



Российская ассоциация искусственного интеллекта
Всероссийский семинар по искусственному интеллекту
для студентов, аспирантов и молодых ученых — 20 мая 2024

Эволюция идей машинного обучения от Фрэнсиса Бэкона до больших языковых моделей

Воронцов Константин Вячеславович

д.ф.-м.н., профессор РАН,
лаборатория машинного обучения и семантического анализа Института ИИ МГУ,
кафедра математических методов прогнозирования ВМК МГУ,
кафедра машинного обучения и цифровой гуманитаристики МФТИ,

k.vorontsov@iai.msu.ru

Докладчик: *Воронцов Константин Вячеславович*

http://www.machinelearning.ru/wiki/index.php?title=User:Vokov

<http://www.MachineLearning.ru/wiki?title=User:Vokov>

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> — короткая ссылка на эту страницу.

1 Учебные материалы [\[править\]](#)

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

- Предыстория машинного обучения
- Терминология машинного обучения
- Примеры задач машинного обучения

2. Методология машинного обучения

- Нейронные сети и глубокое обучение
- Оптимизационные задачи машинного обучения
- Задачи машинного обучения с векторизацией объектов

3. Большие языковые модели

- Модели внимания и трансформеры, эмерджентность
- О тексто-графическом представлении знаний
- Мифы об искусственном интеллекте

Методология эмпирической индукции

От дедуктивного метода познания к индуктивному:

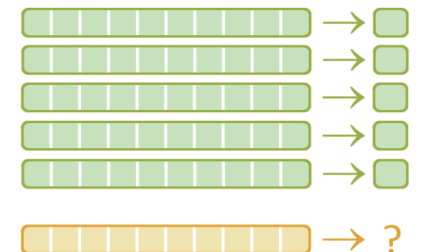
«Не следует полагаться на сформулированные аксиомы и формальные базовые понятия, какими бы привлекательными и справедливыми они не казались. Законы природы нужно «расшифровывать» из фактов опыта. Следует искать правильный метод анализа и обобщения опытных данных; здесь логика Аристотеля не подходит в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

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

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



Фрэнсис Бэкон
(1561--1626)



Задача проведения функции через точки

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

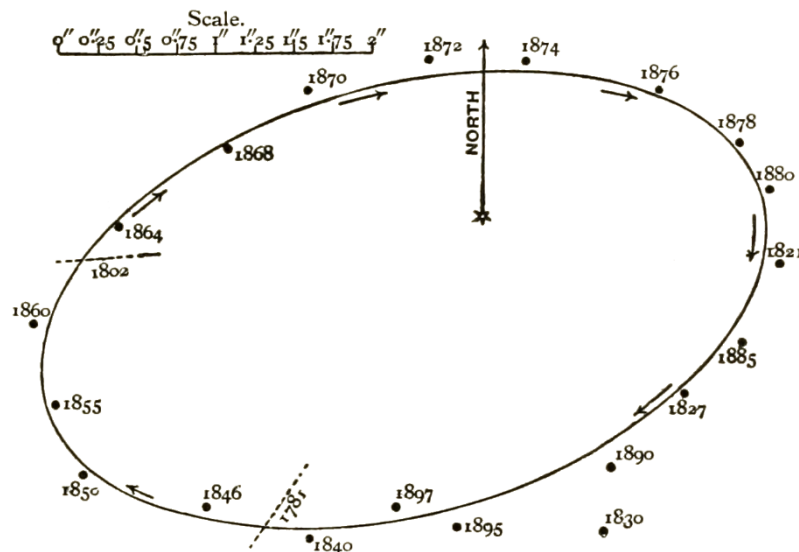
$$a(x, w) = \sum_j w_j f_j(x)$$

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...»

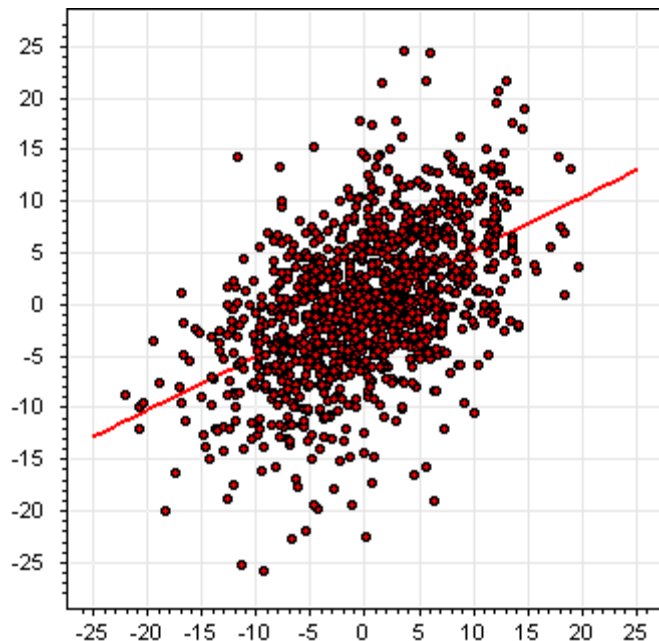
C.F.Gauss. Theory of the motion of the heavenly bodies moving about the Sun in conic sections. 1809.

Задача восстановления регрессии

Исследование наследственности роста (Гальтон, 1886).

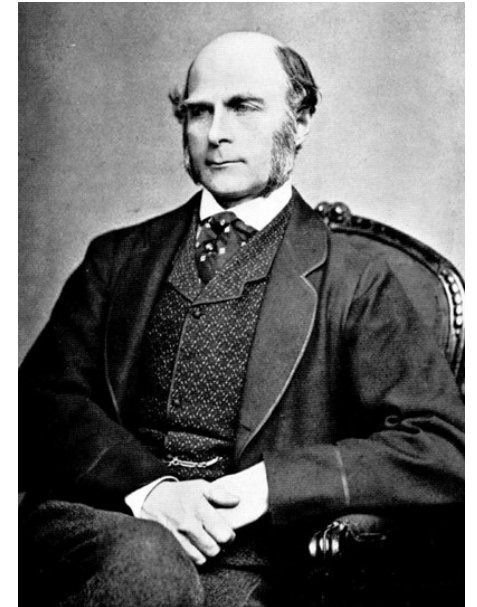
Δ — отклонение роста от среднего в популяции

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



Двойной смысл термина «регрессия»:

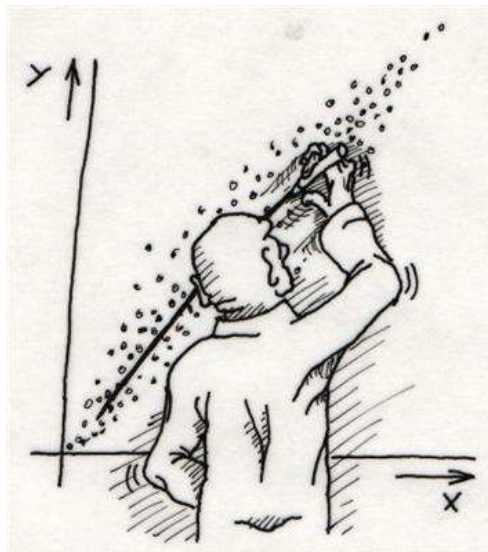
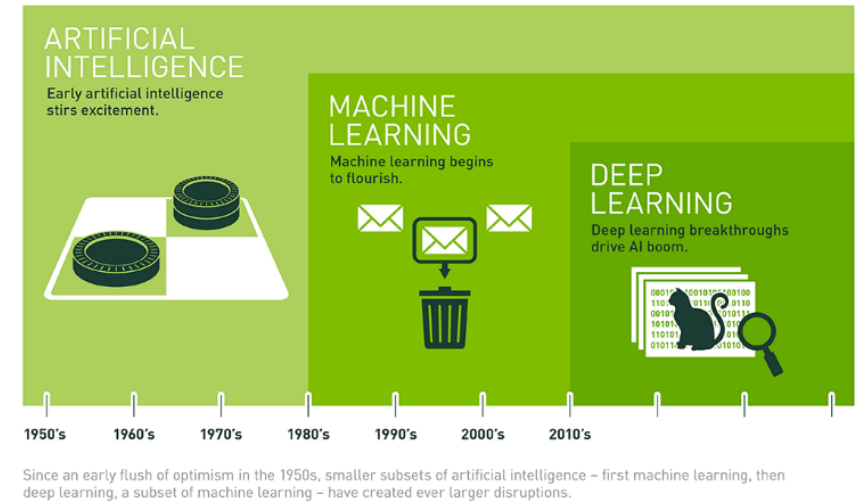
- регрессия (роста) к среднему в популяции
- *необычный «обратный» ход исследования: сначала данные, потом модель*



Фрэнсис Гальтон
(1822--1911)

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

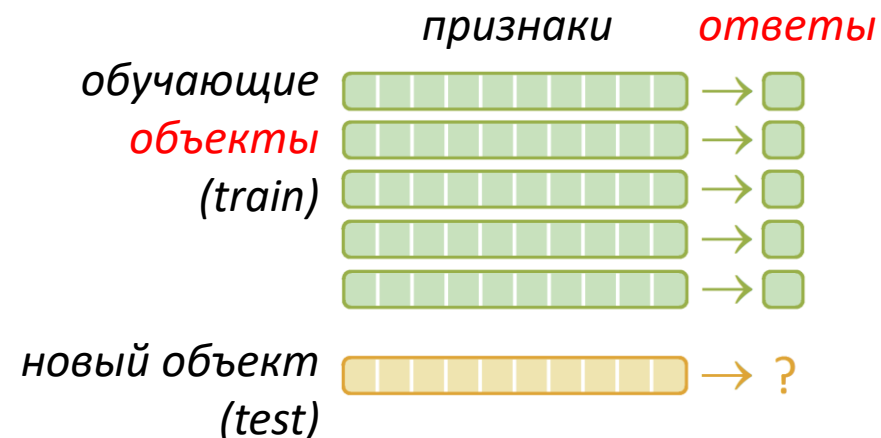
Этап №1 – обучение с учителем

- **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

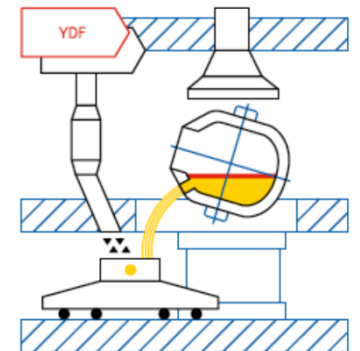
ответ – есть/нет месторождение



- **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



Содержание

1. Задачи машинного обучения

- Предыстория машинного обучения
- Терминология машинного обучения
- Примеры задач машинного обучения

2. Методология машинного обучения

- **Нейронные сети и глубокое обучение**
- **Оптимизационные задачи машинного обучения**
- **Задачи машинного обучения с векторизацией объектов**

