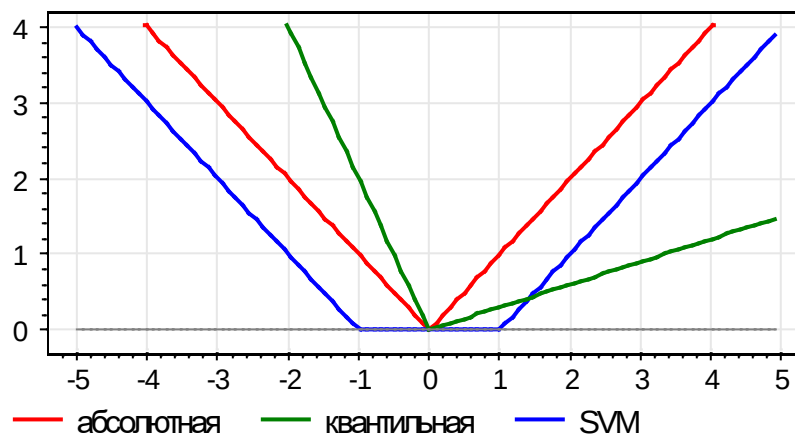
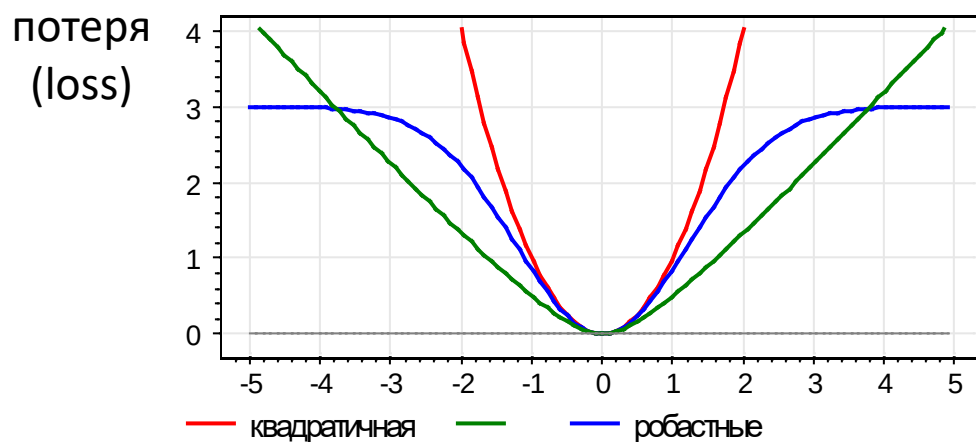
Обучение с учителем (supervised learning): восстановление регрессии (regression)

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



ошибка
(error)

Задача проведения функции через точки

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

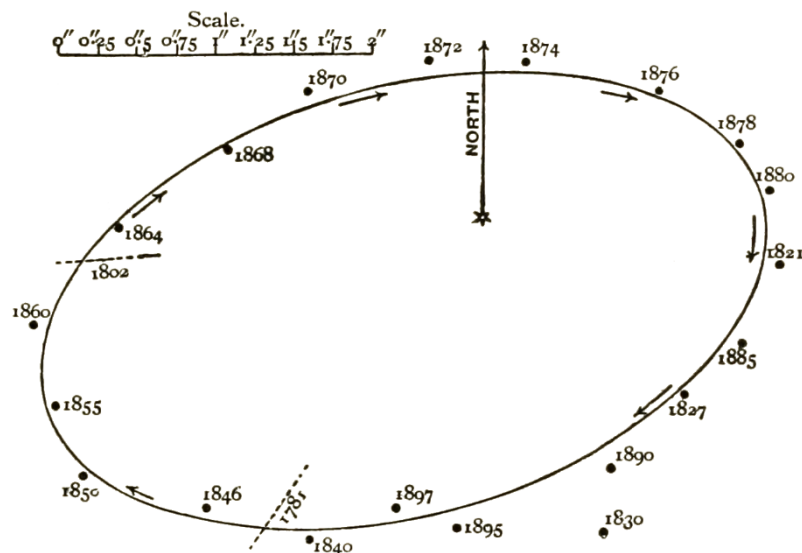
$$a(x, w) = \sum_j w_j f_j(x)$$

Метод наименьших квадратов (Гаусс, 1795):

$$\sum_i (a(x_i, w) - y_i)^2 \rightarrow \min_w$$



Карл Фридрих Гаусс
(1777--1855)



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

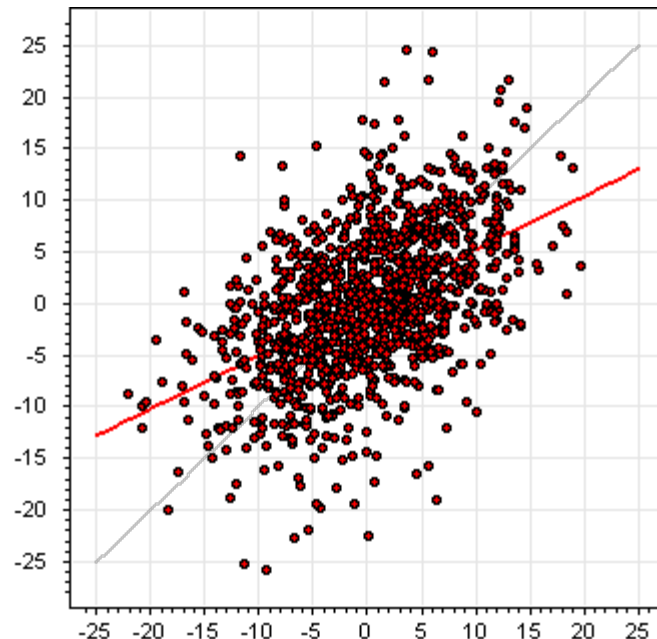
C.F.Gauss. Theory of the motion of the heavenly bodies moving about the Sun in conic sections. 1809.

Задача восстановления регрессии

Исследование наследственности роста (Гальтон, 1886).

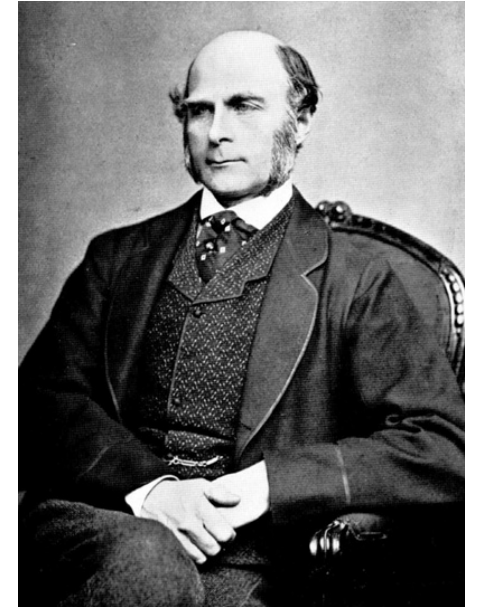
Δ — отклонение роста от среднего в популяции

Зависимость (линейная?) Δ взрослого сына от Δ отца:



Двойной смысл термина «регрессия»:

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



Фрэнсис Гальтон
(1822--1911)

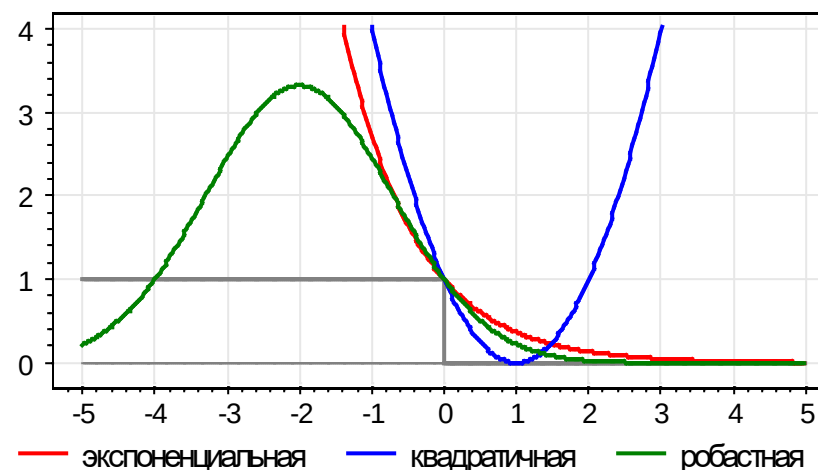
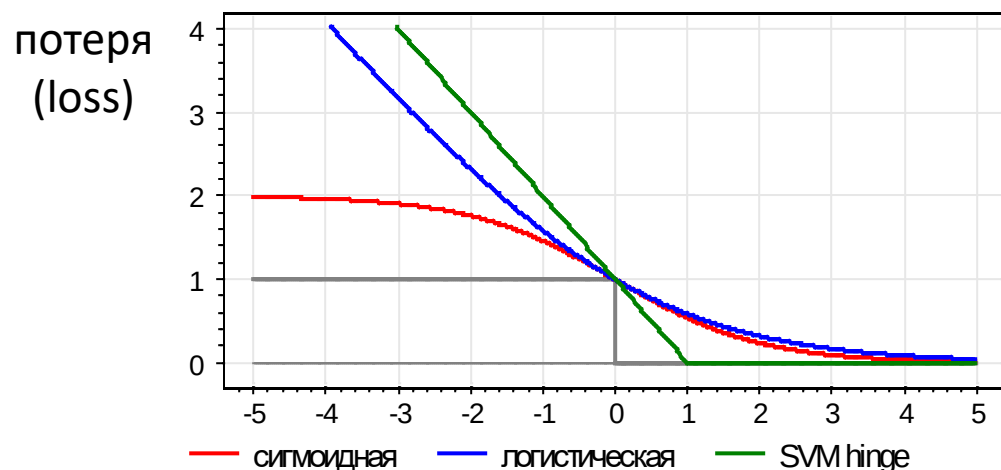
Обучение с учителем (supervised learning): классификация (classification)

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)

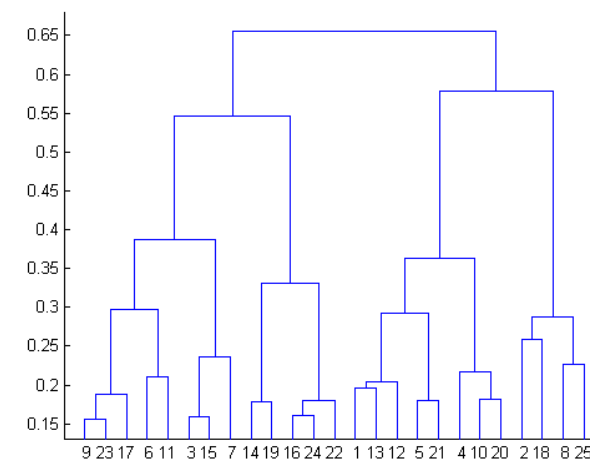
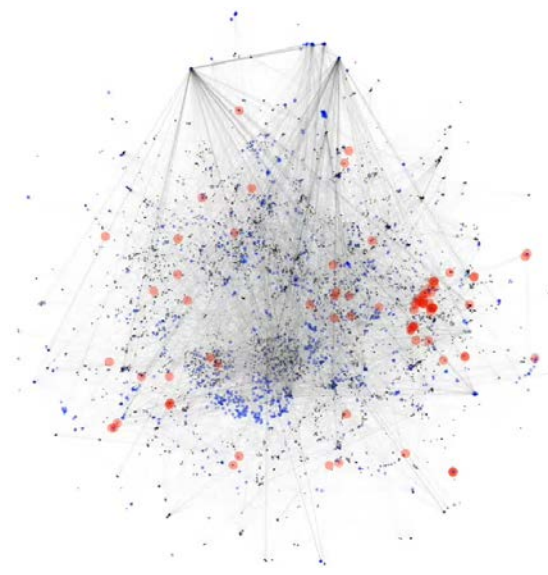
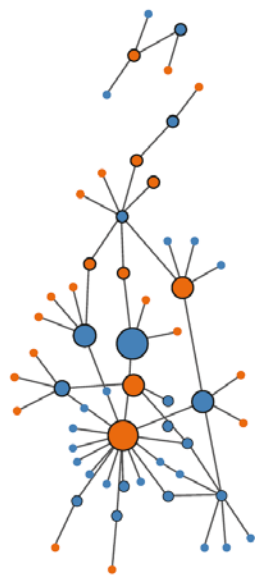
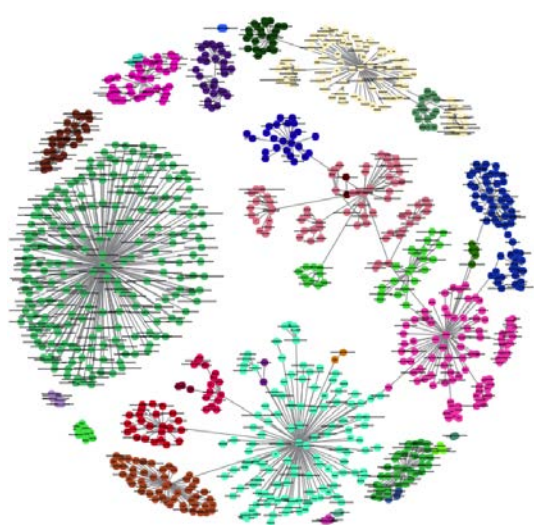
Обучение без учителя (unsupervised learning): кластеризация (clustering)

