

Лекция 01

Введение в машинное обучение

Кирилл Валерьевич Святов

Ульяновский Государственный

Технический Университет

k.svyatov@ulstu.ru

План

Кратко об истории технологических
революций

Технологии будущего

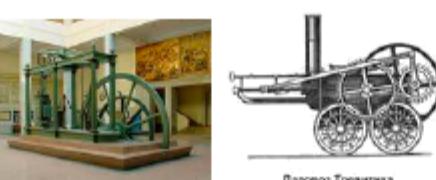
Определения

Формулировка проблем Определений

Приложения технологии ML

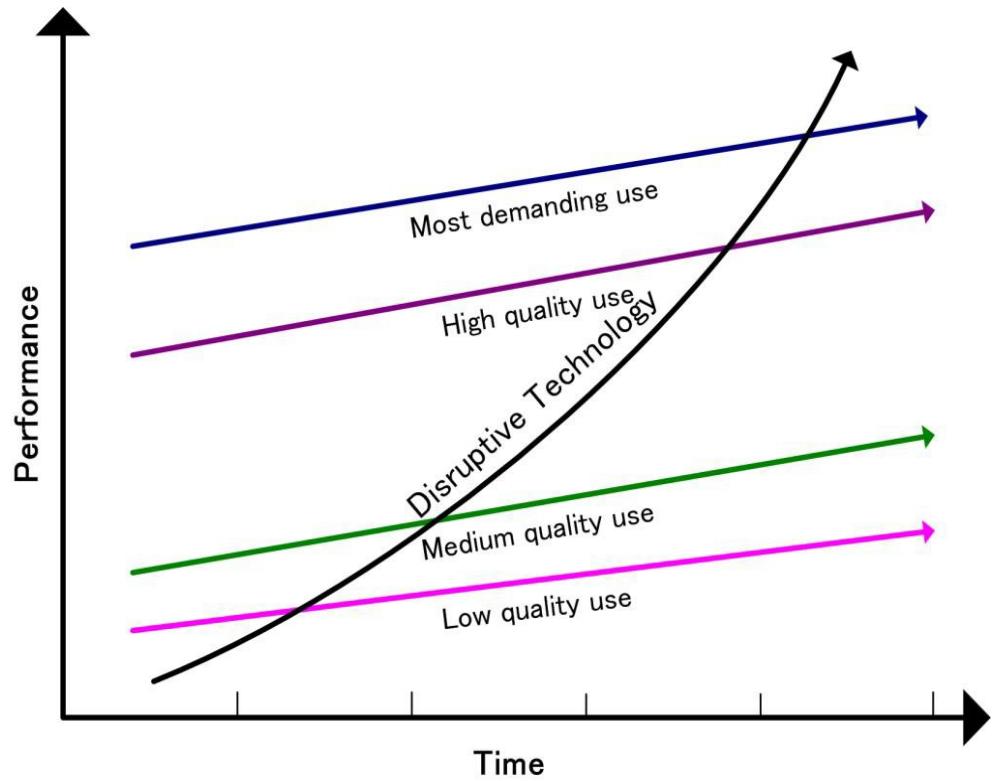
Методы ML

Промышленные революции

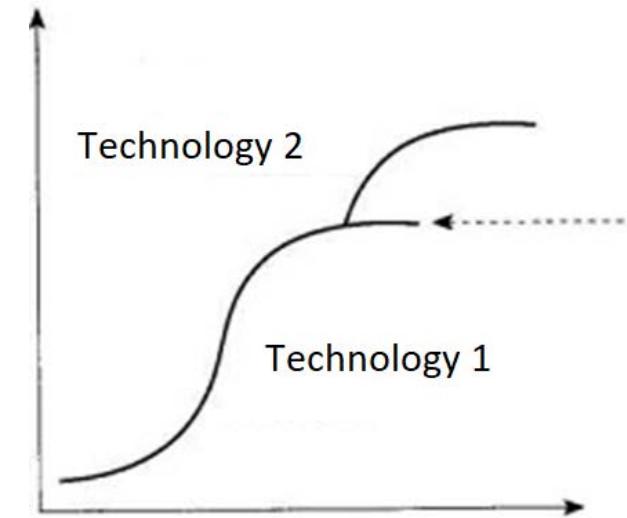
«I» IR XVII century	II IR XVIII – XIX	III IR XIX – XX	IV IR XXI
<p>Technologies</p> <ul style="list-style-type: none">▶ wood▶ peat burn▶ wind energy 	<ul style="list-style-type: none">▶ Iron, steel▶ coal burn▶ steam▶ surgery, anesthesia▶ agriculture machines 	<ul style="list-style-type: none">▶ plastic▶ oil, electricity▶ cars, airplains▶ satellites▶ antibiotics, early diagnosis▶ mineral fertilizers 	<ul style="list-style-type: none">▶ 3D printing & materials▶ thin films▶ clean energy▶ robots▶ small satellites▶ organic agriculture 
<i>Netherlands</i>	<i>England</i>	<i>USA</i>	