3. Большие языковые модели

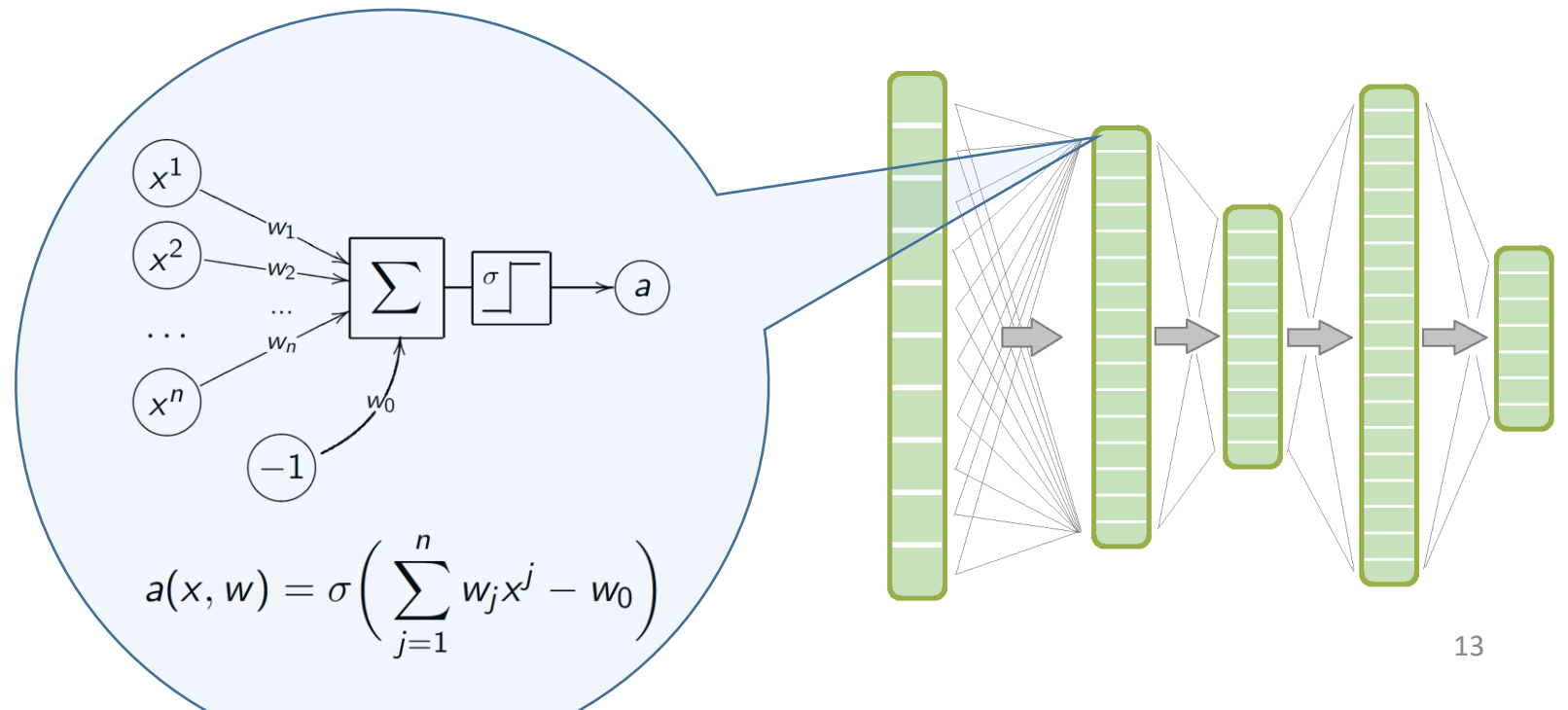
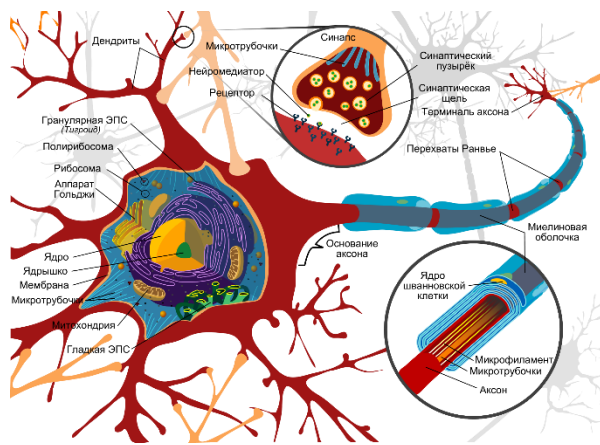
- Модели внимания и трансформеры, эмерджентность
- О тексто-графическом представлении знаний
- Мифы об искусственном интеллекте

Модели искусственных нейронных сетей

На каждом слое сети вектор объекта преобразуется в новый вектор

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

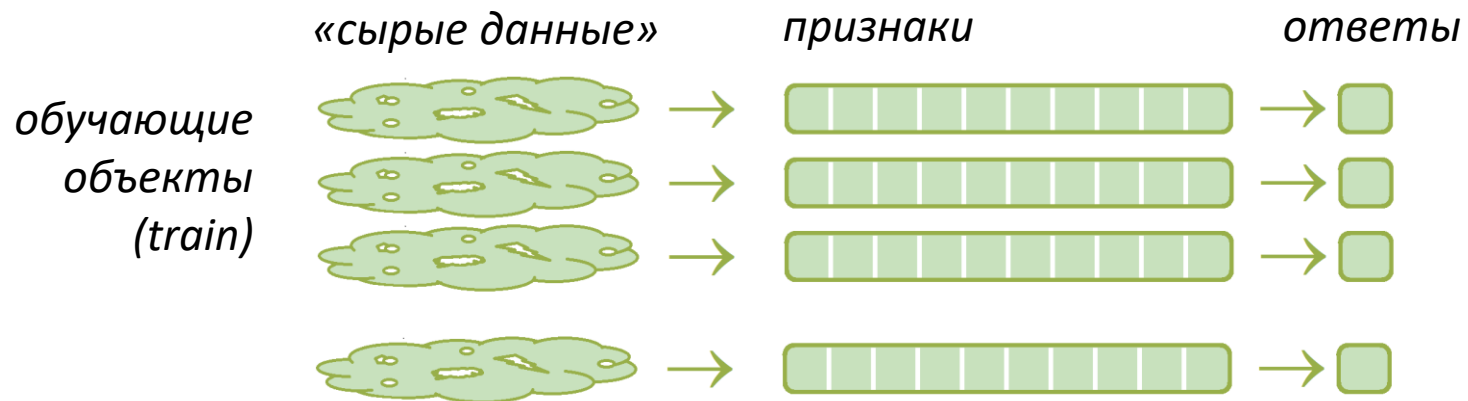
Веса w являются обучаемыми параметрами модели



Модели глубоких нейронных сетей

Вход: сложно структурированные «сырые» данные объектов

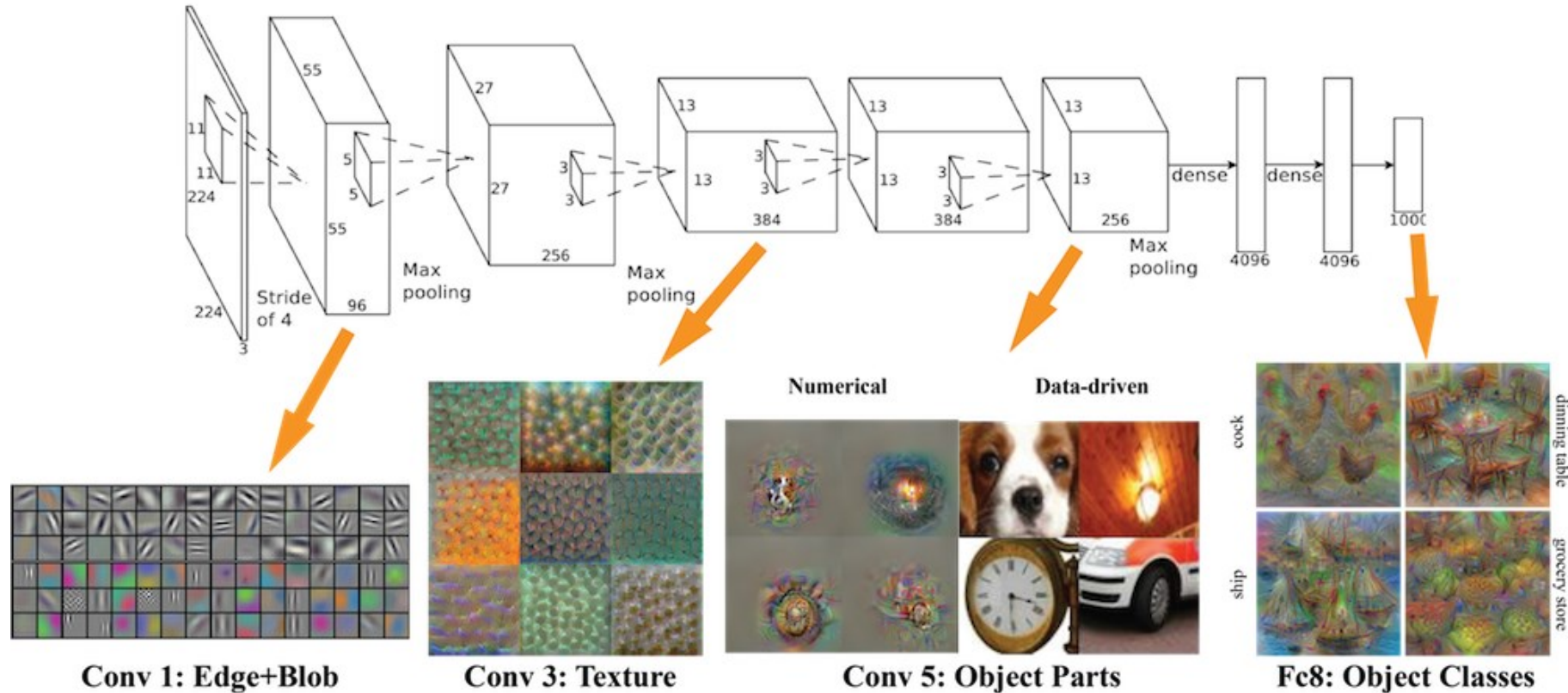
Выход: векторные признаковые представления объектов, затем ответы



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

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

Глубокие свёрточные нейронные сети для классификации объектов на изображениях

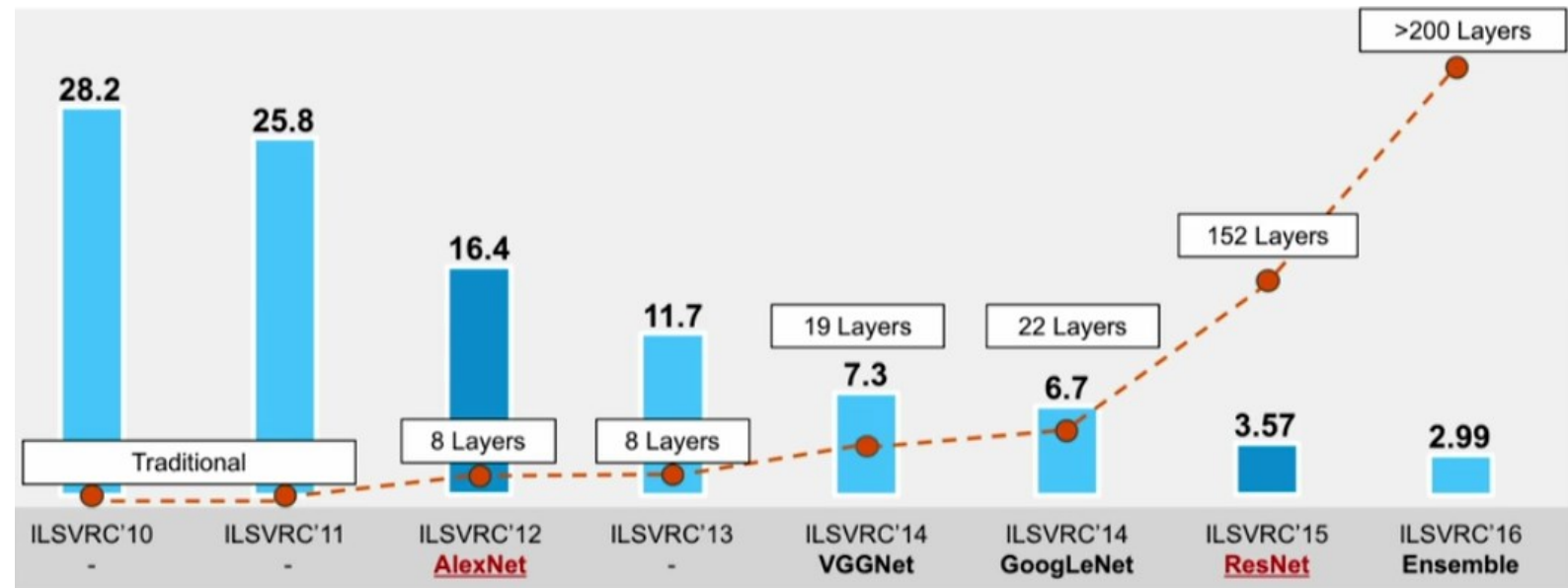


Krizhevsky A., Sutskever I., Hinton G. ImageNet classification with deep convolutional neural networks. 2012.