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



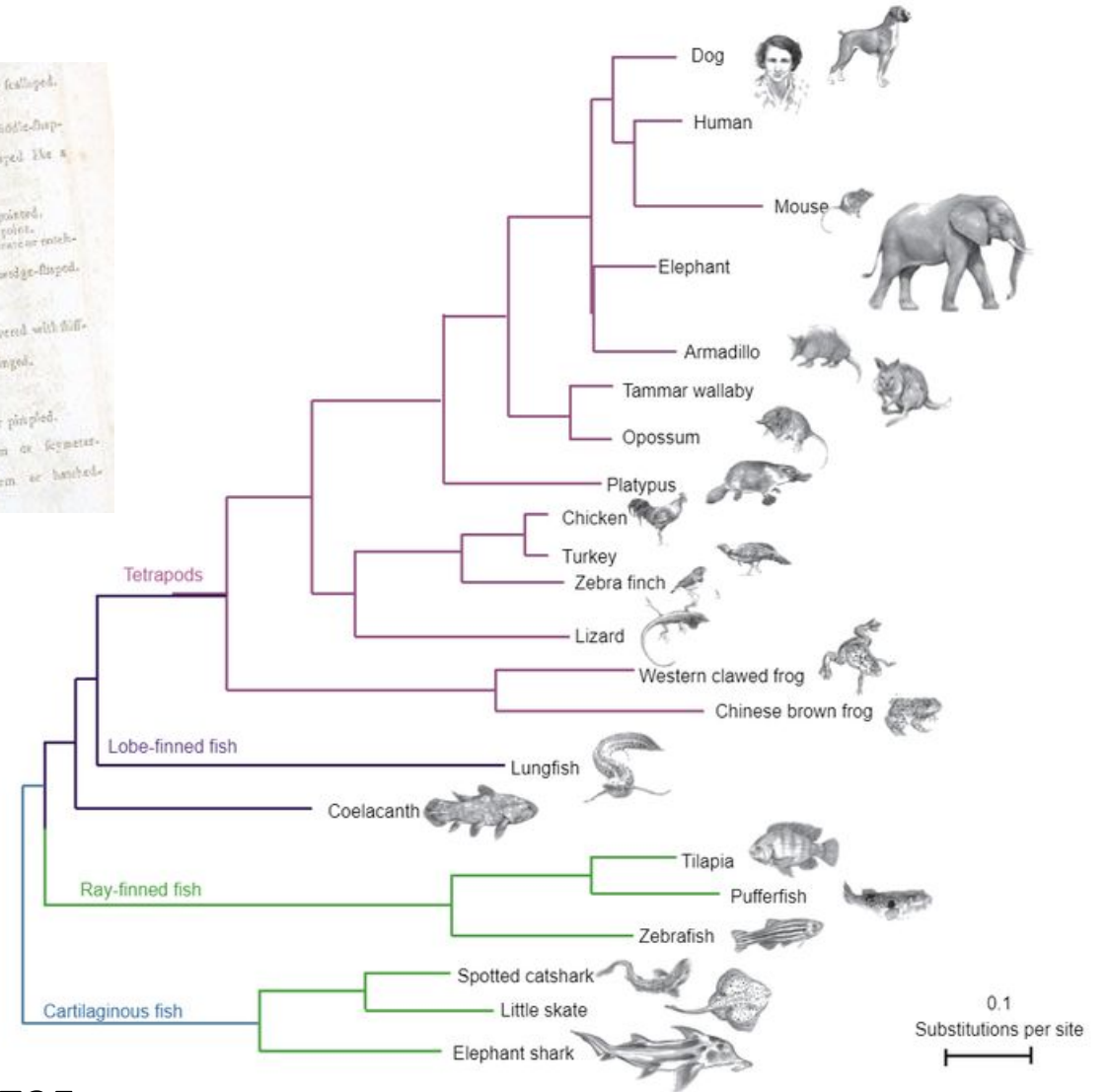
Задача иерархической кластеризации



Карл Линней
(1707--1778)

Система классификации животных, растений и минералов на основе сходства признаков (4 уровня: класс, отряд, род, вид)

Carl Linnaeus. Systema naturae (Система природы). 1735.



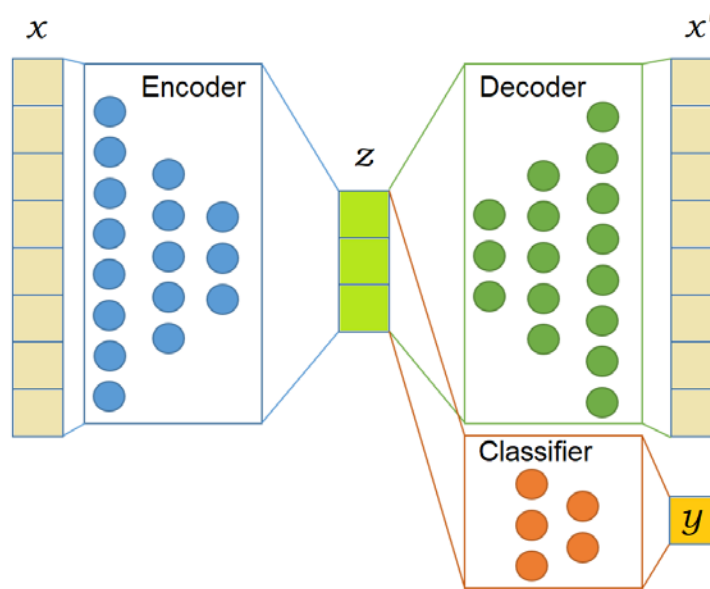
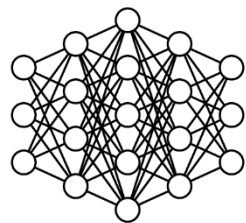
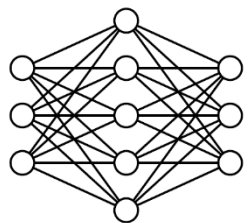
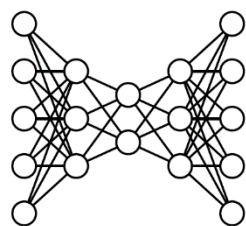
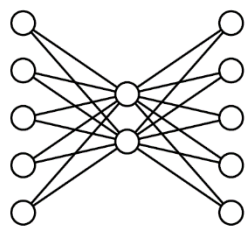
Обучение без учителя (unsupervised learning), векторизация: автокодировщик (autoencoder)

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

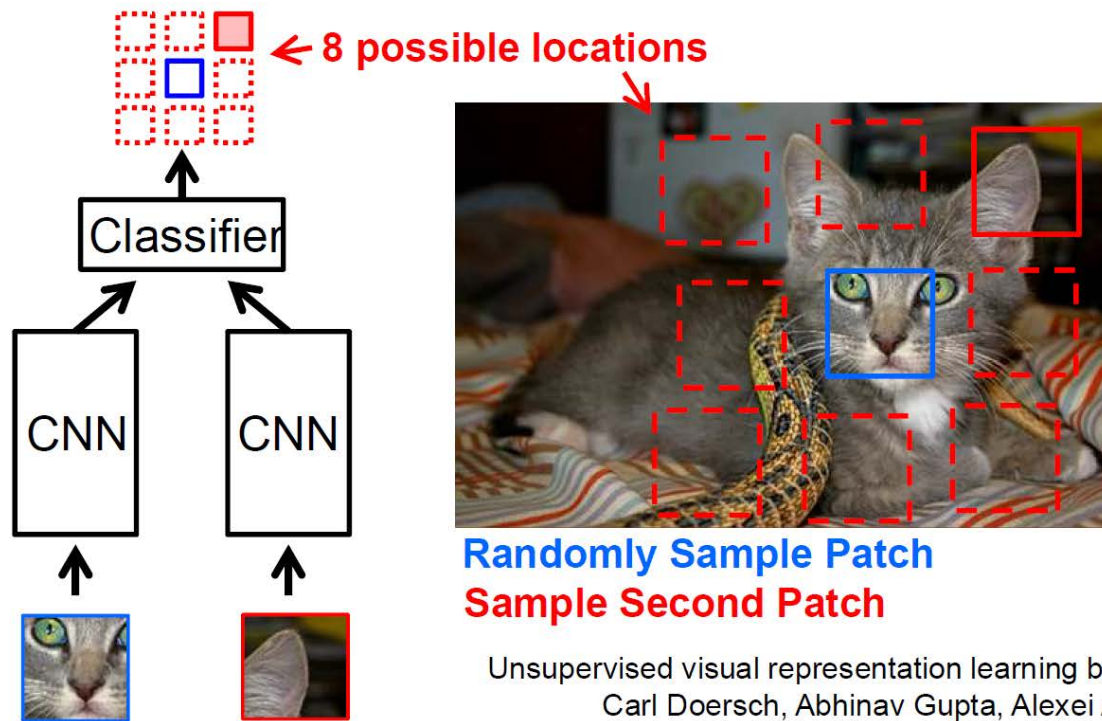


обучаемая
векторизация
сложных
объектов

Самостоятельное обучение (self-supervised)

x — изображение

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



Преимущество:

сеть выучивает векторные представления объектов без размеченной обучающей выборки

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

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



Многозадачное обучение (multi-task learning)

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

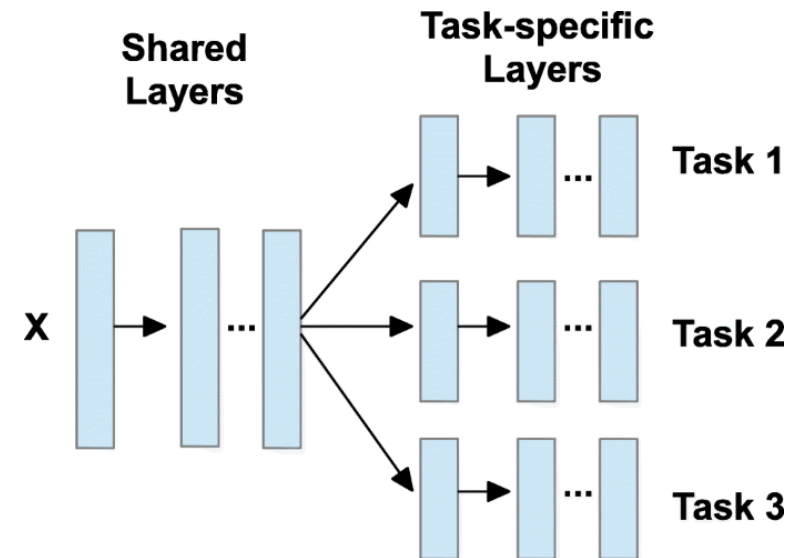
$y = g_t(z, w'_t)$ – часть модели, специфичная для t -й задачи