¹ <http://www.tv2.tomsk.ru/sites/www.tv2.tomsk.ru/files/mobil.jpg>

Подрывные технологии

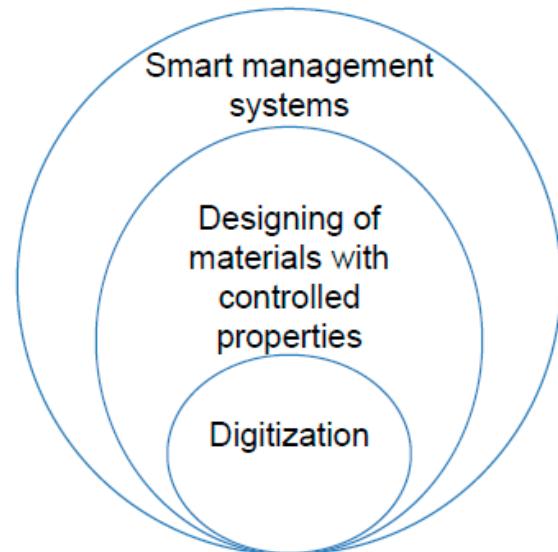


Clayton Cristensen



Промышленность 4.0

Corridors for the development of technology and activities:



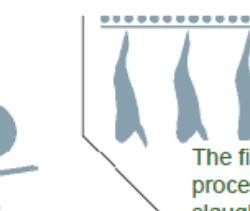
4 stages of industrial revolution

Industrie 1.0 follows the introduction of water and steam mechanical manufacturing enterprises



Mechanical loom, 1784

Industrie 2.0 comes as a result of the introduction of electrified mass production, based on the division of labor



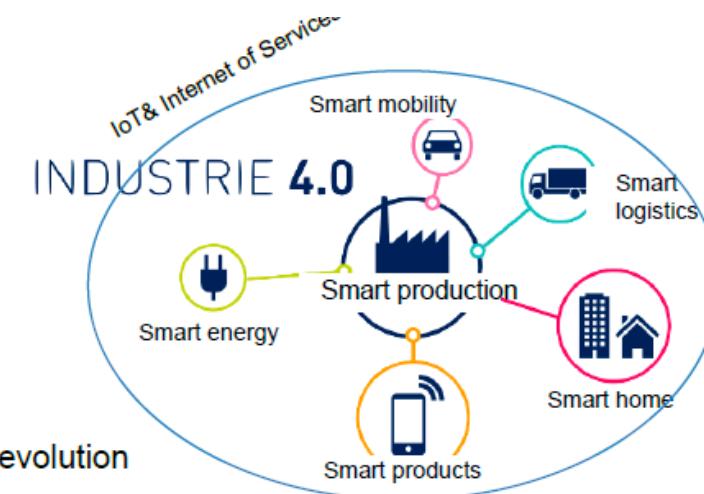
The first processing line, slaughterhouse 1870

Industrie 3.0 uses electronics and information technology for further automation of production



First PLC, 1969

Industrie 4.0 based on cyber physical systems



Производительность труда

Assessment of the growth of labor productivity from the development of new technologies in production