Роль больших данных

ImageNet: открытая выборка 14М изображений, 20К категорий

IMAGENET



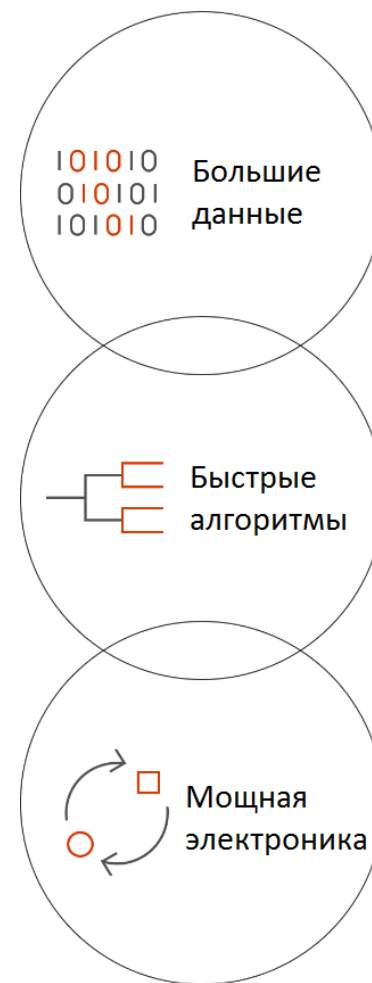
Старт в 2009 г. Человеческий уровень ошибок 5% пройден в 2015 г.

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, закон Мура*
в частности, графические ускорители (GPU)



Машинное обучение – это оптимизация

x – вектор объекта обучающей выборки

$a(x, w)$ – предсказательная модель

w – параметры модели

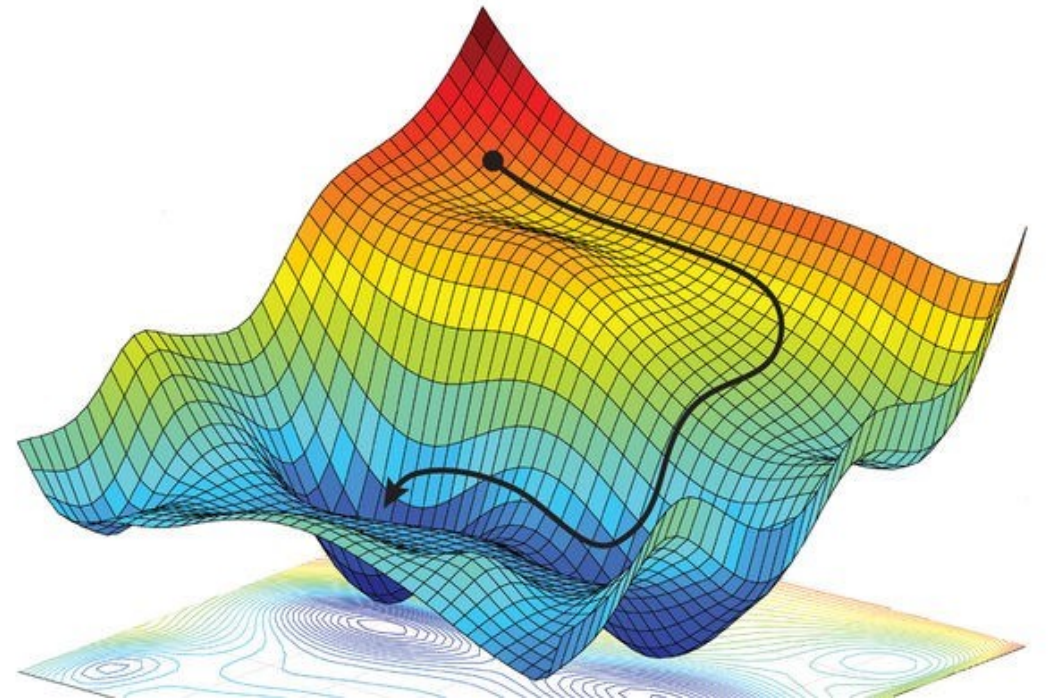
$\text{Loss}(x, w)$ – функция потерь

$Q(w)$ – критерий качества модели

Задача обучения параметров модели:

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

Способ решения – численные методы оптимизации



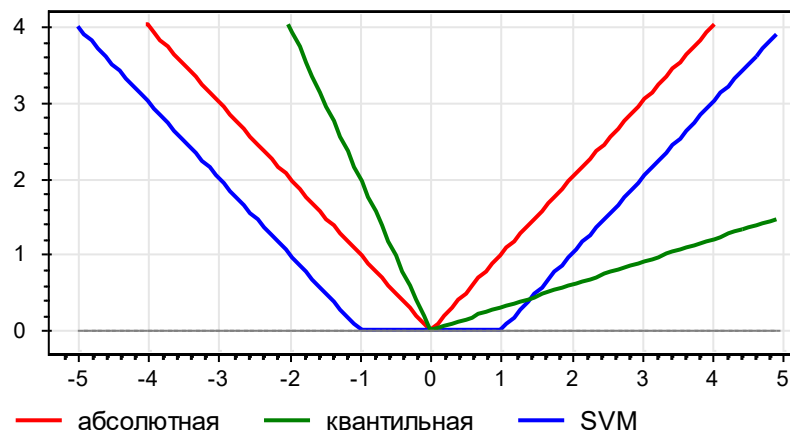
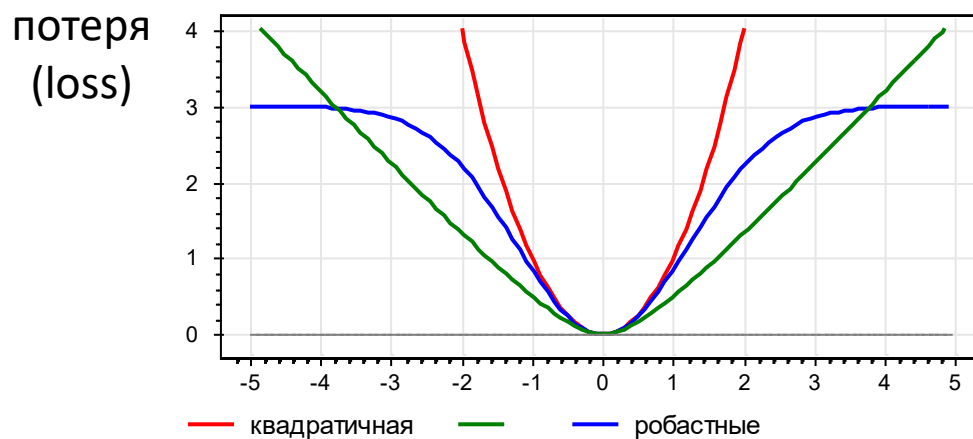
Обучение с учителем (supervised learning): восстановление регрессии (regression)

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)

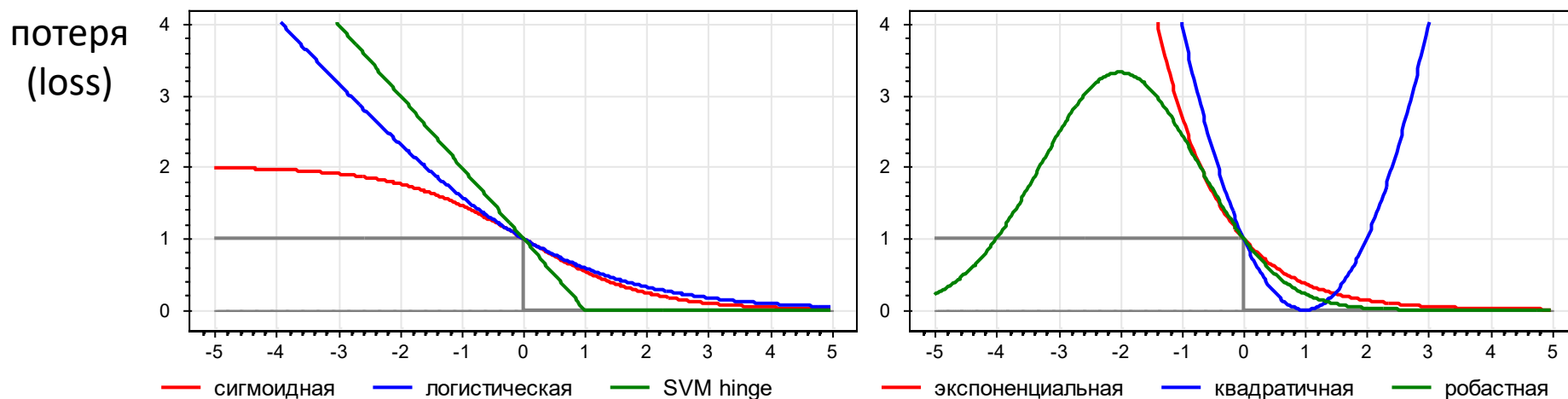
Обучение с учителем (supervised learning): классификация (classification)

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)

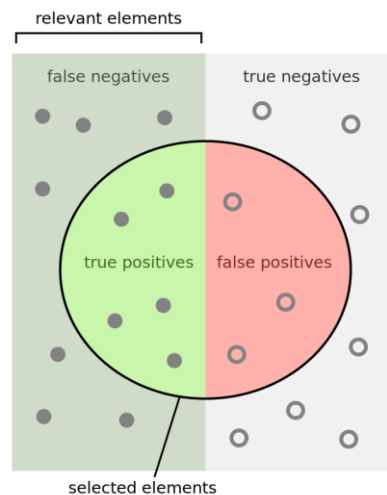
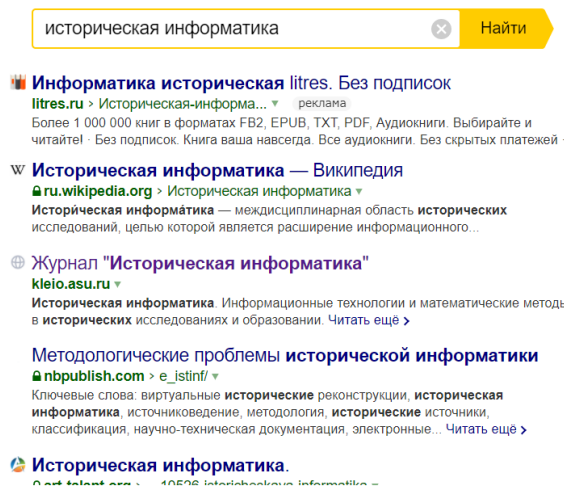
Обучение с учителем (supervised learning): обучение ранжированию (learning to rank)

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 из вариантов*

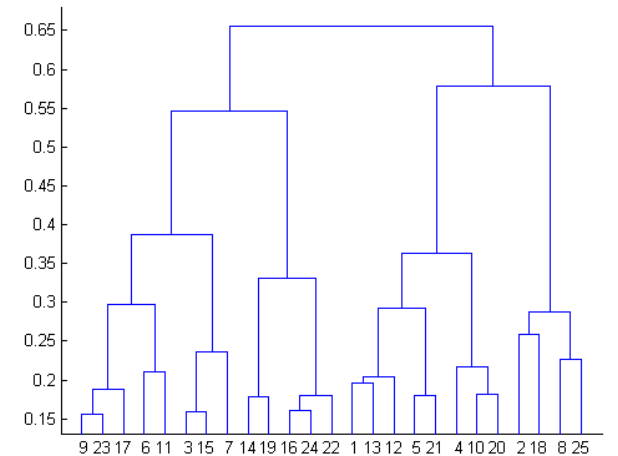
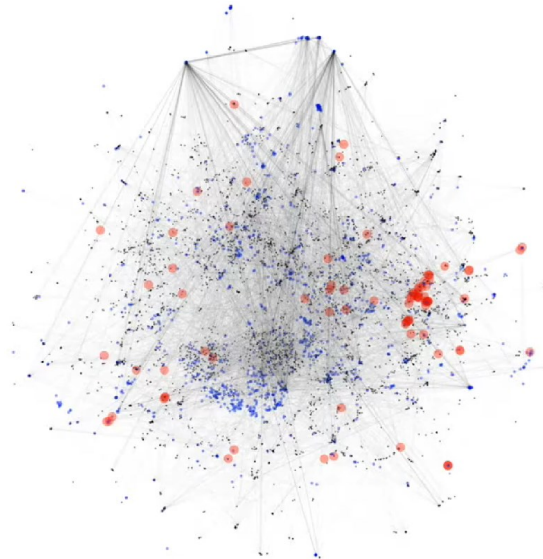
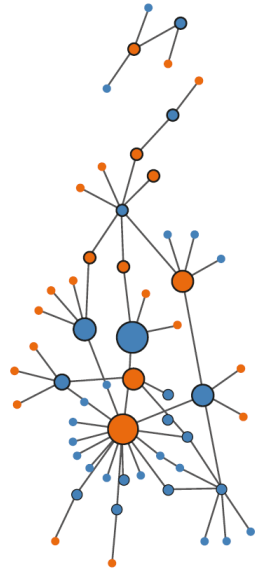
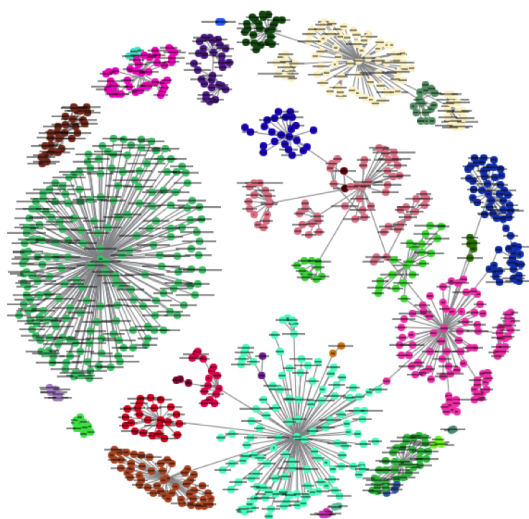
Обучение без учителя (unsupervised learning): кластеризация (clustering)