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

few-shot learning – обучение по малому числу примеров

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

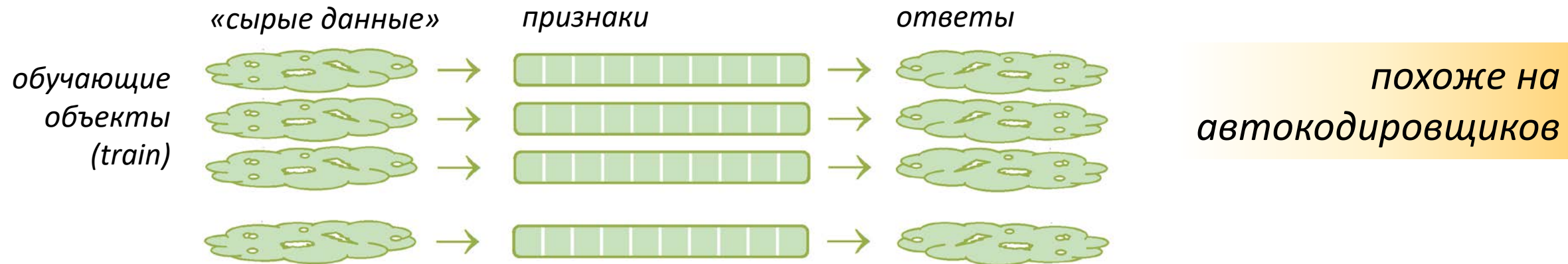
Y.Wang et al. Generalizing from a few examples: a survey on few-shot learning. 2020



Нейронные сети для синтеза объектов

Вход: сложно структурированные объекты

Выход: сложно структурированные ответы



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 «поддельного» объекта x из шума z

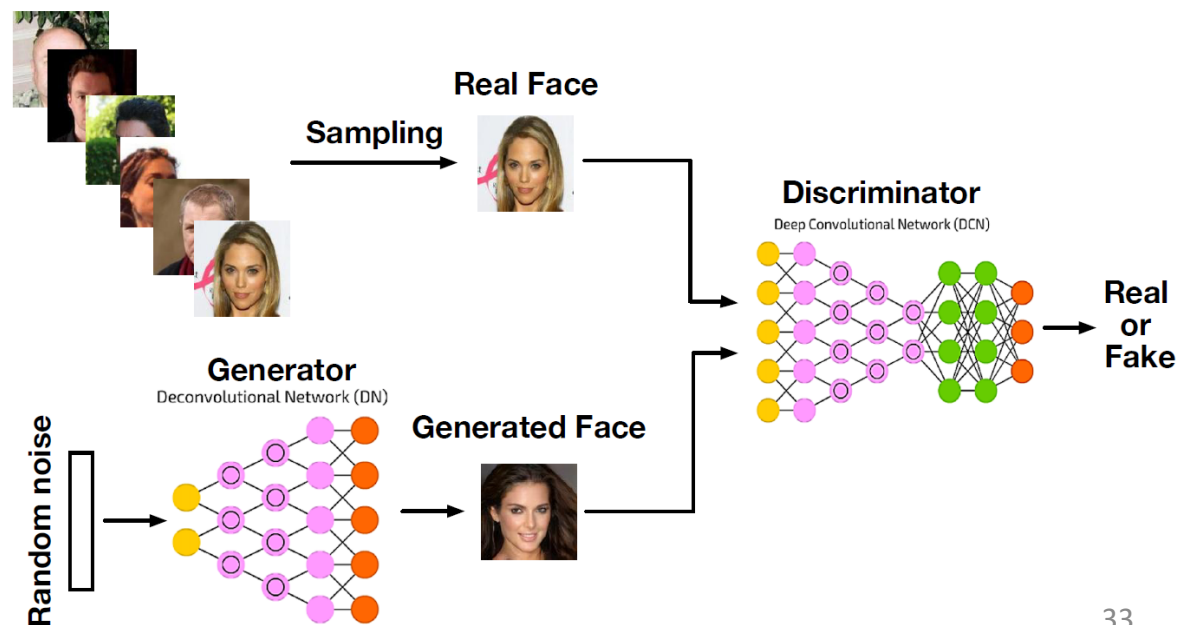
$f(x, w')$ — модель классификации x «реальный/поддельный»

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

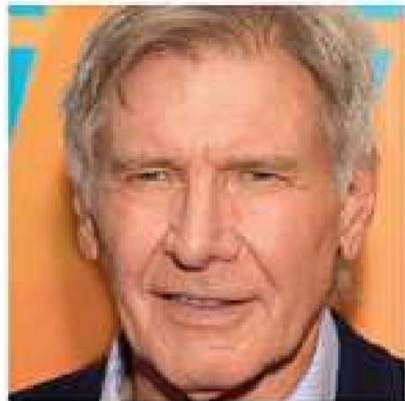
Antonia Creswell et al. Generative Adversarial Networks: an overview. 2017.

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

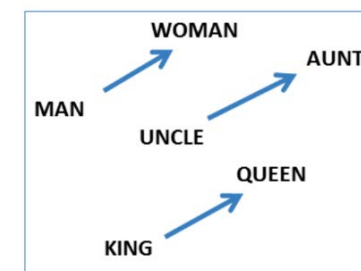
Декомпозиция задач по уровням «пирамиды NLP»

- 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], ...

$$\text{softmax} \left(\frac{\begin{matrix} \text{Q} \\ \text{3x3 grid} \end{matrix} \times \begin{matrix} \text{K}^T \\ \text{3x3 grid} \end{matrix}}{\sqrt{d}} \right) \begin{matrix} \text{V} \\ \text{3x1 grid} \end{matrix}$$

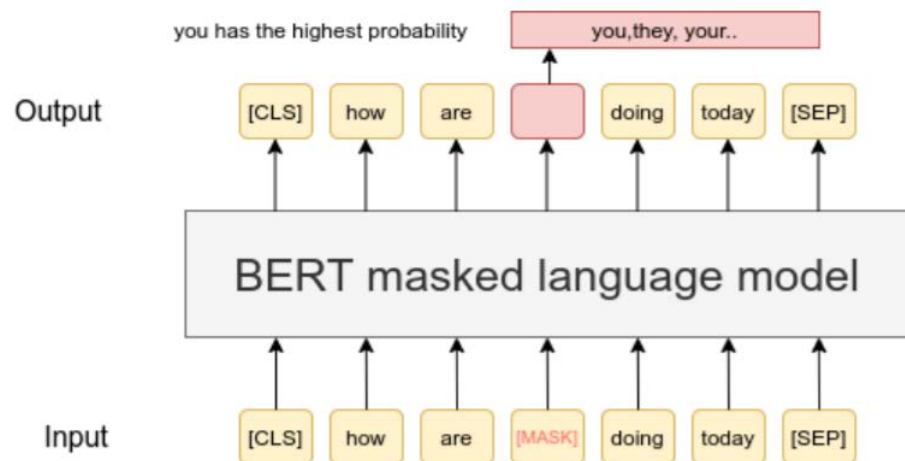
Обучение контекстной векторизации слов

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

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

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

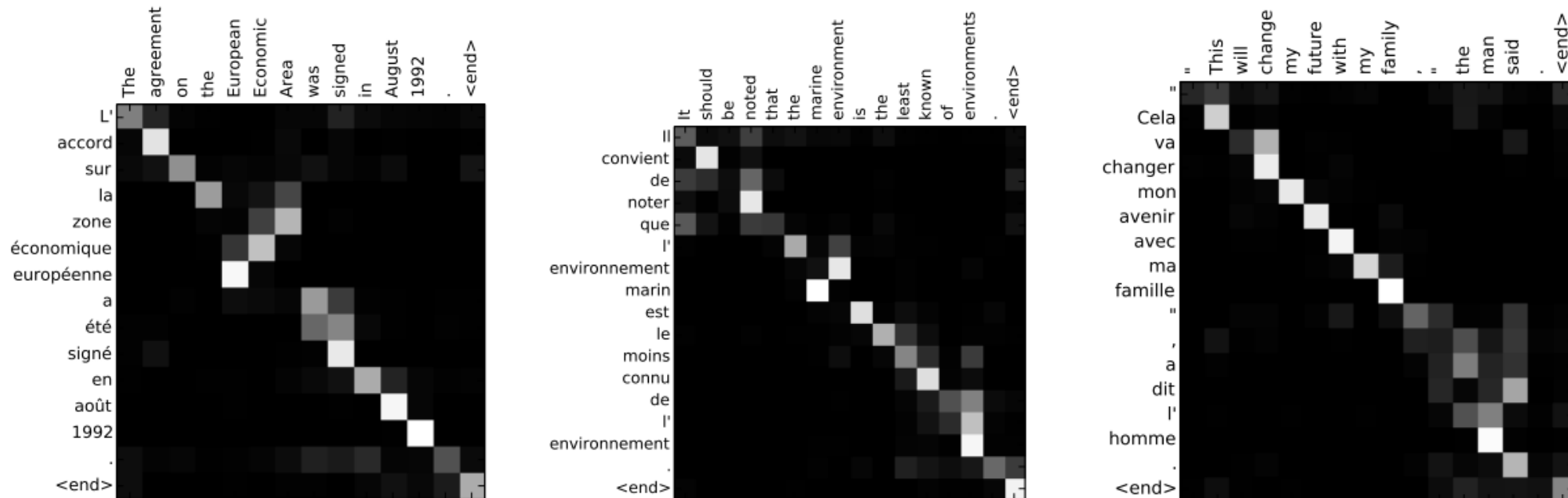


Vaswani et al. (Google) Attention is all you need. 2017.

Jacob Devlin et al. (Google AI Language)

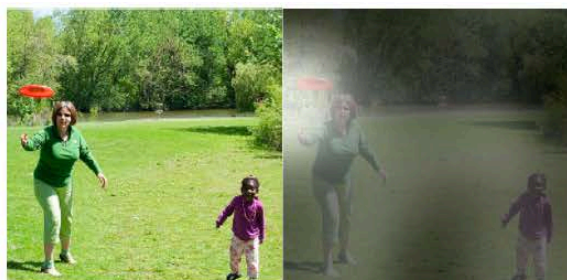
BERT: pre-training of deep bidirectional transformers for language understanding. 2019.

Модели внимания: машинный перевод

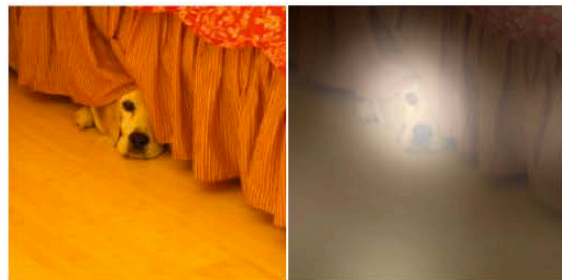


Интерпретация моделей внимания: *матрица семантического сходства* $A[t,i]$ показывает, на какие слова $x[i]$ входного текста модель обращает внимание, когда генерирует слово перевода $y[t]$

Модели внимания: аннотирование изображений



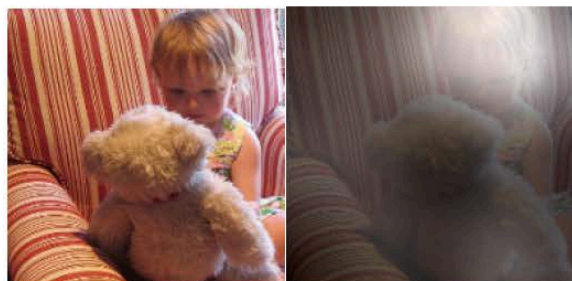
A woman is throwing a frisbee in a park.