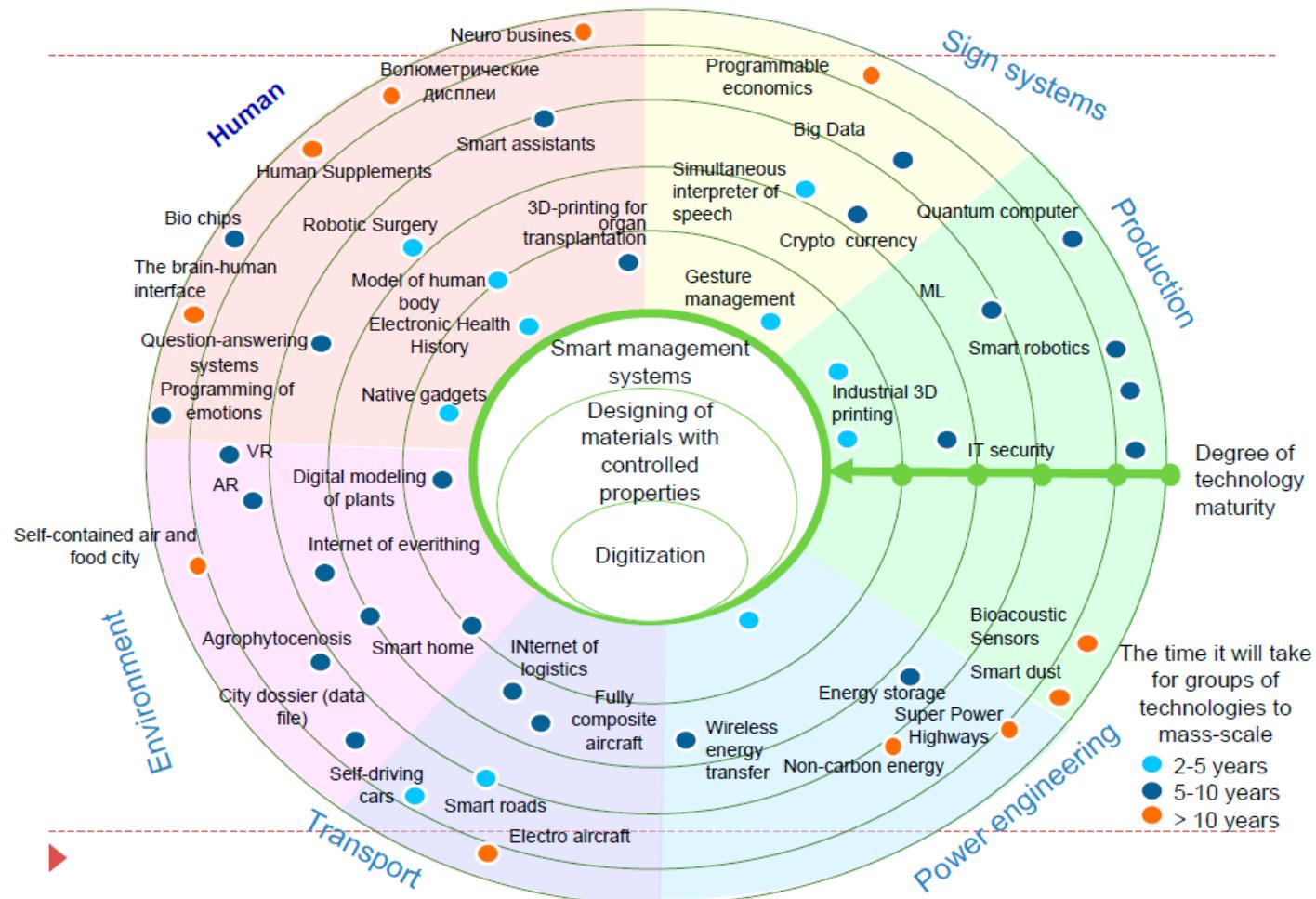


Steam engine	Early robotics	IT	Robots, AI, ML
1850-1910	1993-2007	1995-2005	2015-2065
Growth in labor productivity per year			
0,3%	0,4%	0,6%	0,8 - 1,4%