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



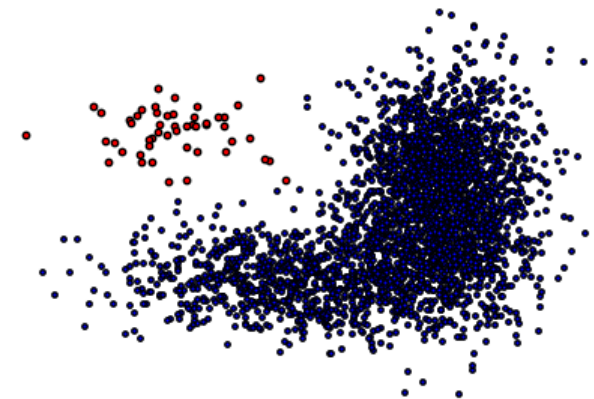
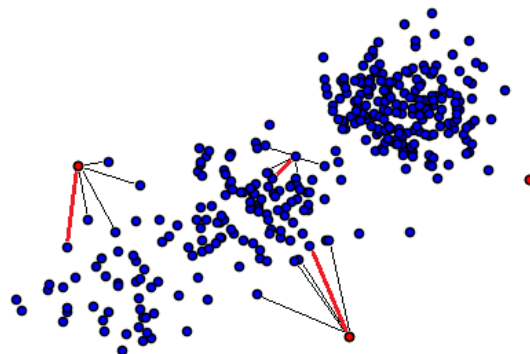
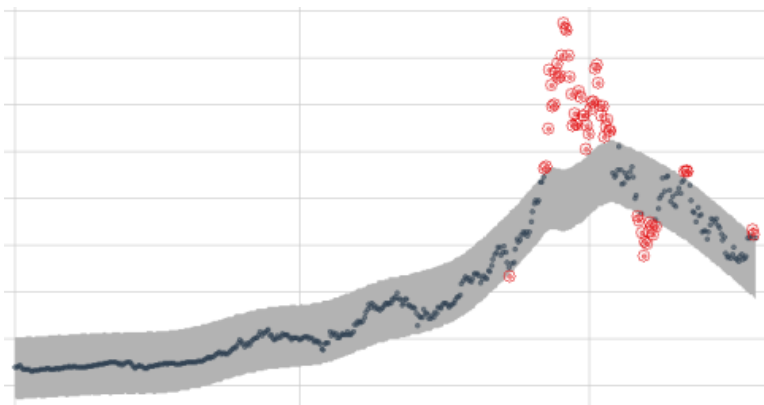
Обучение без учителя: выявление аномалий (anomaly / outlier / novelty detection)

x — вектор объекта

$a(x, w)$ — модель регрессии / классификации / кластеризации

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

Объекты ранжируются по убыванию потерь, анализируются top- k



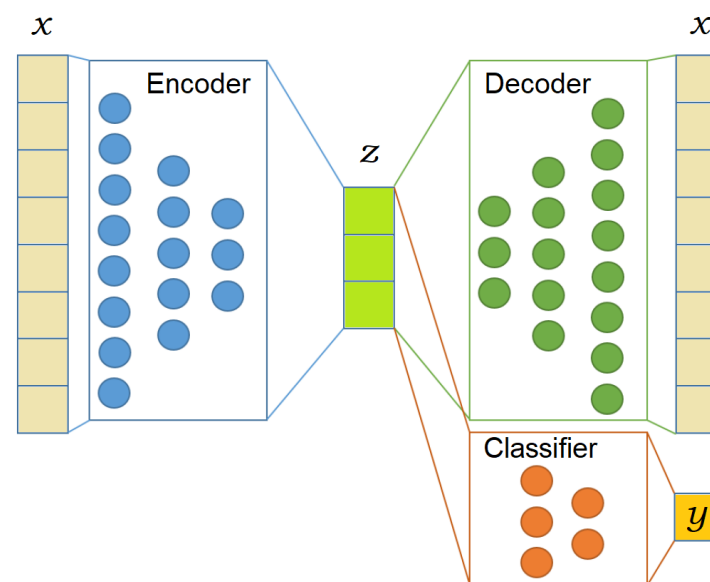
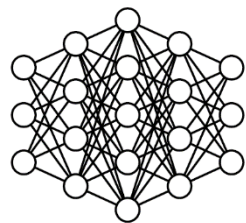
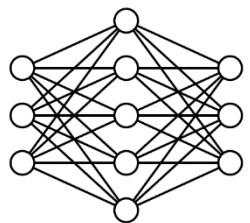
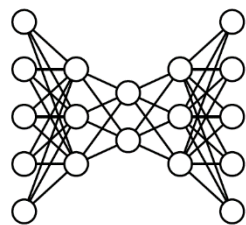
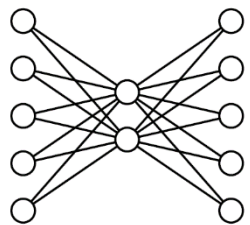
Обучение без учителя: векторизация данных, автокодировщики (autoencoders)

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

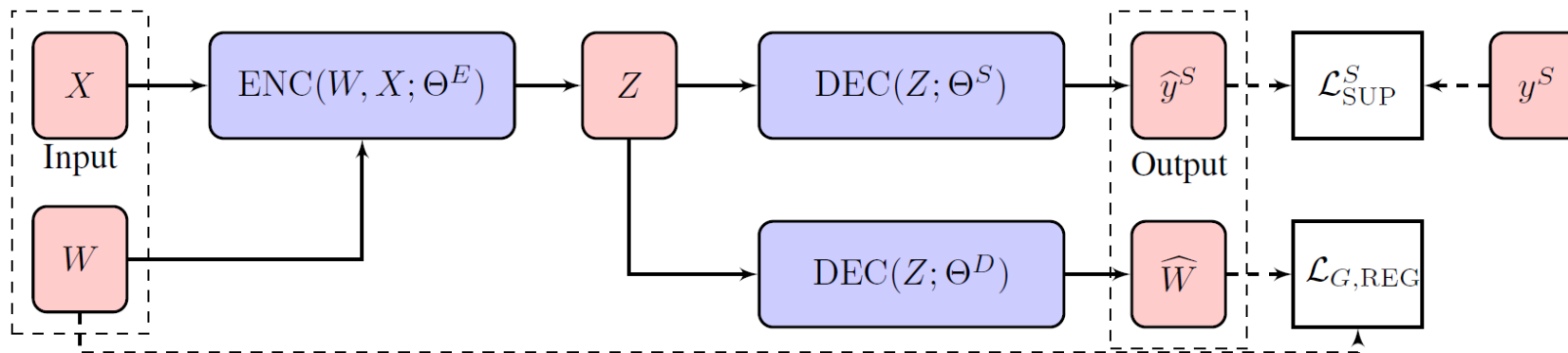
Частичное обучение (semi-supervised learning): векторизация графов (graph embeddings)

$x; (x, x')$ — данные об объектах и взаимодействиях между объектами

$z = f(x, \theta^E)$ — модель векторизации объектов x (вершин графа)

$x' = g(z, \theta^D)$ — модель декодирования z в реконструкцию x'

$\text{Loss}(x, w) = \|g(f(x, \theta^E), \theta^D) - x\| + \tau L_{\text{SUP}}^S(x, \theta^S)$ — два критерия



обучаемая
векторизация
сложных объектов
по данным об их
взаимодействиях

T.Mikolov et al. Efficient estimation of word representations in vector space, 2013.

I.Chami et al. Machine learning on graphs: a model and comprehensive taxonomy. 2020.

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

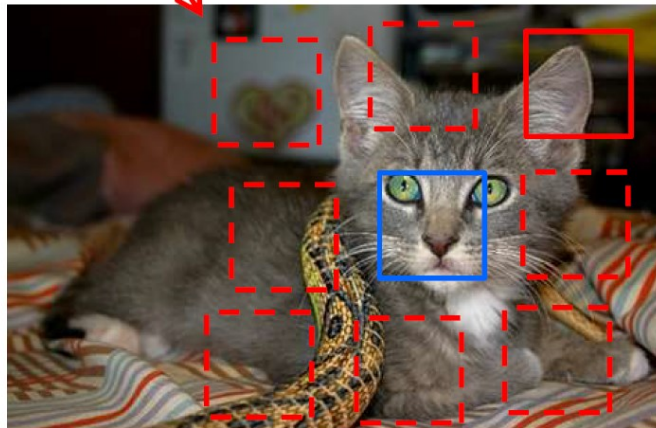
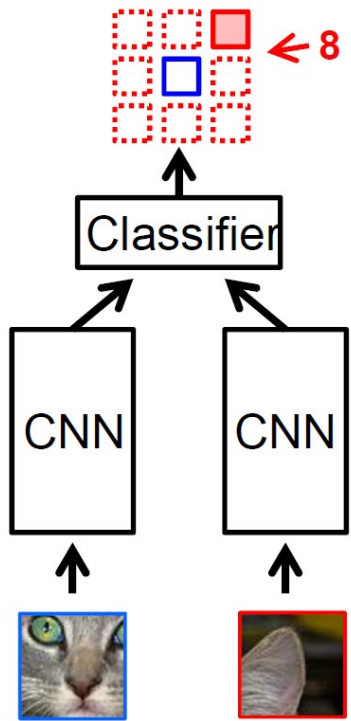
$\min_{w'}: \sum_{x'} \text{Loss}_2(g_2(f(x', w), w'))$ — обучение по своим данным



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

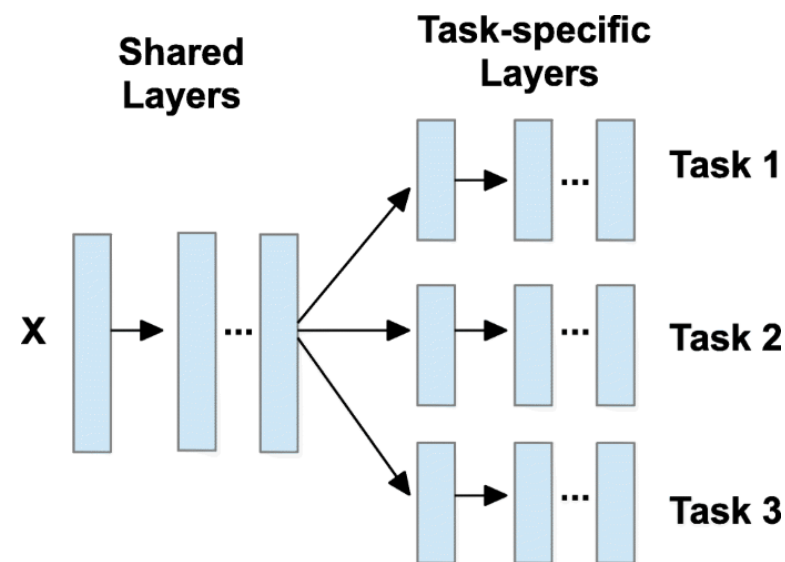
$y = g_t(z, w'_t)$ – часть модели, специфичная для t -й задачи