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.

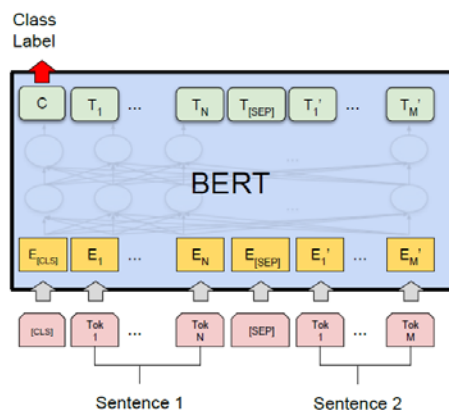


A giraffe standing in a forest with trees in the background.

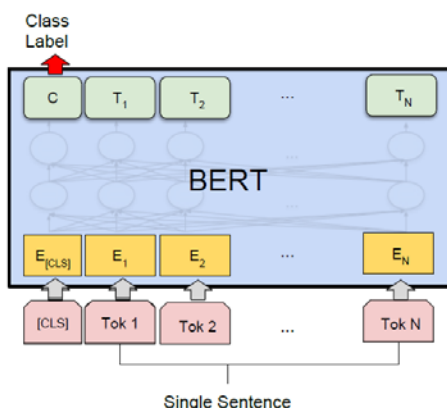
Интерпретация: на какие области модель обращает внимание, генерируя подчёркнутое слово в описании изображения

Трансформеры: большие языковые модели

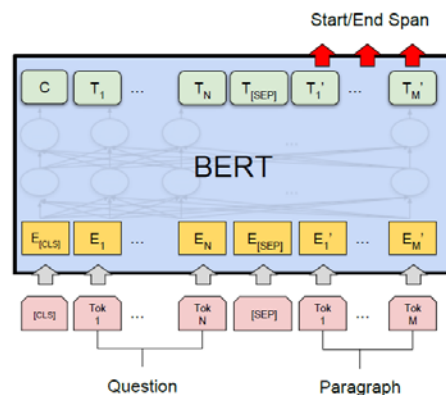
- обучаются векторизовать и предсказывать слова по контексту
- обучаются по терабайтам текстов, «они видели в языке всё»
- мультиязычны: обучаются на десятках языков
- мультизадачны: для каждой новой задачи NLP/NLU достаточно предобученной модели или дообучения на небольшой выборке



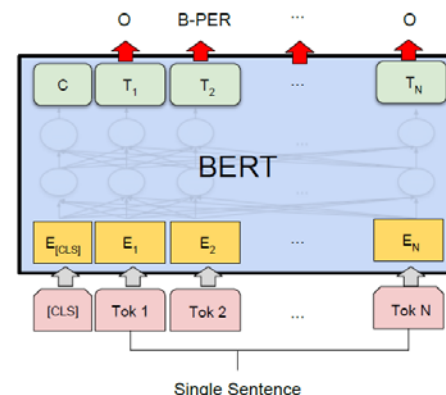
(a) Sentence Pair Classification Tasks:
MNLI, QQP, QNLI, STS-B, MRPC,
RTE, SWAG



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



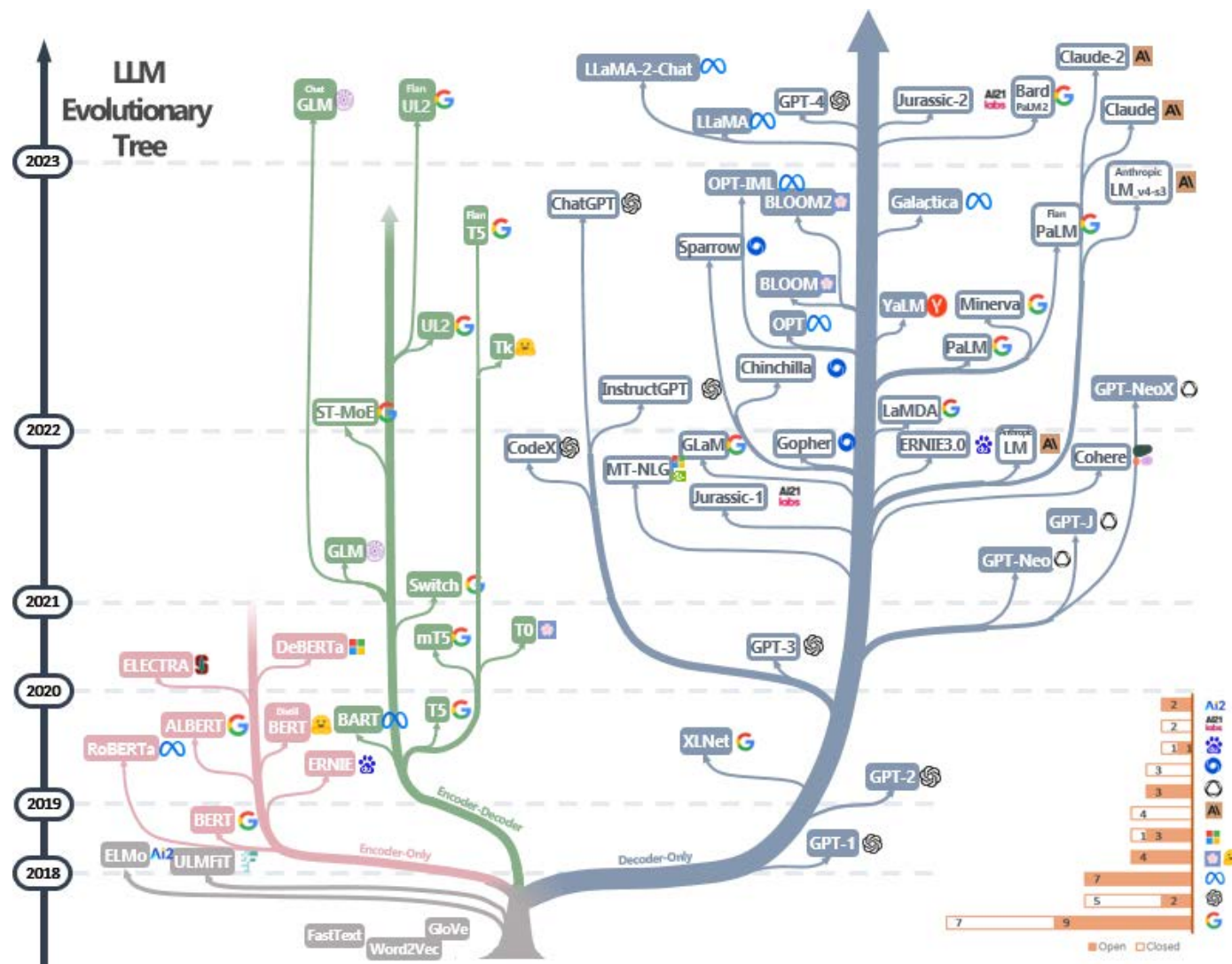
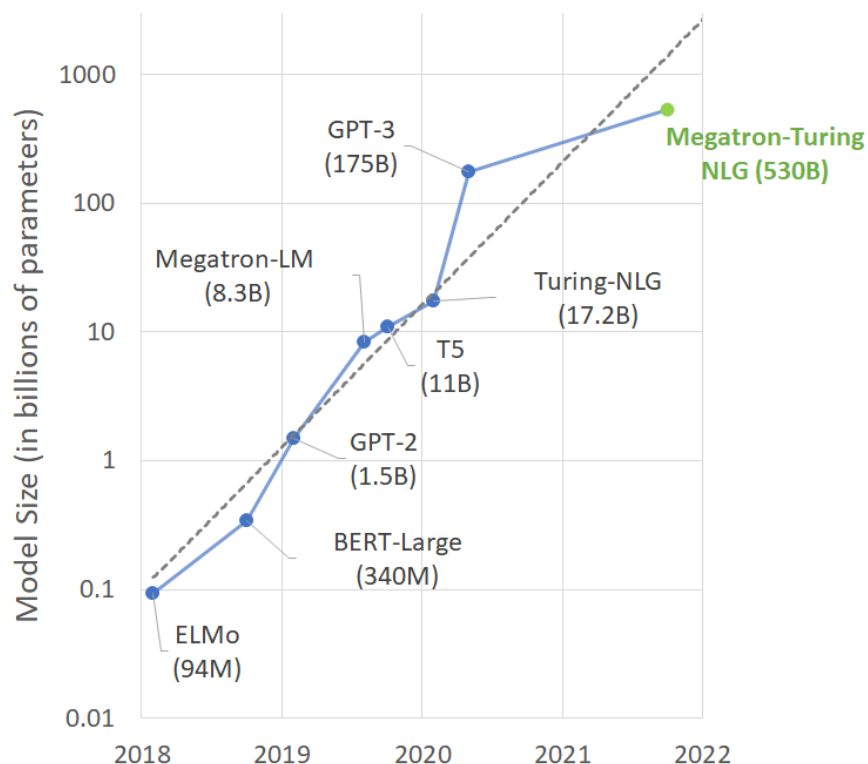
(c) Question Answering Tasks:
SQuAD v1.1



(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

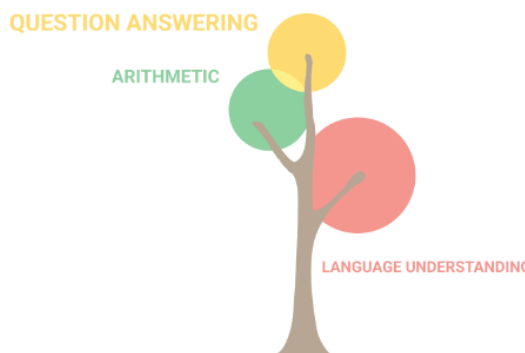
Sébastien Bubeck Varun Chandrasekaran Ronen Eldan Johannes Gehrke
Eric Horvitz Ece Kamar Peter Lee Yin Tat Lee Yuanzhi Li Scott Lundberg
Harsha Nori Hamid Palangi Marco Tulio Ribeiro Yi Zhang