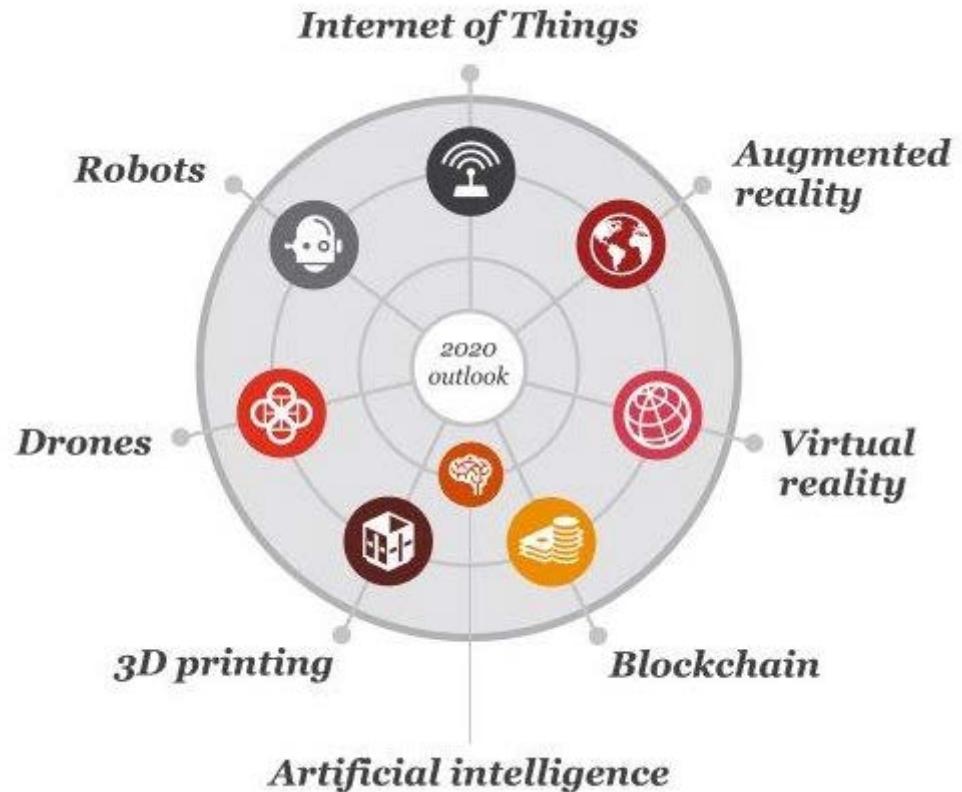
¹ McKinsey & Company, <http://www.mckinsey.com/global-themes/digital-disruption/harnessing-automation-for-a-future-that-works>