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

few-shot learning – обучение по малому числу примеров

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

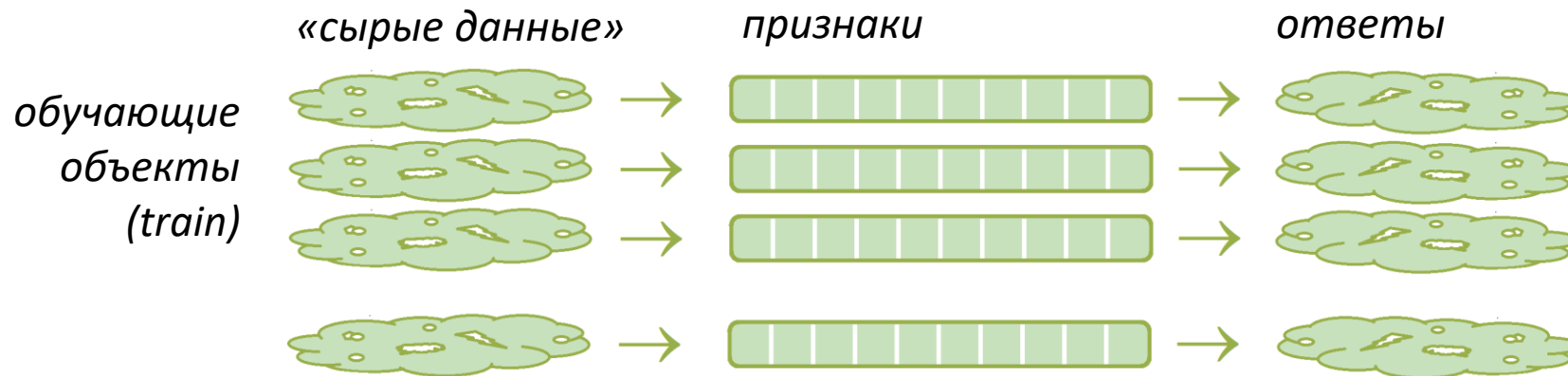
Y.Wang et al. Generalizing from a few examples: a survey on few-shot learning. 2020



Нейронные сети для синтеза объектов

Вход: сложно структурированные объекты

Выход: сложно структурированные ответы



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

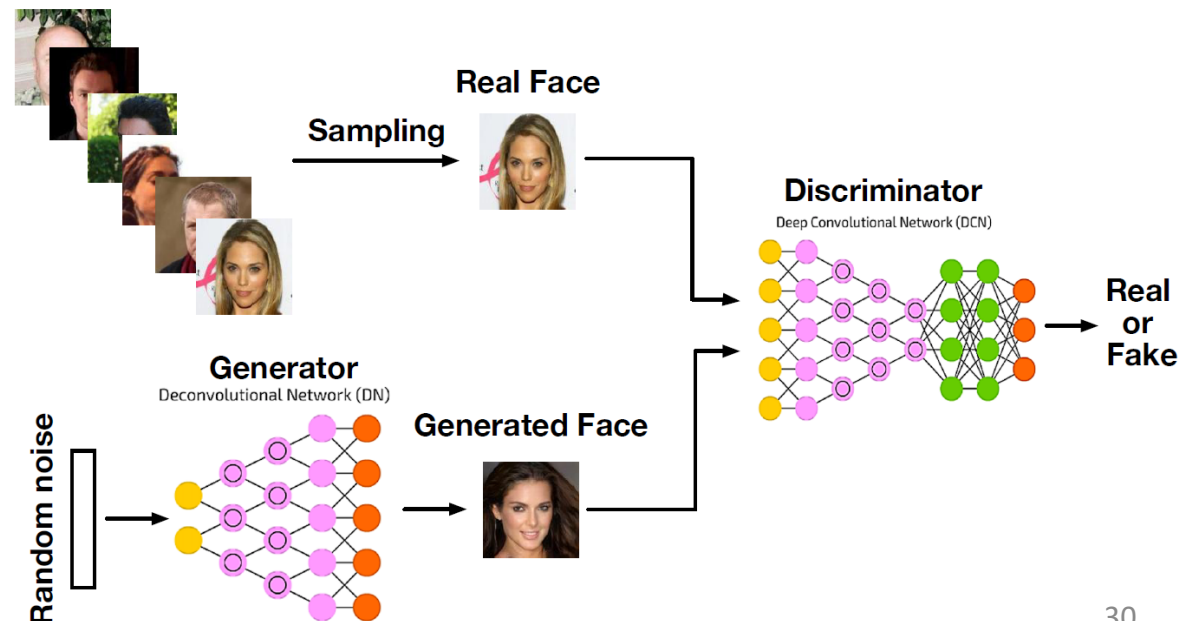
$f(x, w')$ — модель классификации x «реальный/сгенерированный»

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

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

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

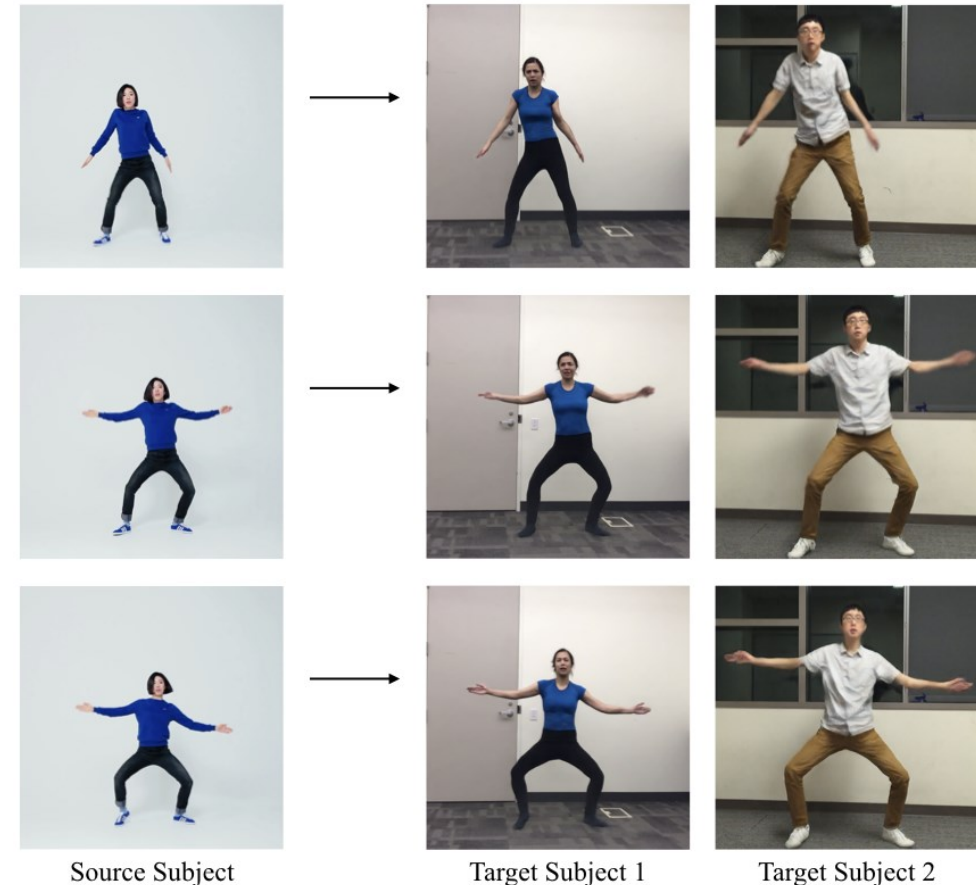
Chris Nicholson. A Beginner's Guide to Generative Adversarial Networks. 2019.



Синтез изображений и видео



(d) input image (e) output 3d face (f) textured 3d face



Содержание

1. Задачи машинного обучения

- Предыстория машинного обучения
- Терминология машинного обучения
- Примеры задач машинного обучения

2. Методология машинного обучения

- Нейронные сети и глубокое обучение
- Оптимизационные задачи машинного обучения
- Задачи машинного обучения с векторизацией объектов

3. Большие языковые модели

- Модели внимания и трансформеры, эмерджентность
- О тексто-графическом представлении знаний
- Мифы об искусственном интеллекте

Эволюция подходов в обработке текстов

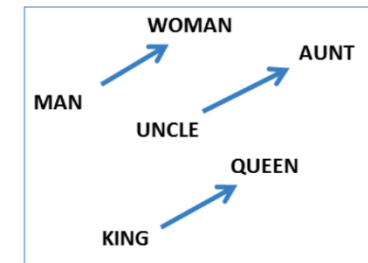
Декомпозиция задач по уровням «пирамиды NLP»

- 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], ...

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

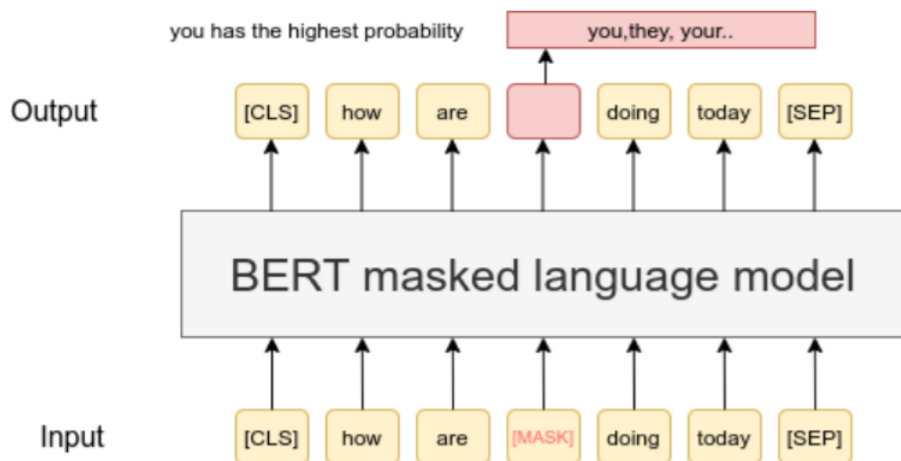
Обучение контекстной векторизации слов

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 в i -й позиции по его контексту (Masked Language Model)

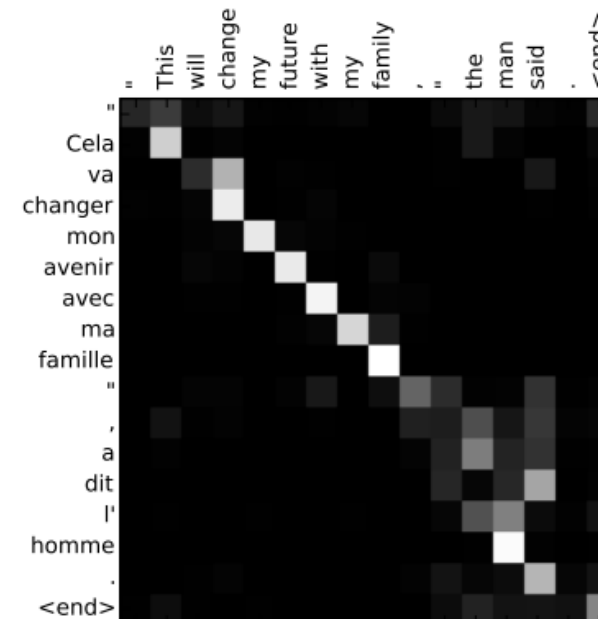
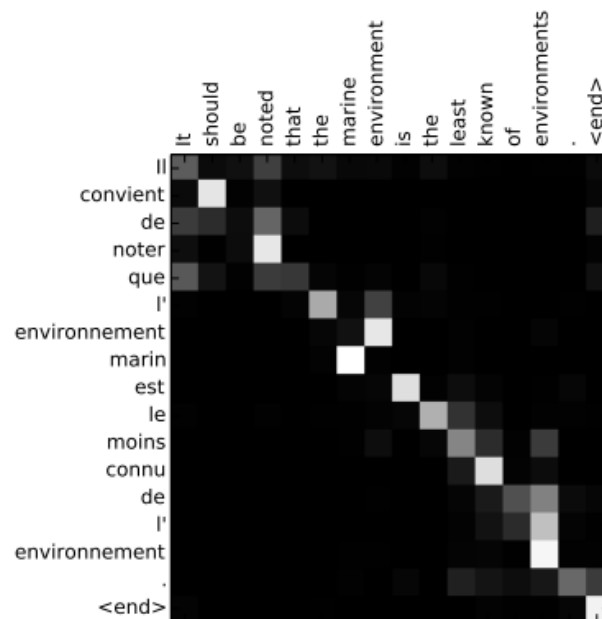
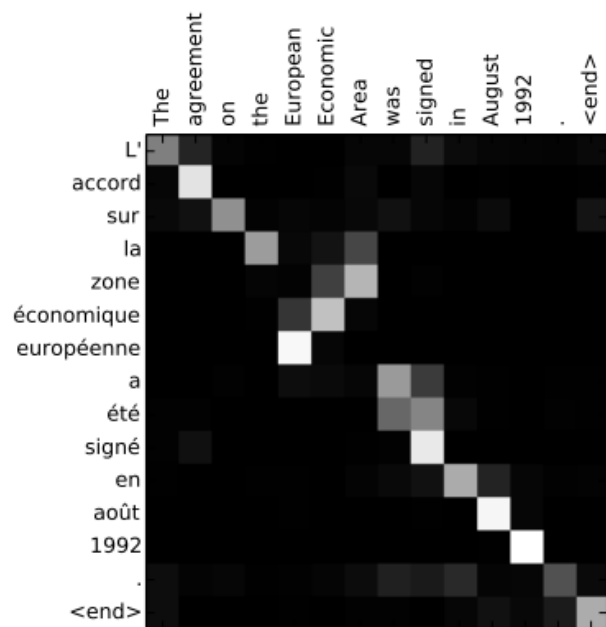


Vaswani et al. (Google) Attention is all you need. 2017.