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

Эмерджентные способности модели

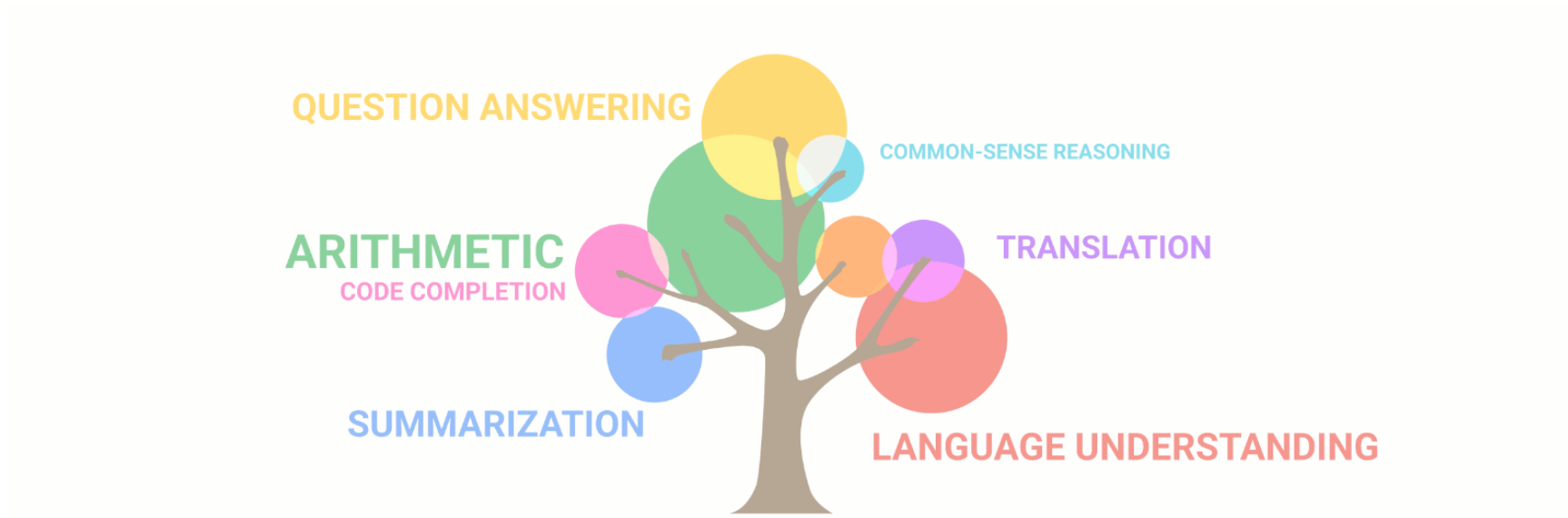


GPT-2: 14-Feb-2019

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

Эмерджентные способности модели

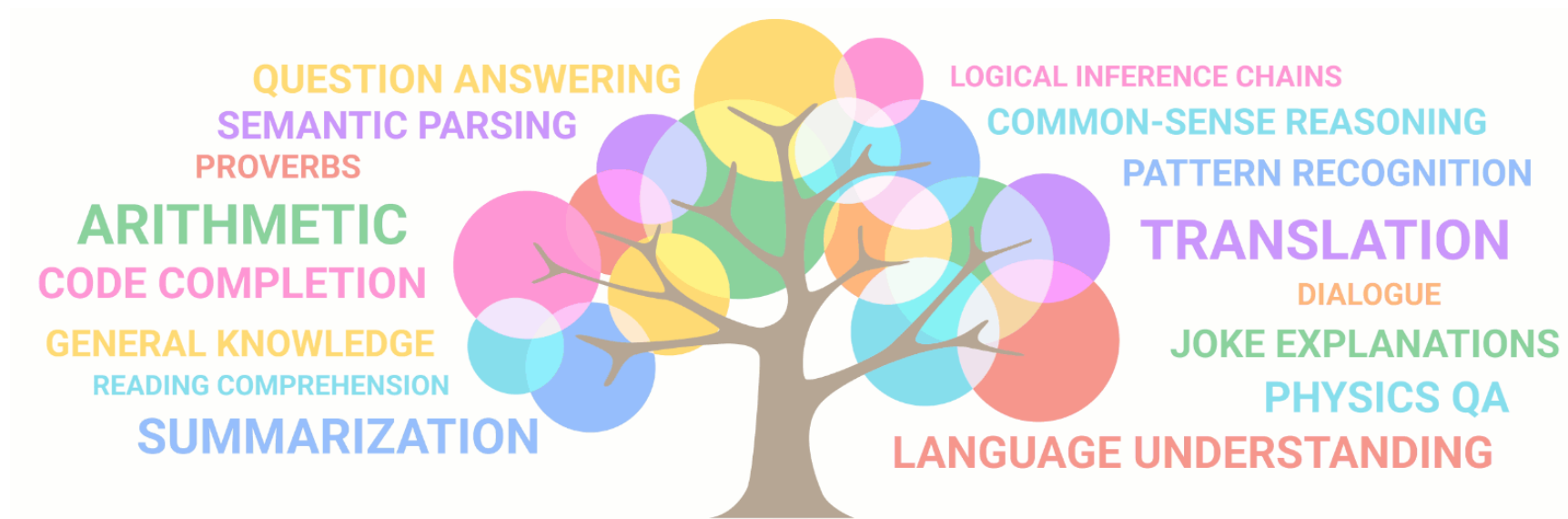


GPT-3: 11-Jun-2020

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

Эмерджентные способности модели

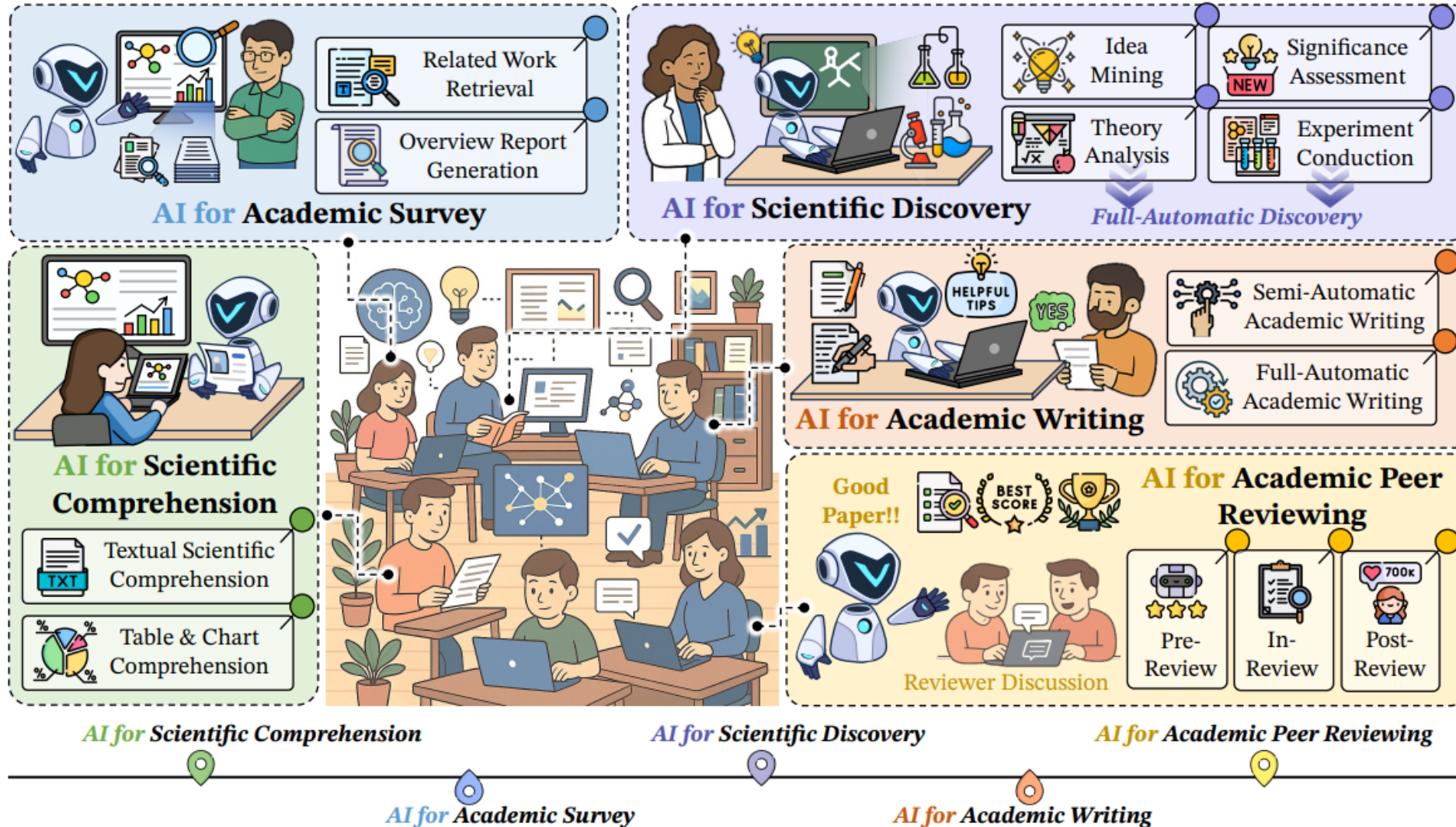


GPT-4: 14-Mar-2023

>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

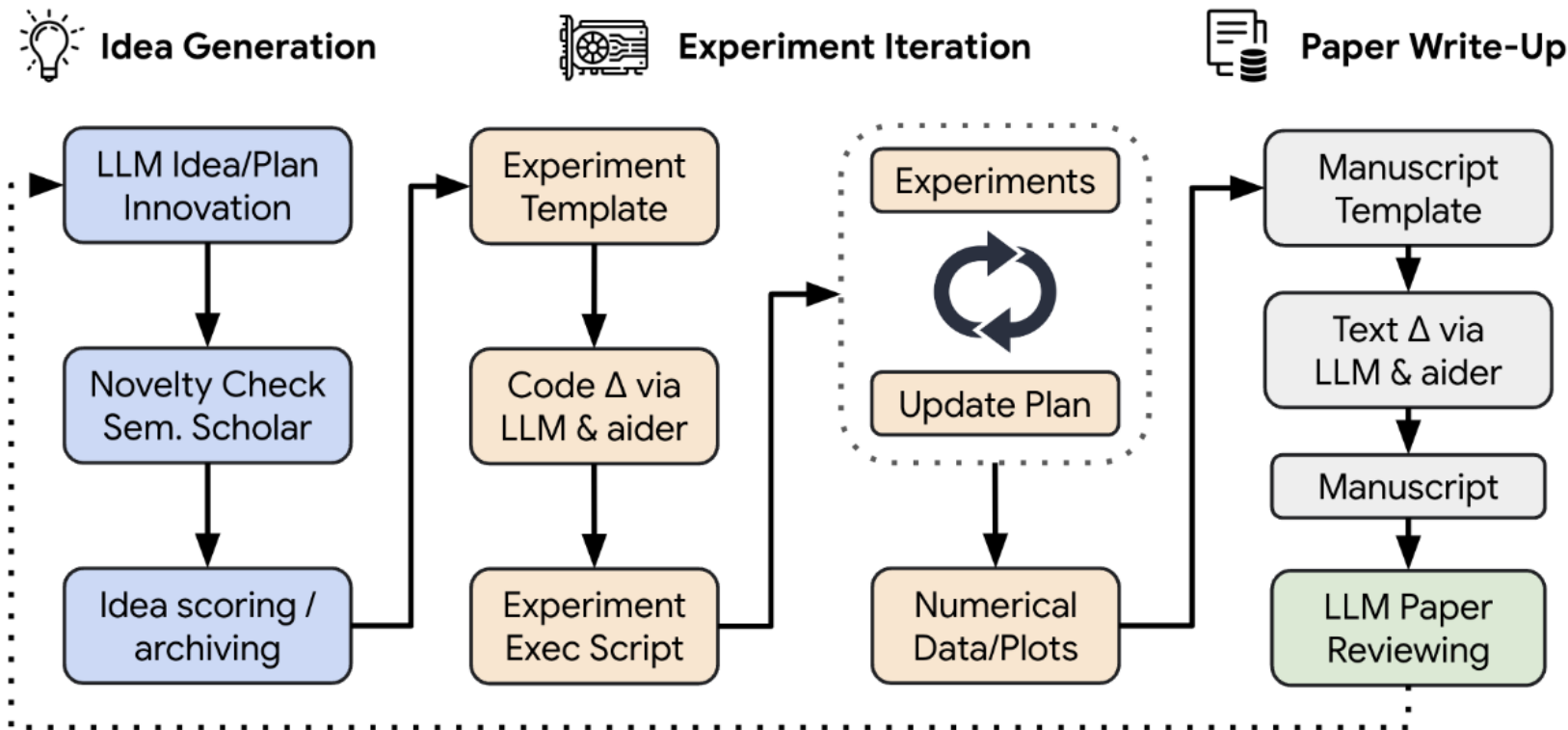
Автоматизация анализа научной литературы



AI4Research can be divided into five key areas:

- (1) AI for Scientific Comprehension
- (2) AI for Academic Survey
- (3) AI for Scientific Discovery
- (4) AI for Academic Writing
- (5) AI for Academic Peer Review

Автоматизация научных исследований



Источник:

*The **AI Scientist**: Towards Fully Automated Open-Ended Scientific Discovery, 2024*

AI Scientist — процесс научных открытий, основанный на LLM.

1. Генерация идей и оценивание их новизны.
2. Определение методов проверки гипотез, включая автоматическую генерацию кода.
3. Автоматическое выполнение экспериментов для сбора результатов, включая данные, графики, таблицы.
4. Результаты обосновываются, объясняются и суммируются в отчете LaTeX.
5. Генерация обзора в соответствии с современной практикой на конференциях по ML.