Современные технологии

Today there is a new platform of technologies of the New Industrial Revolution



8 ключевых технологий

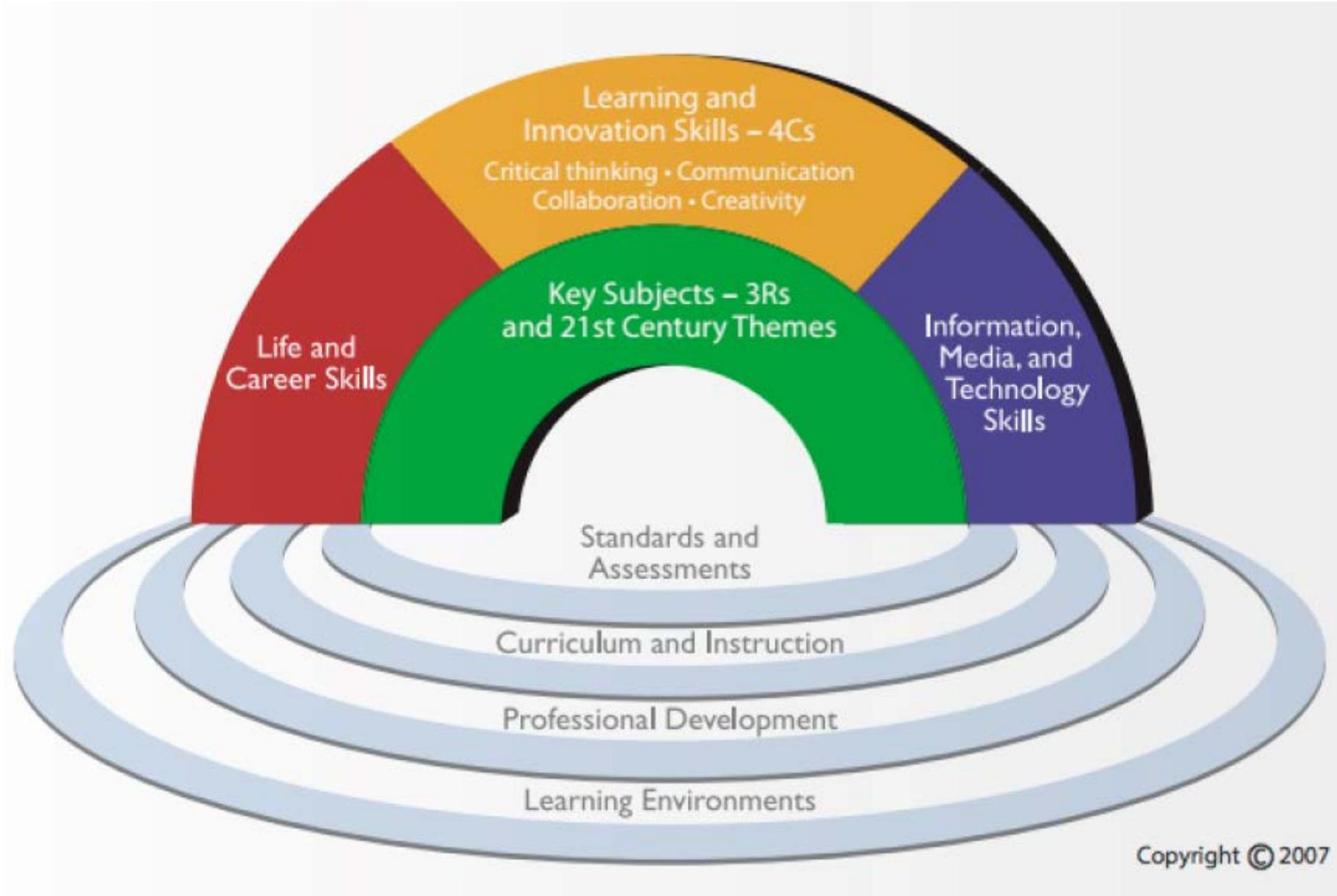