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.

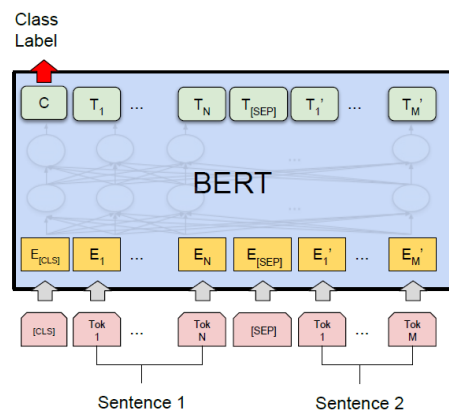


A giraffe standing in a forest with trees in the background.

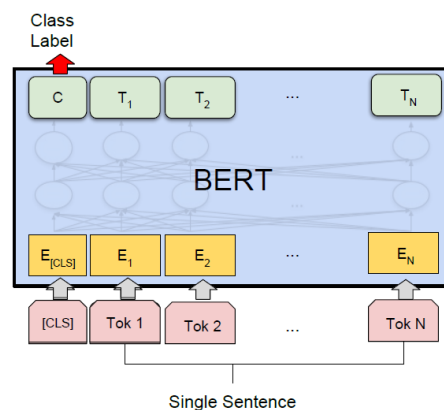
Интерпретация: на какие области модель обращает внимание, генерируя подчёркнутое слово в описании изображения

Трансформеры: нейросетевые модели языка

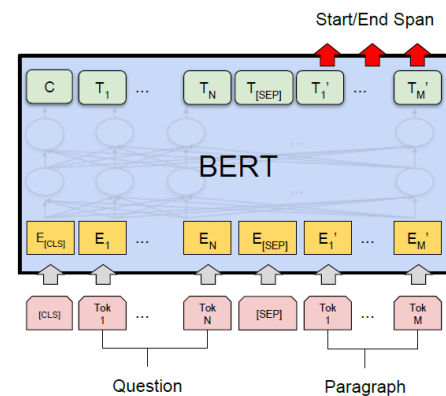
- Обучаются векторизовать и предсказывать слова по контексту
- Обучаются по терабайтам текстов, «они видели в языке всё»
- Мультиязычны: обучаются на десятках языков
- Мультизадачны: для каждой новой задачи NLP/NLU достаточно предобученной модели или дообучения на небольшой выборке



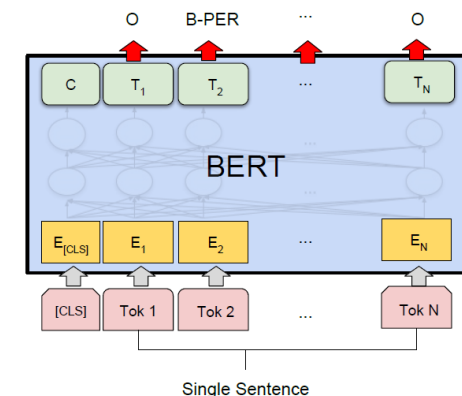
(a) Sentence Pair Classification Tasks:
MNLI, QQP, QNLI, STS-B, MRPC,
RTE, SWAG



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



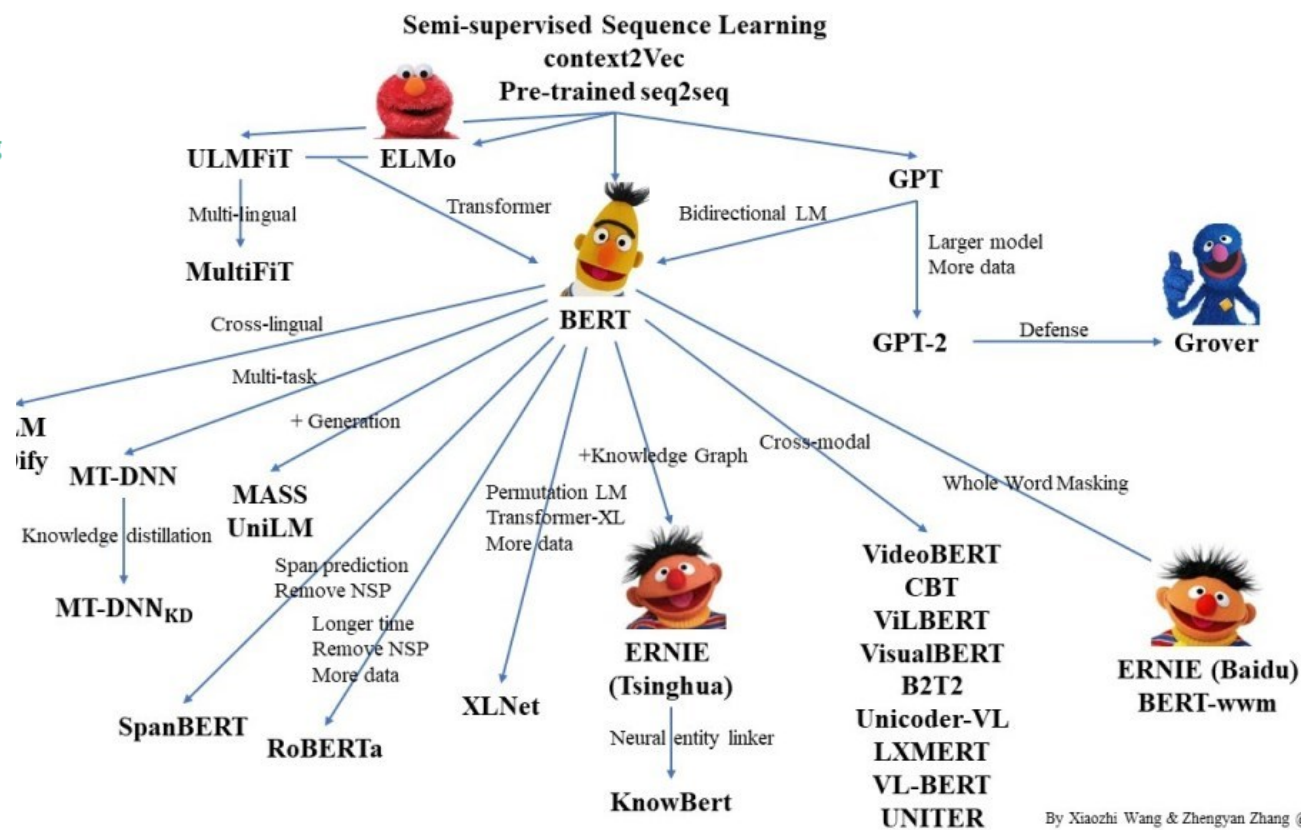
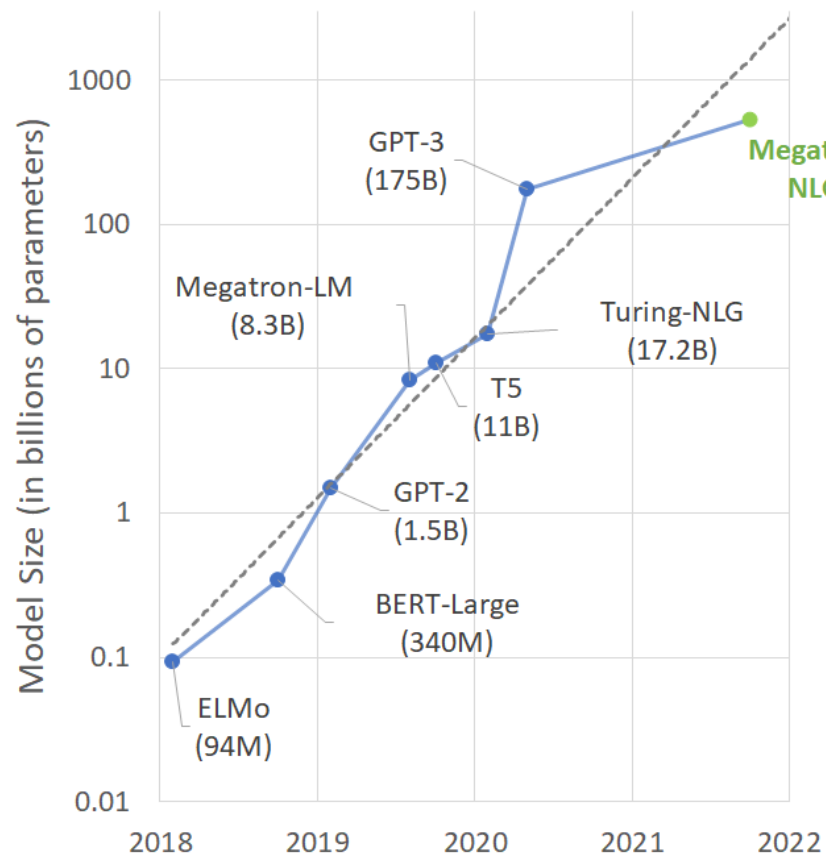
(c) Question Answering Tasks:
SQuAD v1.1



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

Трансформеры: нейросетевые модели языка

Рост числа параметров нейросетевых трансформерных моделей языка



By Xiaozhi Wang & Zhengyan Zhang @THUNLP

Проблески общего искусственного интеллекта

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

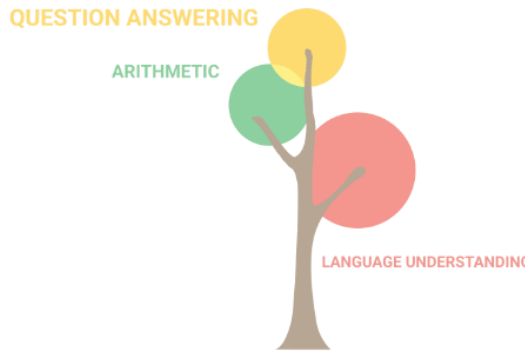
Sébastien Bubeck Varun Chandrasekaran Ronen Eldan Johannes Gehrke
Eric Horvitz Ece Kamar Peter Lee Yin Tat Lee Yuezhi Li Scott Lundberg
Harsha Nori Hamid Palangi Marco Tulio Ribeiro Yi Zhang

Microsoft Research (27 March 2023)

Новые способности модели, не закладывавшиеся при 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