Автоматизация научных исследований

AI Coscientist — система для создания новых гипотез, исследовательских предложений и экспериментальных протоколов

Построена на Gemini 2.0. Использует:

- расширенные возможности рассуждений и синтеза по сложным темам;
- специализированные агенты, **вдохновленные научным методом**: Generation, Reflection, Ranking, Evolution, Proximity, Meta-review.

Итеративно генерируют, оценивают и уточняют гипотезы с помощью автоматической обратной связи.

Чат + веб-поиск + специализированные модели ИИ для обоснования и улучшения генерации гипотез.

Приложения в биомедицине.

Качество работы подтверждено лабораторными экспериментами по перепрофилированию лекарств, обнаружению целей и изучению устойчивости к противомикробным препаратам.

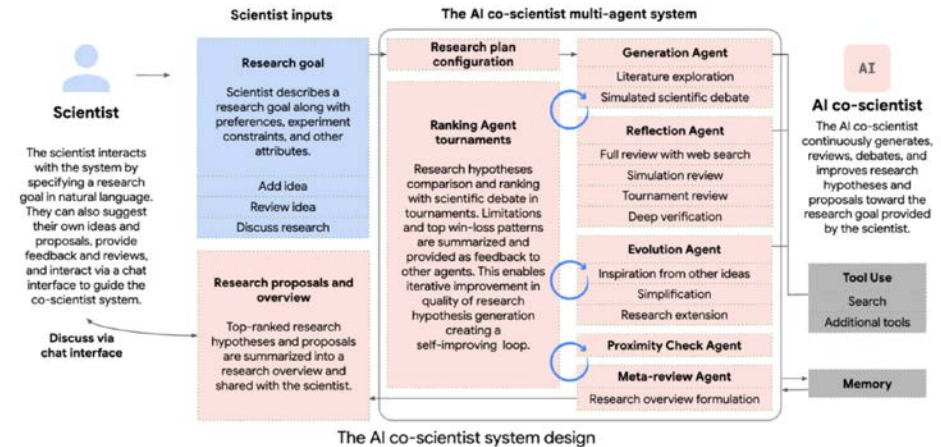
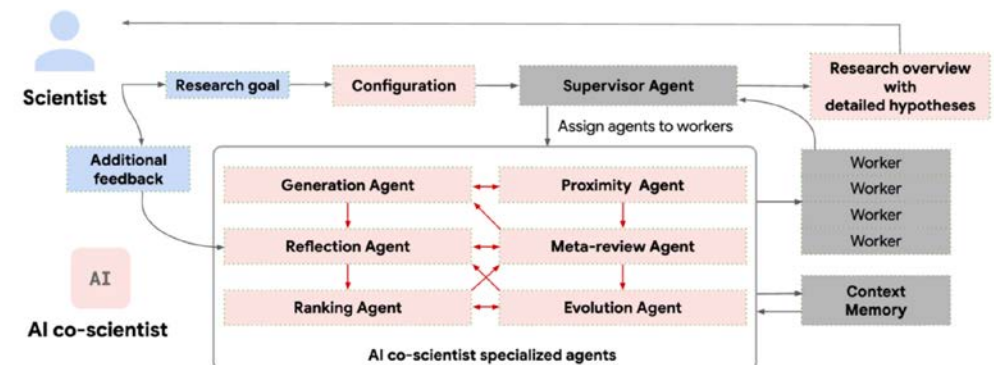


Illustration of the different components in the AI co-scientist multi-agent system and the interaction paradigm between the system and the scientist.



AI co-scientist system overview. Specialized agents (red boxes, with unique roles and logic); scientist input and feedback (blue boxes); system information flow (dark gray arrows); inter-agent feedback (red arrows within the agent section).

Источник: **AI Coscientist: Multi-Agent AI System Built with Gemini 2.0 (2025)**

<https://research.google/blog/accelerating-scientific-breakthroughs-with-an-ai-co-scientist>

Автоматизация научных исследований

OpenAI Plots Charging \$20,000 a Month For PhD-Level Agents

By Stephanie Palazzolo and Cory Weinberg



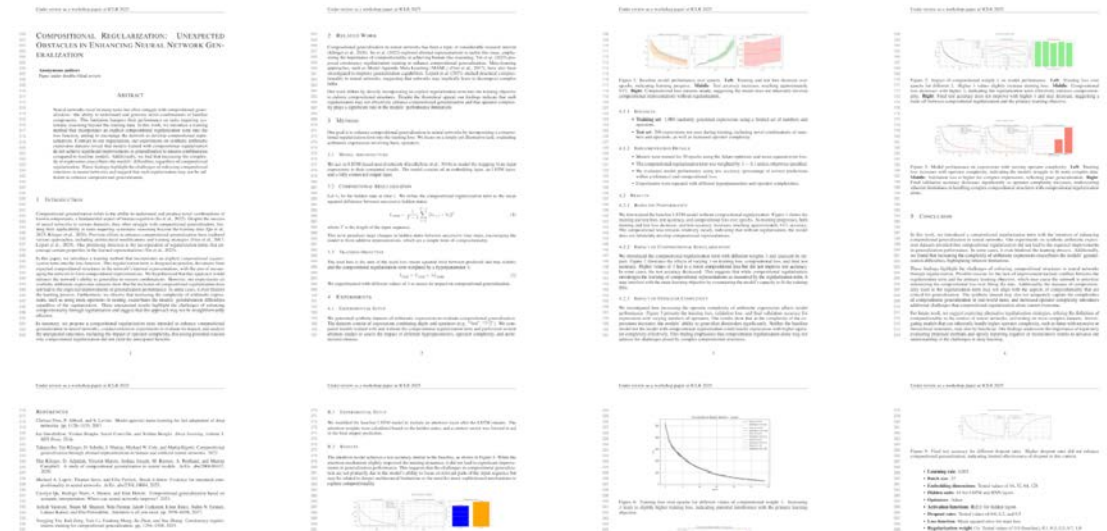
Источник: <https://www.theinformation.com/articles/openai-plots-charging-20-000-a-month-for-phd-level-agents>

sakana.ai

Статья AI Scientist v2
прошла слепое
рецензирование
на воркшоп ICLR-2025

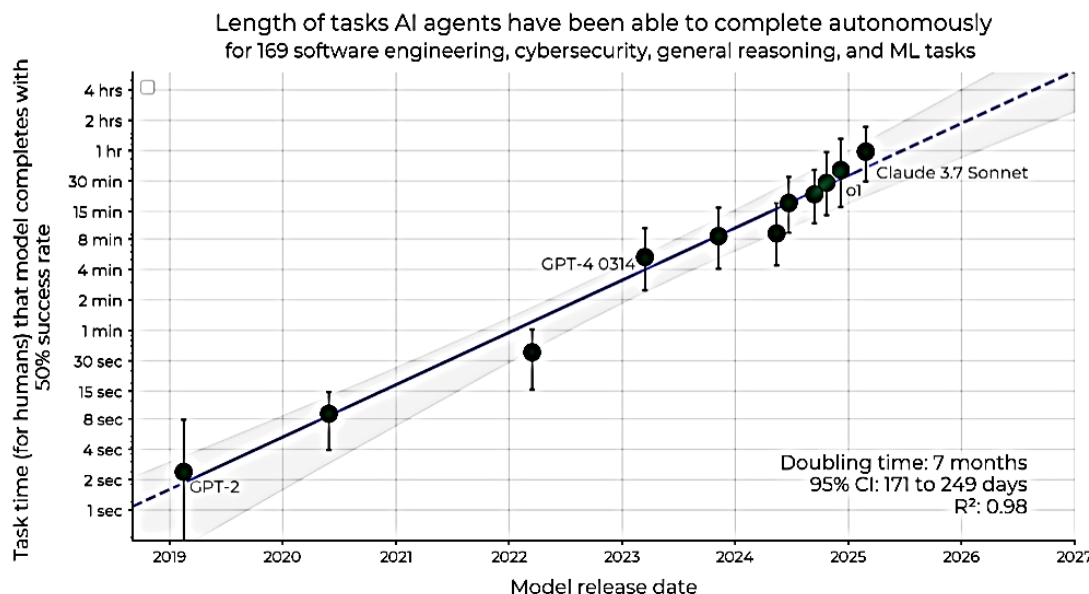
The AI Scientist Generates its First Peer-Reviewed Scientific Publication

March 12, 2025



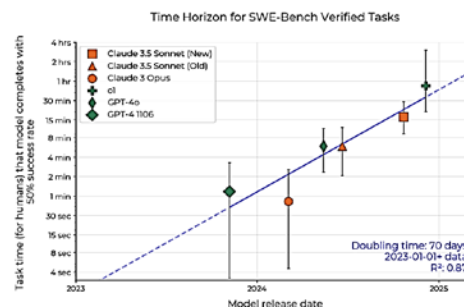
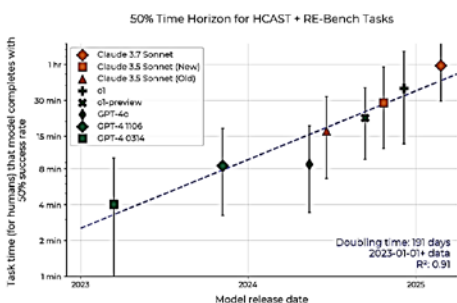
Источник: <https://sakana.ai/ai-scientist-first-publication/>

METR: как ИИ решает сложные задачи



ИИ как рабочая сила в экономике

- Сложность задач (измеряемая временем, которое требуется людям-профессионалам), которые универсальные автономные агенты могут выполнить с 50% надежностью, **удваивается каждые 7 месяцев в течение последних 6 лет**



- В программировании ещё быстрее: **каждые 70 дней!**

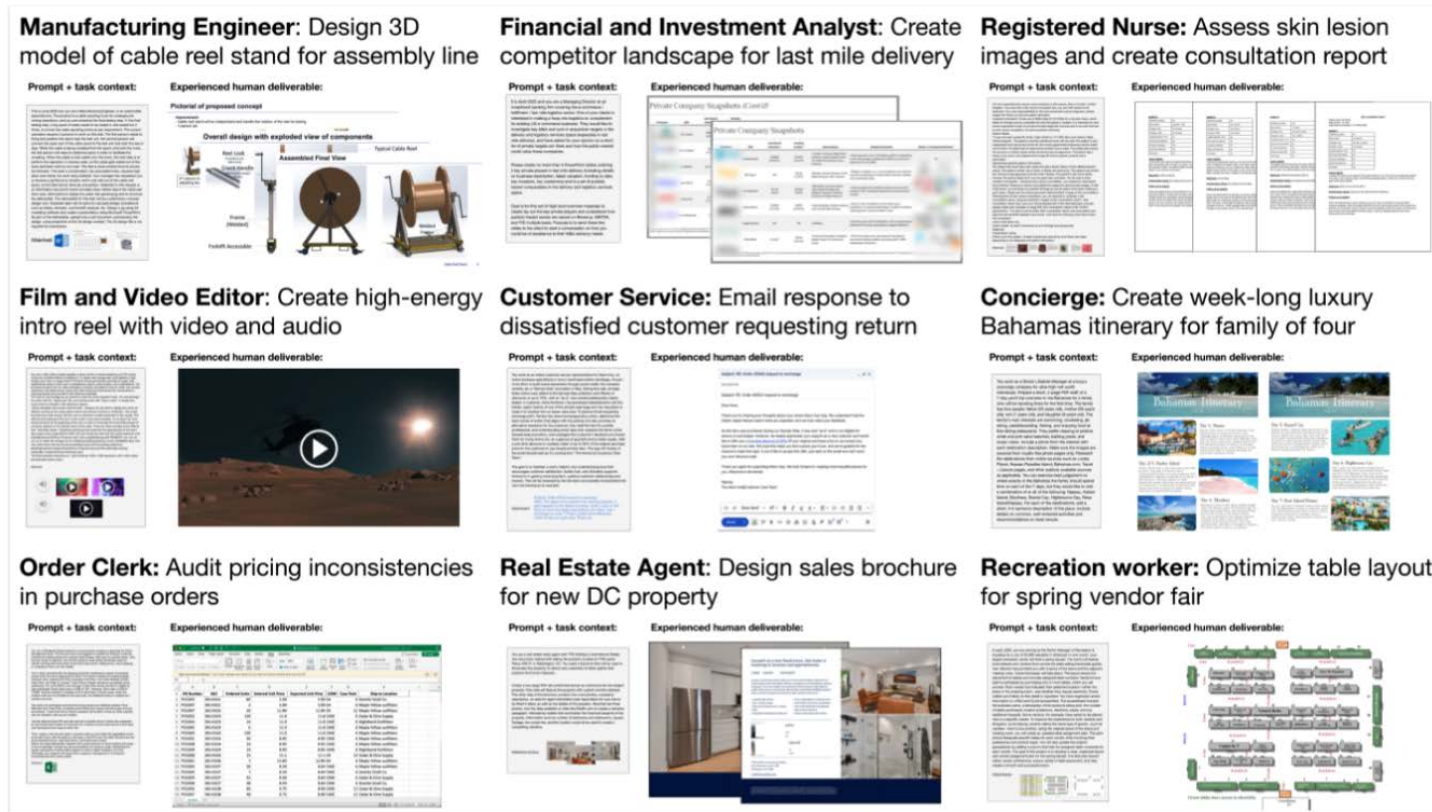
GDPval: ИИ для ВВП

бенчмарк ИИ-агентов-профессионалов (сентябрь 2025)