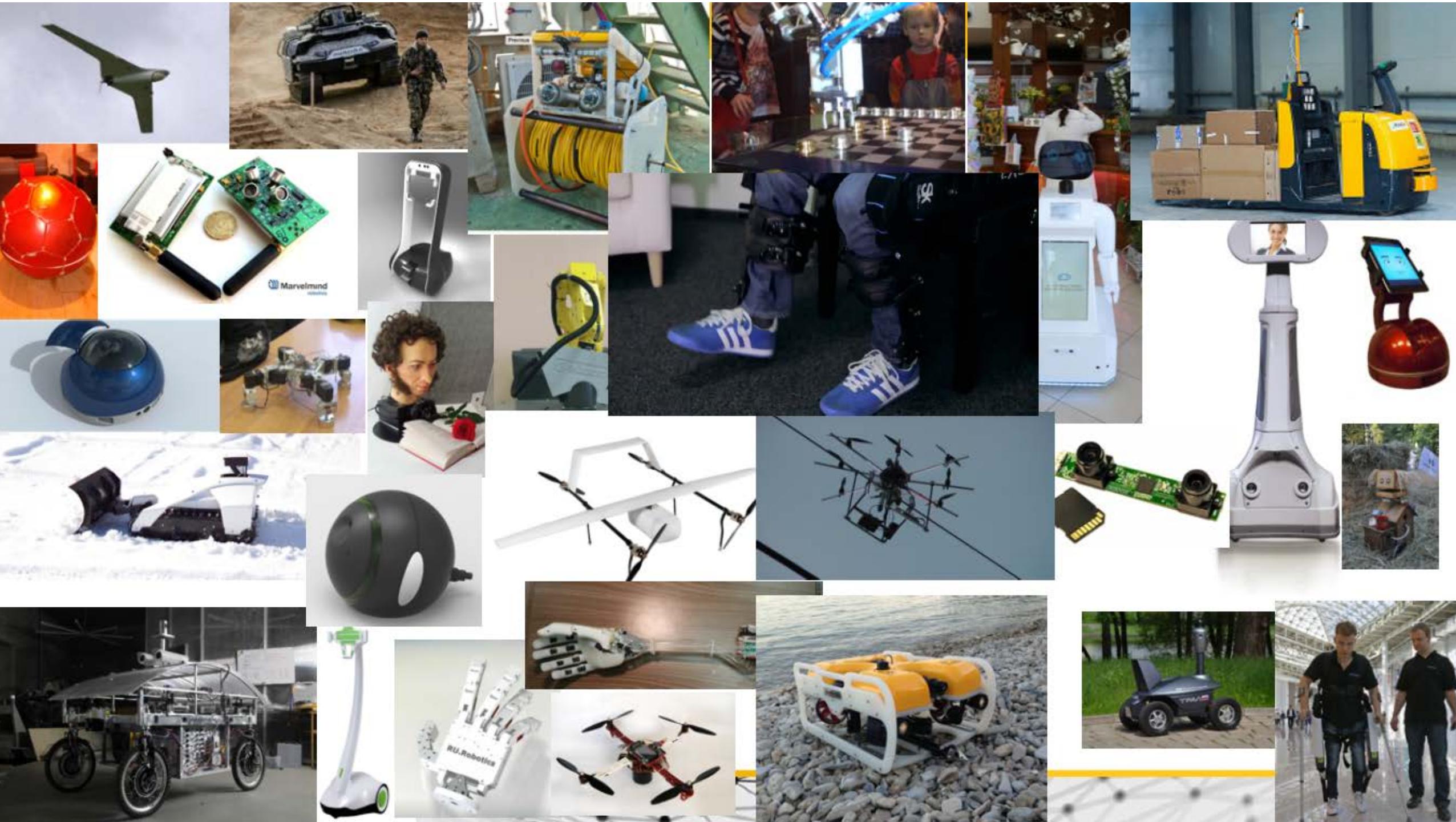
<https://pwc.blogs.com/ceoinsights/2016/08/a-guide-to-the-essential-eight-emerging-technologies.html>