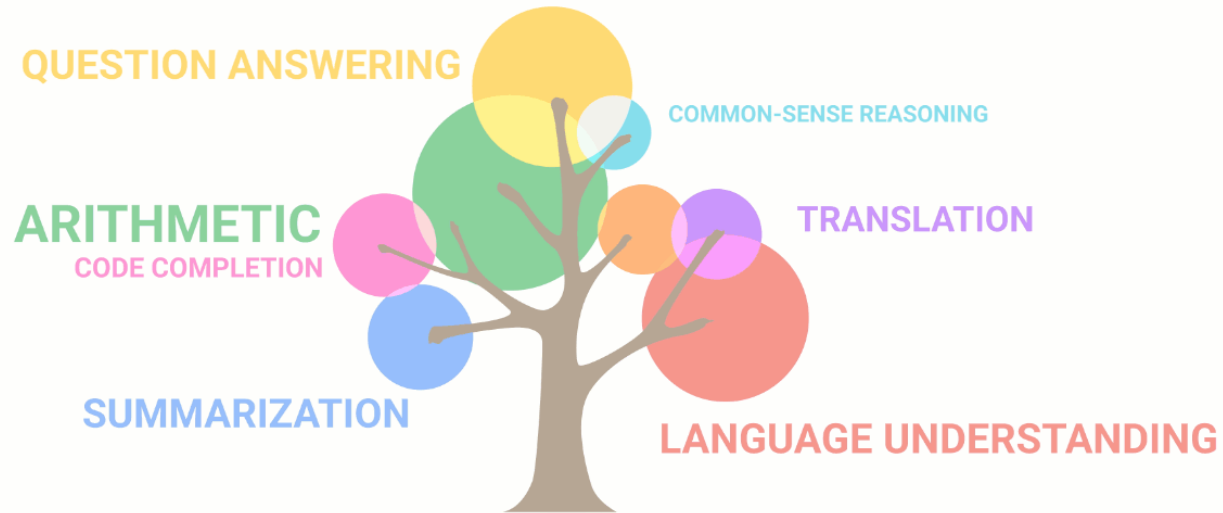


GPT-2: 14-Feb-2019

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

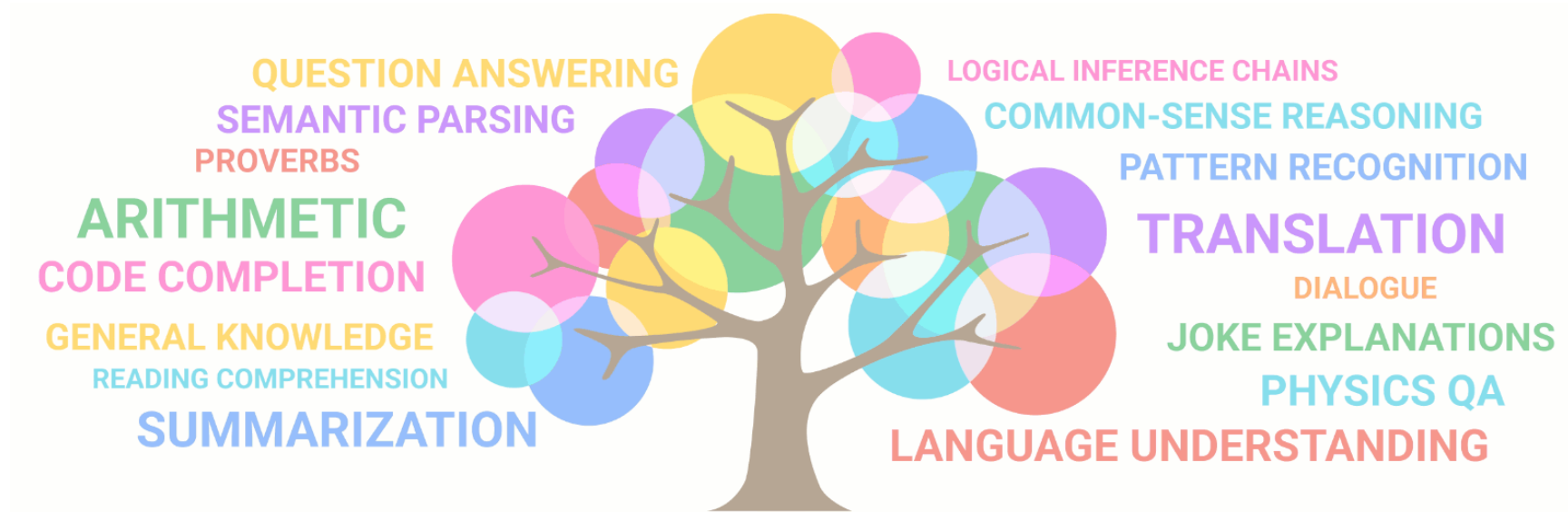


GPT-3: 11-Jun-2020

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

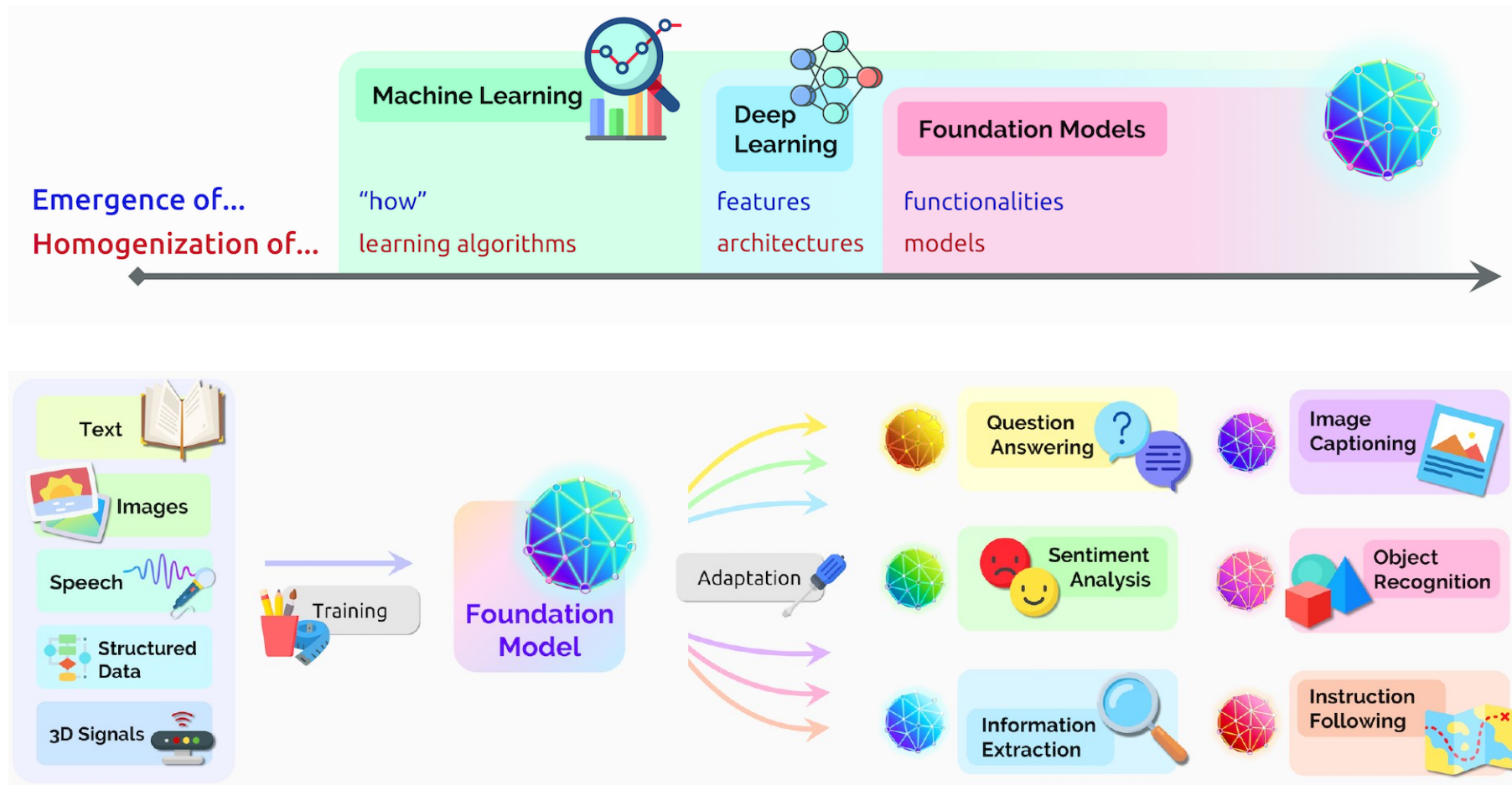


GPT-4: 14-Mar-2023

>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

Фундаментальные модели (Foundation Models)



Возможности и угрозы

Чаты GPT уже способны помогать с рутинно-творческой работой:

- генерировать документы или сайты по техническому заданию
- в том числе медицинские, юридические документы по шаблонам
- искать и структурировать профессиональную информацию
- делать обзоры, рефераты, сводки на разных языках
- 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

Возможности и угрозы

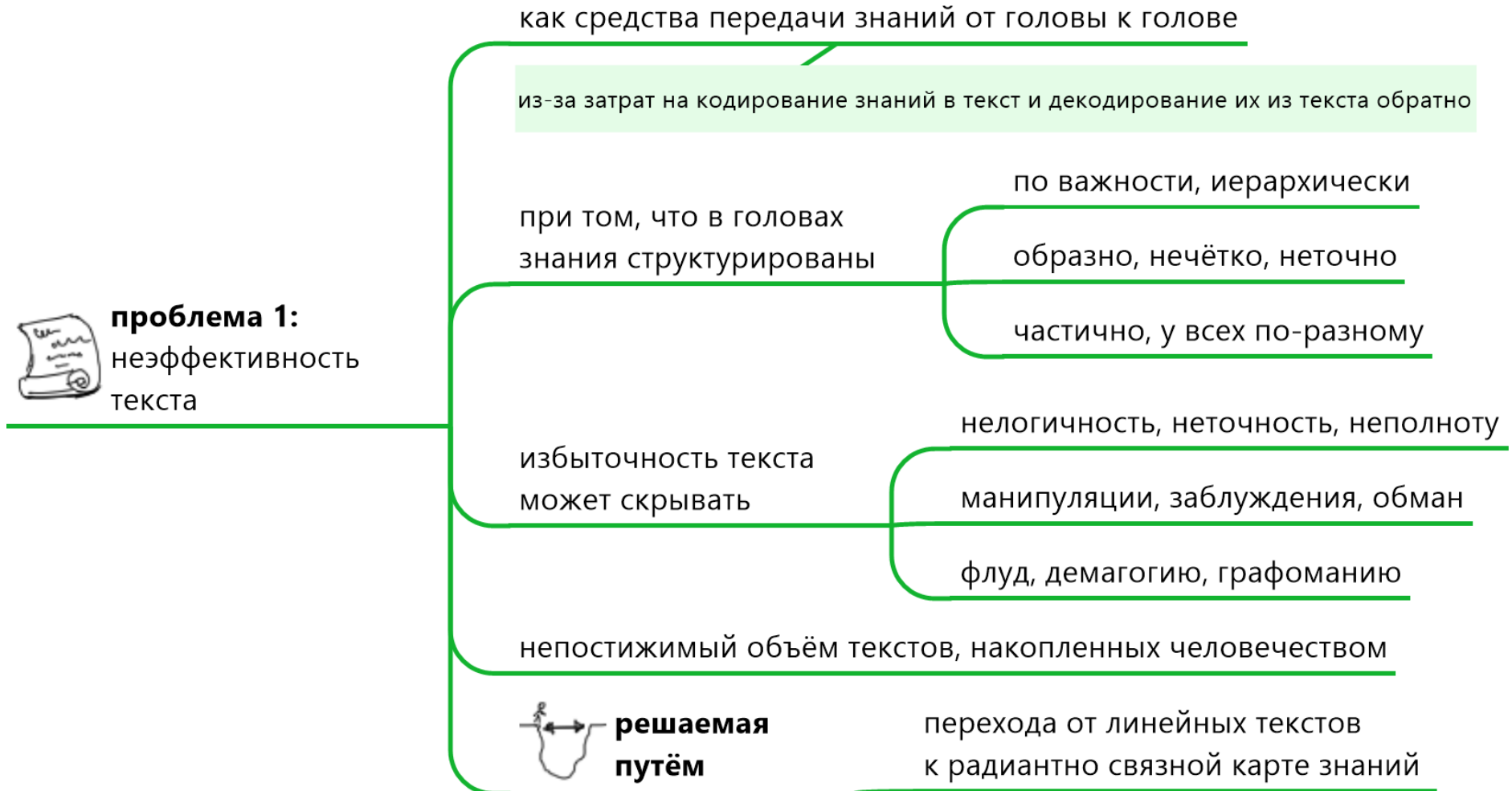
Чаты GPT способны (непреднамеренно, не обладая автономностью):

- «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

Перспектива развития ИИ через большие языковые модели

- Устранение ошибок и галлюцинаций, выравнивание (alignment)
- Интеграция с поиском в надёжных источниках, проверкой фактов
- Интеграция с отраслевыми решателями
- Мультимодальные фундаментальные модели (foundation models)
- Навыки логических рассуждений (решение математических задач)
- **Тексто-графическое представление знаний (от mind-map к картам знаний)**
- *Сократические модели*: несколько пред-обученных LLM договариваются на человеческом языке о выполнении новых мультимодальных задач в незнакомой среде (от ML к подлинному AI).