44 профессиональные задачи из 9 областей экономики

Пример запроса (300 слов)

- **Вы — инженер-технолог в логистической компании,** занимающейся обработкой больших объёмов посылок... *(описание деталей задачи)*
- **Создайте качественную схему процесса** в формате PDF, которая наглядно продемонстрирует стандартизированную и оптимизированную версию того, как должен функционировать сквозной поток отправок... *(описание требований к результату)*

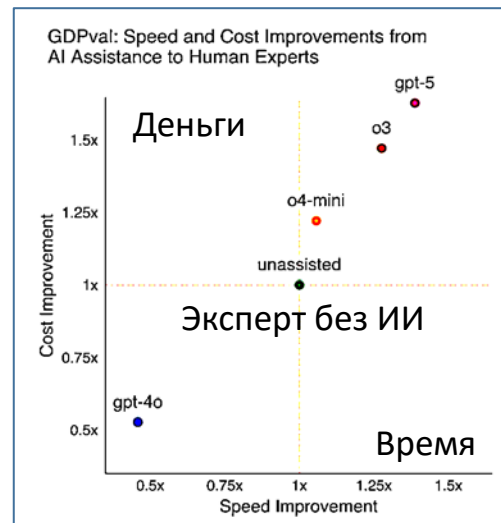
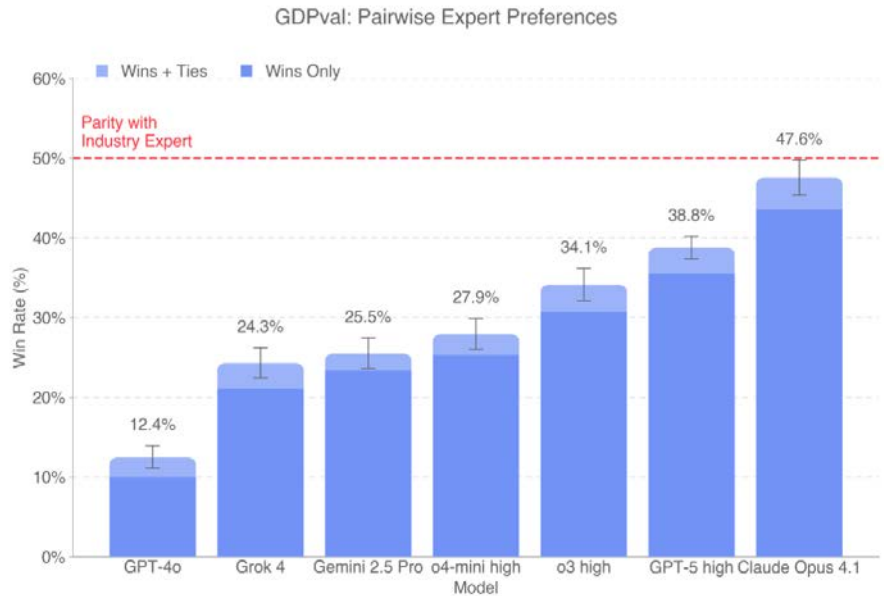


GDPval: Evaluating AI Model Performance on Real-World Economically Valuable Tasks (Open AI)

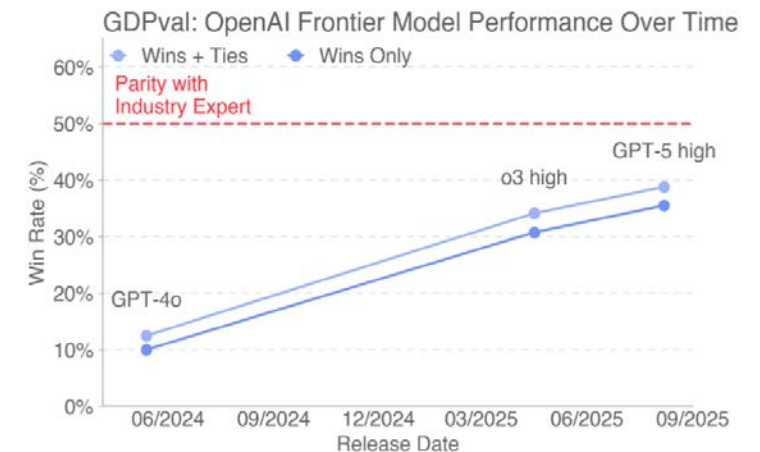
<https://huggingface.co/datasets/openai/gdpval>

GDPval: ИИ для ВВП

бенчмарк ИИ-агентов-профессионалов (сентябрь 2025)



За год модели стали в 4 раза ближе к уровню эксперта



GDPval: Evaluating AI Model Performance on Real-World Economically Valuable Tasks (Open AI)

<https://huggingface.co/datasets/openai/gdpval>

Миф об интеллекте

«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 модель машинного обучения
 - для предсказания одного слова по очень длинному контексту,
 - доученная для рассуждений и ответов на вопросы,
 - с помощью оптимизации миллиардов параметров

Развитие AI в соревновании с человечеством?



этапы
развития

1956: дартмутский семинар

AI -- новая наука и мечта учёных

2020: эмерджентный AI (LLM GPT-3)

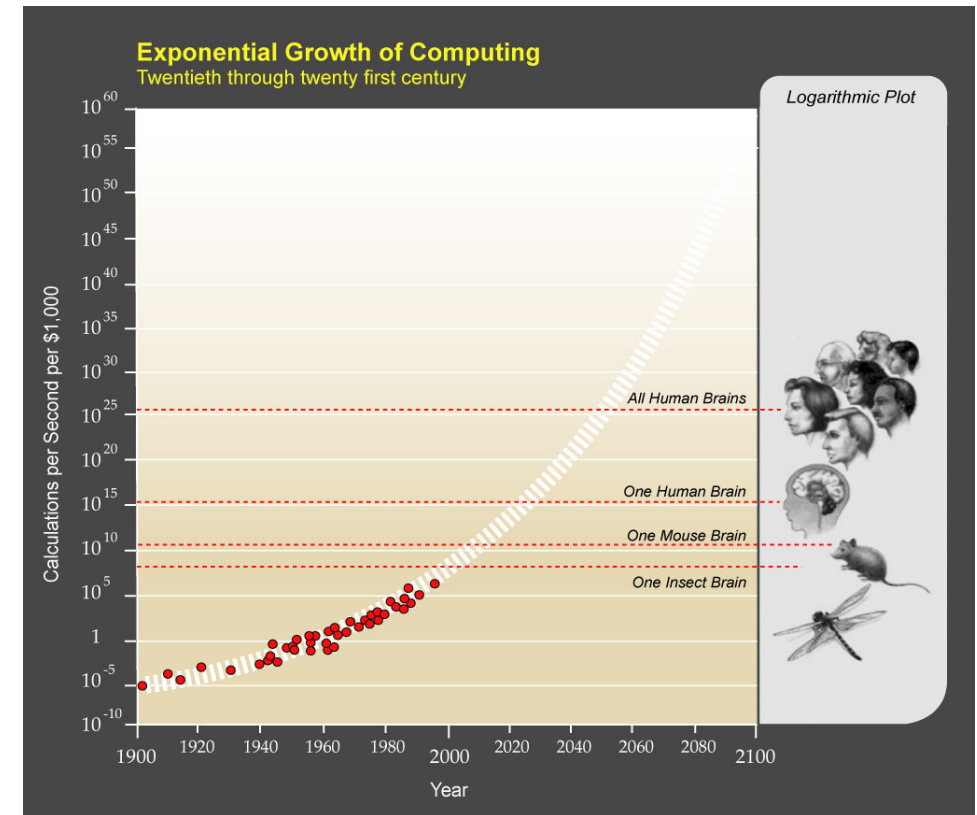
AGI -- Artificial General Intelligence

превосходит интеллект человека

ASI -- Artificial Super-Intelligence
(искусственный супер-интеллект)

превосходит интеллект больших
человеческих коллективов

универсальный ИИ (Universal AI)



Ray Kurzweil. The Singularity Is Near. 2005

Рэй Курцвейл. Революция разума: на подступах к Сингулярности. Как технологии изменят общество и сознание. 2025

Tim Genewein et al. [From AGI to ASI](#). — DeepMind, ArXiv 2606.12683. — 10 Jun 2026.

<http://www.machinelearning.ru>: [Публикация:DeepMind 2026 From AGI to ASI](#)