Проблемы трансформации

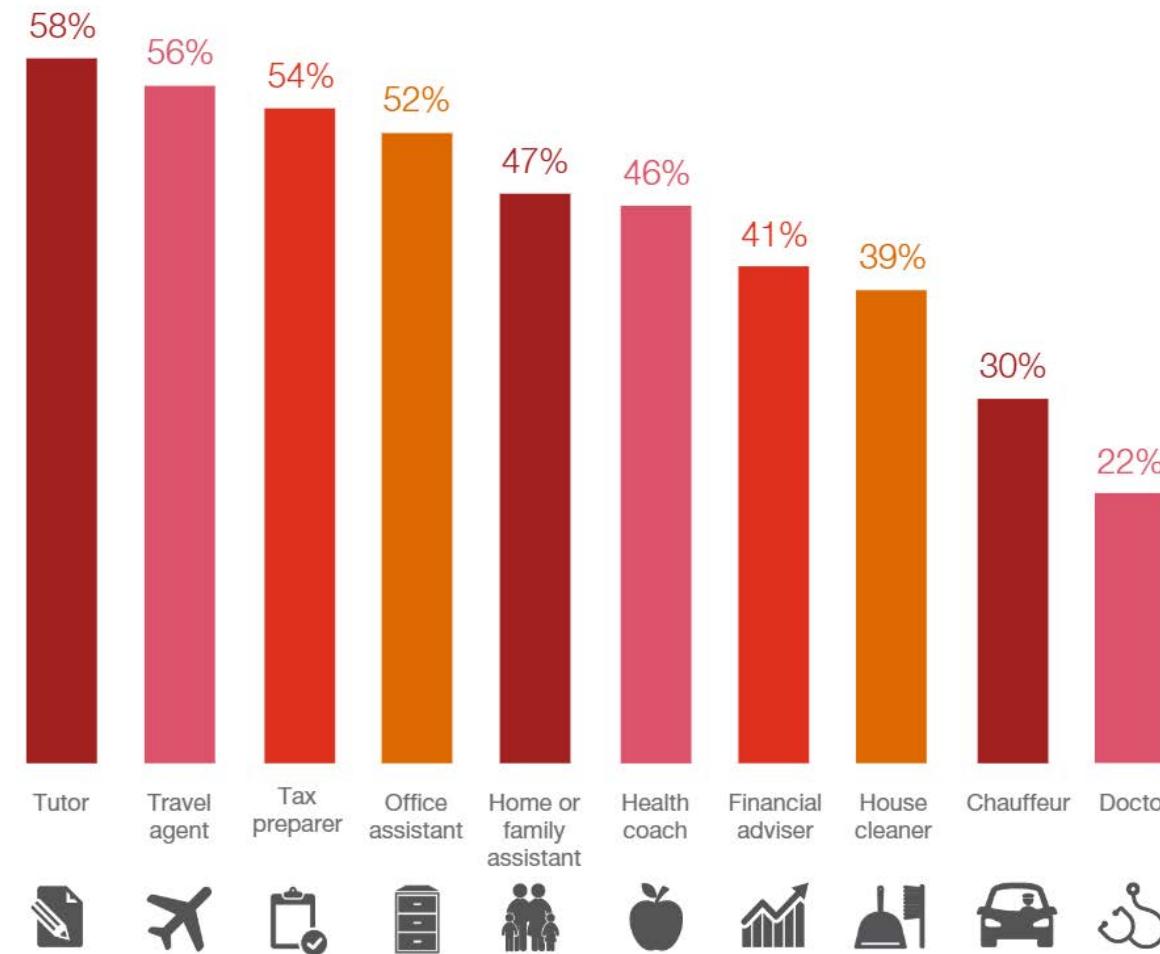


Важные навыки





"Я вижу, что в течение следующих пяти лет ИИ заменит людей в качестве..."