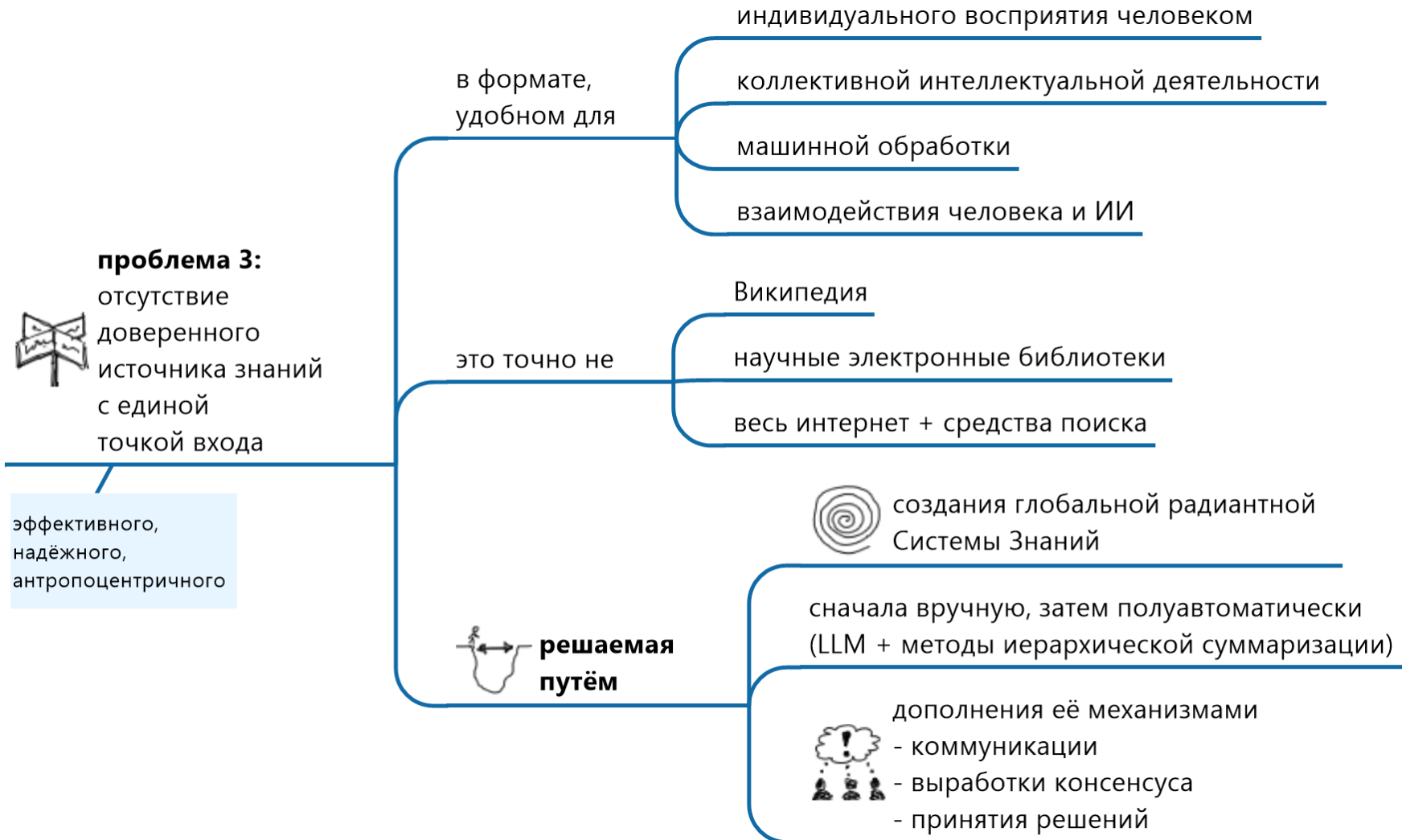
Перспективная задача развития ИИ: устранять барьеры между человеком и знанием



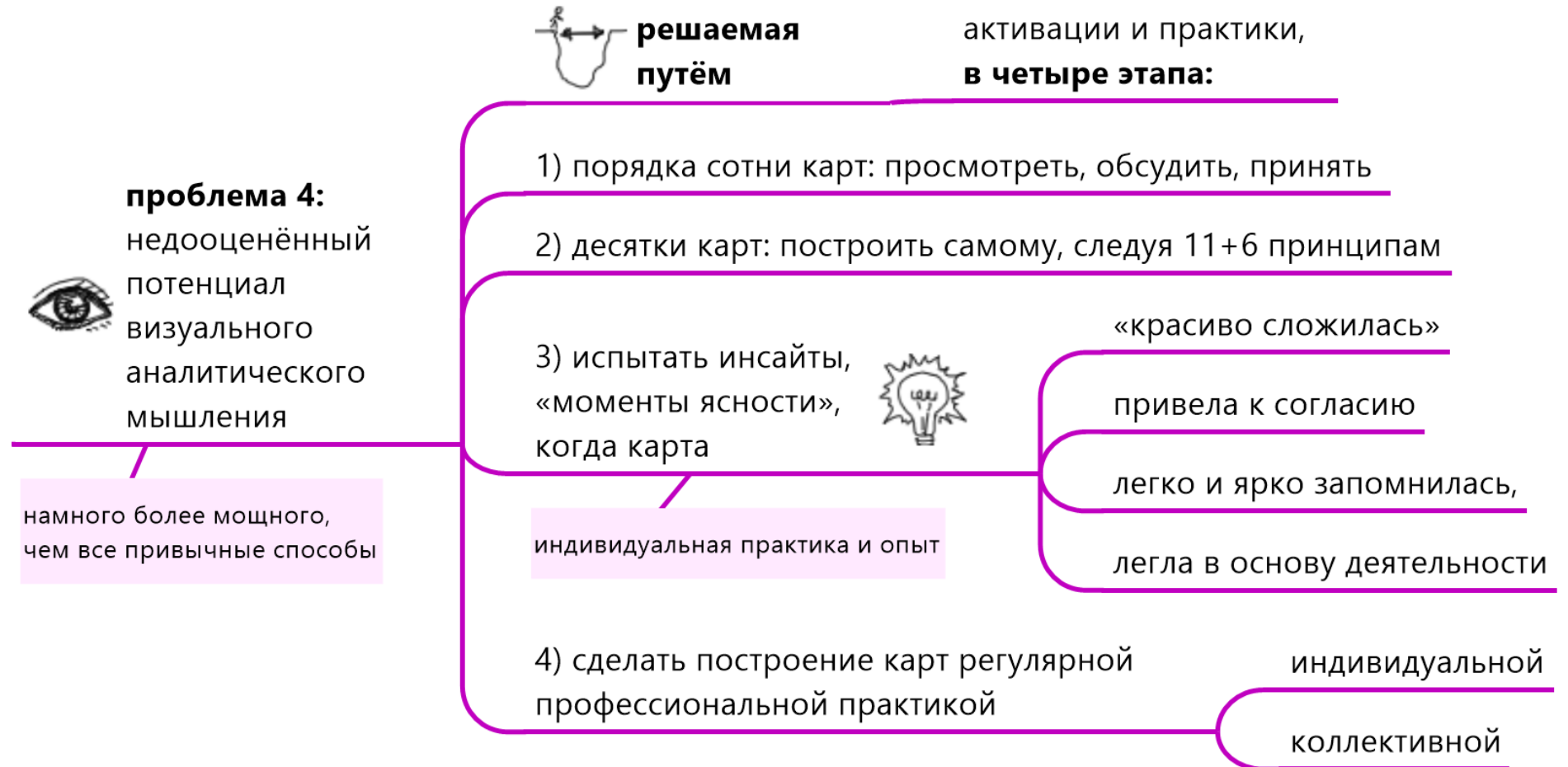
Перспективная задача развития ИИ: устранять барьеры между человеком и знанием



Перспективная задача развития ИИ: устранять барьеры между человеком и знанием



Перспективная задача развития ИИ: устранять барьеры между человеком и знанием

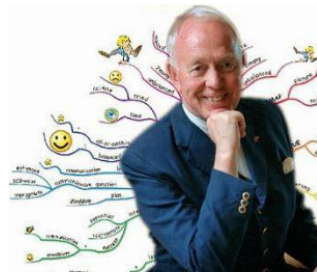


Карты знаний: базовые принципы интеллект-карт (mind map)



способ визуализации того, как темы (мысли, идеи) разбиваются на подтемы иерархически

предложены
в 70-е годы
британским
психологом
Тони Бьюзеном



графическое
оформление

активация зрительной памяти

радиантность: линии
расходятся из центра

размер шрифта
отражает важность

цвет
выделяет поддеревья

картинки
усиливают образность



дополнительные
элементы

ассоциативные связи между темами

комментарии, выноски, теги, (гипер)ссылки



техника
запоминания

посмотреть, понять, обсудить, принять

самостоятельно воспроизвести через
10 минут → сутки → неделю → месяц

Карты знаний: +11 известных принципов



Карты знаний: +6 новых принципов

всех карт через ключевые понятия в единую **Систему Знаний**



(1) глобальная
радиантная
связность

компромисс с обозримостью



в центре находится
смысловое ядро

естественно-научное, цивилизационное
знания, которые важны всегда и для всех

критерии важности тем:
что в теме главное?

для чего?

для кого?

метафора:

источник силовых линий, по которым
ранжируется семантическое поле карты



(2) отторгаемость

компромисс с лаконичностью

комментарии автора не обязательны для понимания карты

карта способна «жить своей жизнью»

Карты знаний: +6 новых принципов

(3) коллективность,
на всех этапах
жизненного цикла

компромиссы между авторами

создание

рецензирование, согласование

развитие

уточнение, реструктуризация

детализация, разрастание

применение

в практической деятельности

с разграничением прав доступа

любой фрагмент карты читается
как связный текст, нарратив

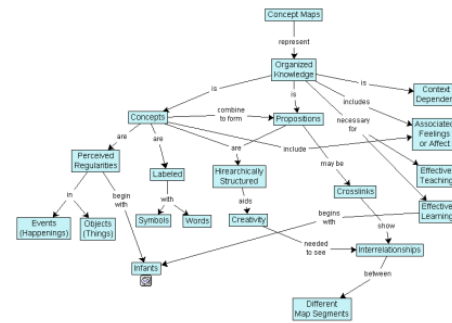
легко и однозначно

даже автоматически

(4) читабельность

компромисс с лаконичностью
и обзорностью

в отличие
от других техник
представления
знаний



онтологий

фреймов и др.

Карты знаний: +6 новых принципов



(5) сворачиваемость

компромисс с читабельностью

любой темы без утраты

читабельности

сбалансированности

позволяет «отложить на потом» любую детализацию

способствует

выделению главного в каждой теме

пониманию и взаимопониманию

на этапе
создания
карт:

подбор

источников, ссылок

картинок по контексту

суммаризация текстов в виде карт знаний

на этапе чтения:
автоматическое

сворачивание карты по слайдам

преобразование карты в нарратив

обучение по картам знаний
больших языковых моделей

думающих как люди

безопасных для людей



(6) возможность
машинной обработки

компромисс с антропоцентричностью

Выводы про карты знаний

- **Универсальный инструмент мышления** для человека и машины.
- **Главные полезные навыки** — во всём выделять главное (7 ± 2),
 - делать это быстро, формулировать лаконично.
- **Прежде чем обучать ИИ** по тексто-графическим представлениям,
 - необходимо освоить их самим,
 - в своей практической деятельности,
 - в том числе (и даже в первую очередь) — в коллективной.

Мифы №1 ... №8

«В будущем Искусственный Интеллект...

... лишит людей работы»

... будет использован для узурпации власти над миром»

... приведёт людей к праздности и деградации»

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

... станет автономным, 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 они хотят 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

Выводы (технологические): о составных частях успеха ИИ

- принцип эмпирической индукции Фрэнсиса Бэкона
- минимизация (и аппроксимация) эмпирического риска
- регуляризация некорректно поставленных задач
- коннекционизм и глубокие нейросетевые архитектуры
- векторизация сложно структурированных данных
- самостоятельное обучение вместо обучения по разметке
- увеличение скорости и параллелизма вычислителей

Выводы (гуманитарные): как относиться к ИИ и его развитию

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

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/ZR6ZeK5B3lL6OxXv>
- *Воронцов К. В.* Лекции по машинному обучению. www.MachineLearning.ru, 2004-2024.
- *Кевин Мэрфи.* Вероятностное машинное обучение. Введение. 2022.
- *Шумский С. А.* Машинный интеллект. РИОР ИНФРА-М, 2020.