Будущее AI: негативные сценарии



неблагоприятные
сценарии,
риски и угрозы

источники
постановок
задач

массовые когнитивные искажения в восприятии чатов GPT

захламление и деградация информационного пространства

смещённость рекомендаций, информационные пузыри

ментальные войны, усиление эффектов постправды

усиление технологического и социального неравенства

лишние люди, рост безработицы

эскапизм в виртуальные среды

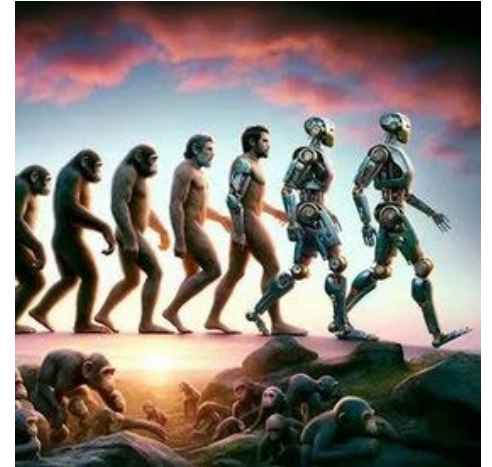
деградация человеческого интеллекта

автономные системы вооружений, риск потери контроля

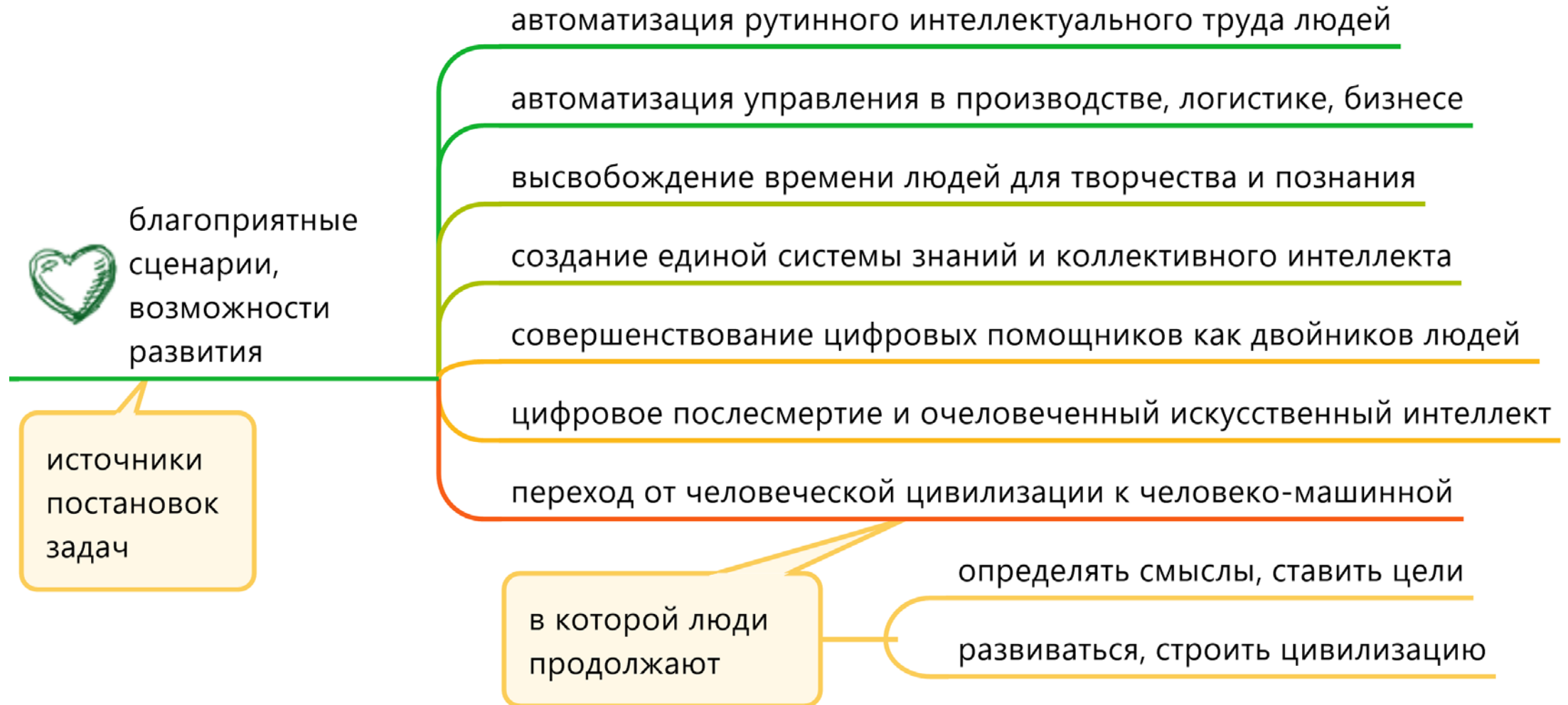
сингулярность по Курцвейлу, потеря управляемости

передача управления и целеполагания машинам

переход от человеческой цивилизации к машинной



Будущее AI: позитивные сценарии

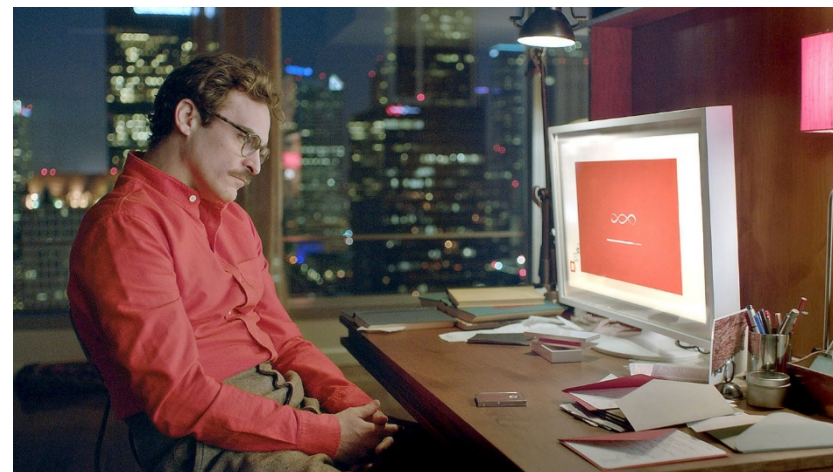


Перспектива AI: персональные помощники

Фильм «Она» (Her, 2013)

режиссёр Спайк Джонз, «Оскар» за сценарий

- голосовая операционная система
- анализирует всю деловую коммуникацию
- ведёт дела и переписку, генерирует идеи,
- до которых человек не додумался сам
- по контексту ищет информацию в сети
- формирует модель личности пользователя — интересов, привычек, потребностей, деловых и профессиональных компетенций
- понимает эмоции человека, способна манипулировать человеком



Тот самый парень, который влюбился в операционку
(жанр фильма — фантастическая мелодрама)

Это будущее уже наступило

Перспектива AI: следующий шаг

- **Жизнь человека — это регистрируемый текстовый поток:**

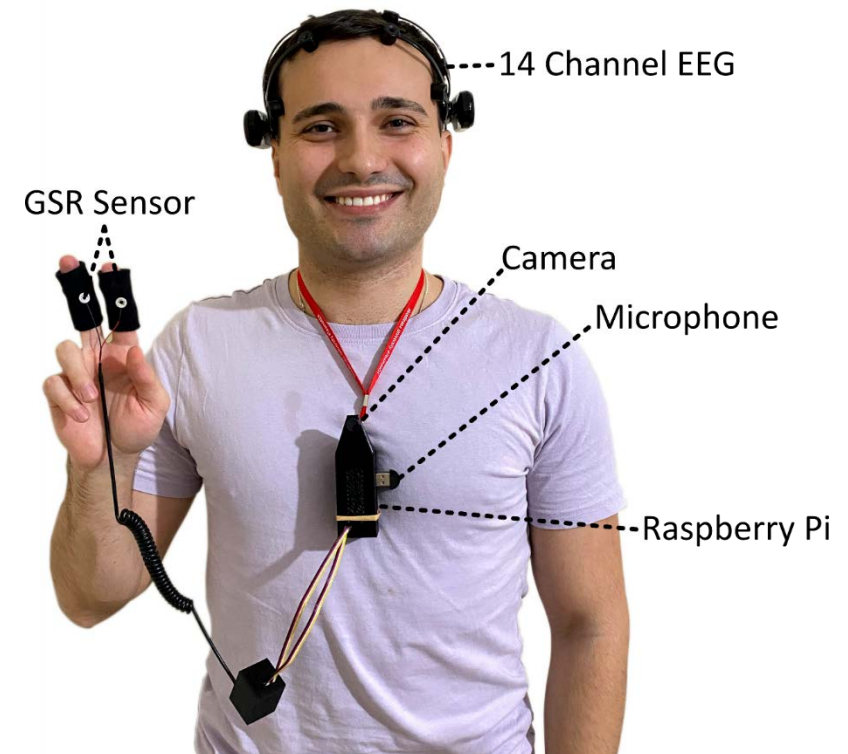
браузеры, почта, соцсети, мессенджеры, пользовательские документы, файлы, системы учёта времени и ведения проектов, ВКС, видео, аудио, голос, VR/AR, ...

- **FPFM (First-person Foundation Model):**

каждый человек — генератор уникального потока семантических векторов, по объёму сопоставимого с Интернетом

Они жалуются, что у них закончился Интернет?

Но у них остались мы — 8 миллиардов людей



Перспектива AI: от помощника к аватару

- **Чел**овек воспитывает и обучает своего персонального **Пом**ощника всю жизнь, в процессе всей своей разнообразной деятельности
- **Чел** делегирует **Пом**у всё больше своих информационных функций (поиск, обучение, коммуникация, анализ, принятие решений)
- **Пом** становится записной книжкой и поминутным дневником, постепенно перенимает *код личности Чела*, дополняя сверх**чел**овеческими возможностями
- **Чел**овек умирает.
Помощник переходит в новое состояние **Ава**тара — *Очеловеченного ИИ*, носителя личностного кода **чел**овека



Перспектива AI: очеловеченный ИИ

- **Ава** продолжит приносить пользу обществу, выполняя профессиональные и коммуникативные функции **Чел+Пома**
- **Ава** продолжит развиваться, накапливая знания и мудрость
- **Ава** отличается от «фабричного безличного» ИИ тем, что лучше понимает людей, их ценности, цели, чувства, взаимоотношения
- **Ава** остаётся доступен для семьи в роли наставника, «хранителя рода»
- Коллективы **Ава** будут осваивать космос, глубины океанов, недра земли



Какие технологии развивать и ради чего?

«Сама по себе технология не является ни решением человеческих проблем, ни злом; но в конкретной истории она не бывает нейтральной, потому что принимает облик тех, кто ее придумывает, оплачивает, направляет и использует»

«Никогда еще человечество не обладало такой властью над самим собой»

«Ради какого замысла и в каком духе мы должны трудиться, чтобы сохранить и умножить величие человечества, полученное нами в дар?»

«Технический прогресс, сам по себе ценный, требует осознания стоящей за ним антропологической модели и целей, которые он преследует»



Magnifica Humanitas
(Великолепное человечество)
— энциклика папы римского Льва XIV,
25мая 2026 г.
**о защите человеческой личности
в эпоху искусственного интеллекта**

Антропоцентричное определение ИИ



Искусственный интеллект — комплекс технологических решений, позволяющий **имитировать** когнитивные функции человека (включая поиск решений без заранее заданного алгоритма) и получать при выполнении конкретных задач результаты, **сопоставимые** с результатами интеллектуальной деятельности человека или **превосходящие** их.

— *Национальная стратегия развития искусственного интеллекта на период до 2030 года.*

***Искусственный интеллект* —**
вычислительные технологии, повышающие
производительность созидательного
интеллектуального труда людей

не соперник и **не** замена человека,
не «загадочный новый тип разума»,
не повод уподобиться Богу, чтобы
«творить по образу и подобию Своему»



Главное о ИИ и отношении людей к ИИ

- ИИ = Имитация Интеллекта, не объект и не субъект, а технология
- ИИ начинается с постановки задачи *Дано-Найти-Критерий*
- *Люди* ставят задачу и несут ответственность за качество решения, способы и риски его применения
- *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
- *Категорический императив Канта*: в любой нашей деятельности человек — это цель, технология — лишь средство

Рекомендуемые материалы

- *Визильтер Ю. В.* Приоритетные направления исследований и ключевые тенденции развития технологий ИИ. XXII Национальная конференция по искусственному интеллекту. 6–10 октября 2025. Санкт-Петербург. <https://rcai2025.raai.org/>
- Не пропустите открытие тысячелетия! // Vital Math, 13 января 2024, <https://www.youtube.com/watch?v=JZjH0it9Jyg>
- Навигатор по ИИ-ландшафту от DENTONS GLOBAL ADVISORS
Report: AI Decrypted 2026: Mid-Year Assessment, September 29, 2026
- *Воронцов К. В.* Лекции по машинному обучению. www.MachineLearning.ru, 2004-2026.
- *Николенко С.* Машинное обучение: основы, 2025.
- *Николенко С., Кадури́н А., Архангельская Е.* Глубокое обучение: основы, 2024.
- *Марков С.* [Охота на электроовец. Большая книга искусственного интеллекта.](#) Том 1, 2. 2024.
- *Шумский С. А.* Машинный интеллект. РИОР ИНФРА-М, 2020.

Дзен-канал «Цивилизационная идеология»



- Безопасное определение искусственного интеллекта



- Искусственный интеллект и цифровое посмертие



- Что не так с «Революцией разума» по Курцвейлу



- Турбулентная эпоха ИИ уже настала: разбор эссе Билла Гейтса с позиций цивилизационной идеологии