зима ИИ & весна

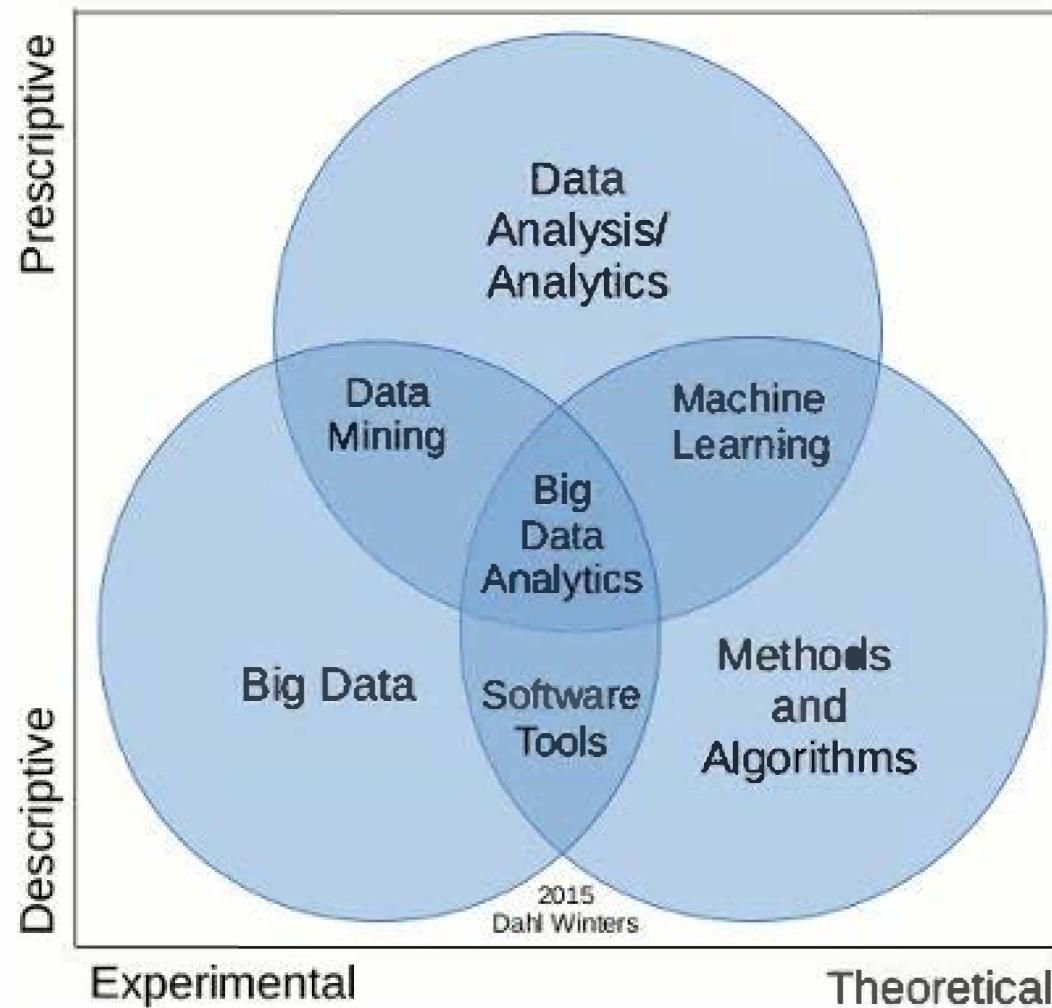


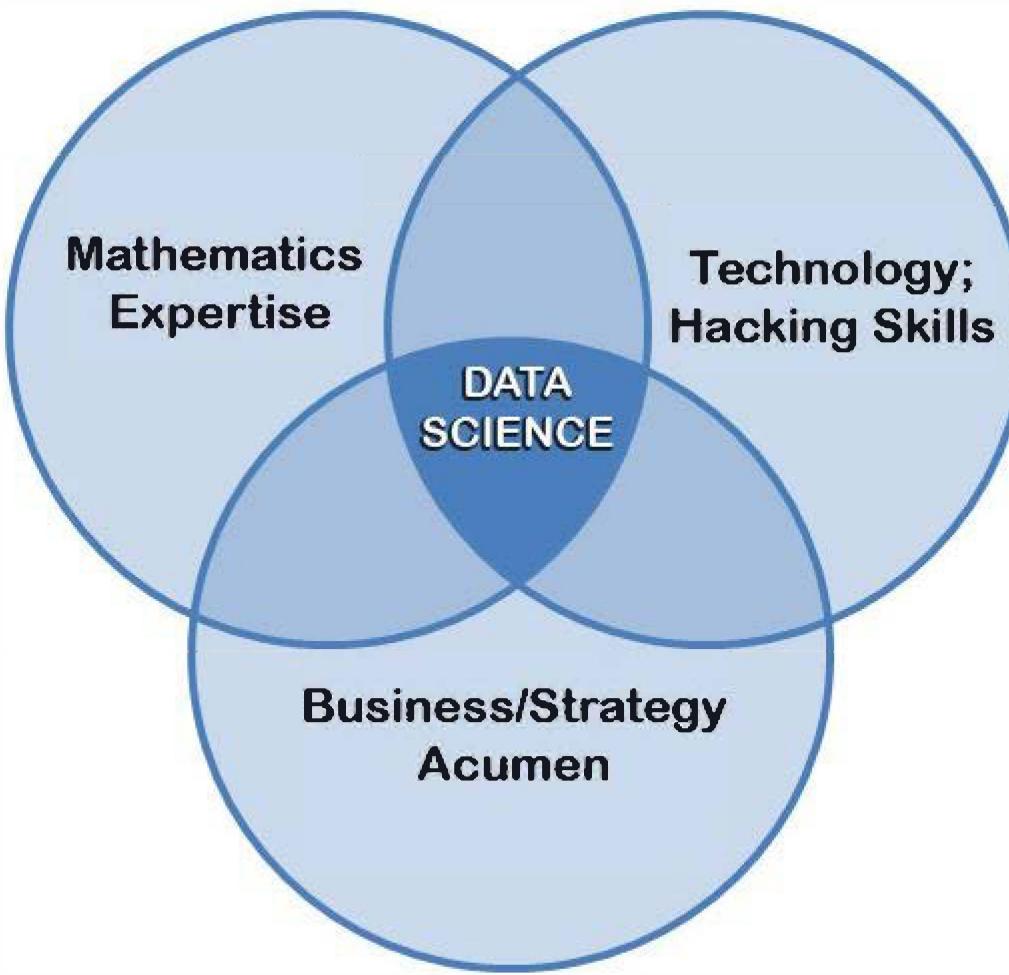
Определение ML

Machine Learning:

- Machine learning - это:
 - раздел искусственного интеллекта, математическая дисциплина, которая использует математическую статистику, численные методы оптимизации, теорию вероятности , дискретный анализ, для извлечения сведений из данных.
 - Область изучения, которая предоставляет компьютерам возможность обучаться без предварительного явного программирования.
-

The Fields of Data Science





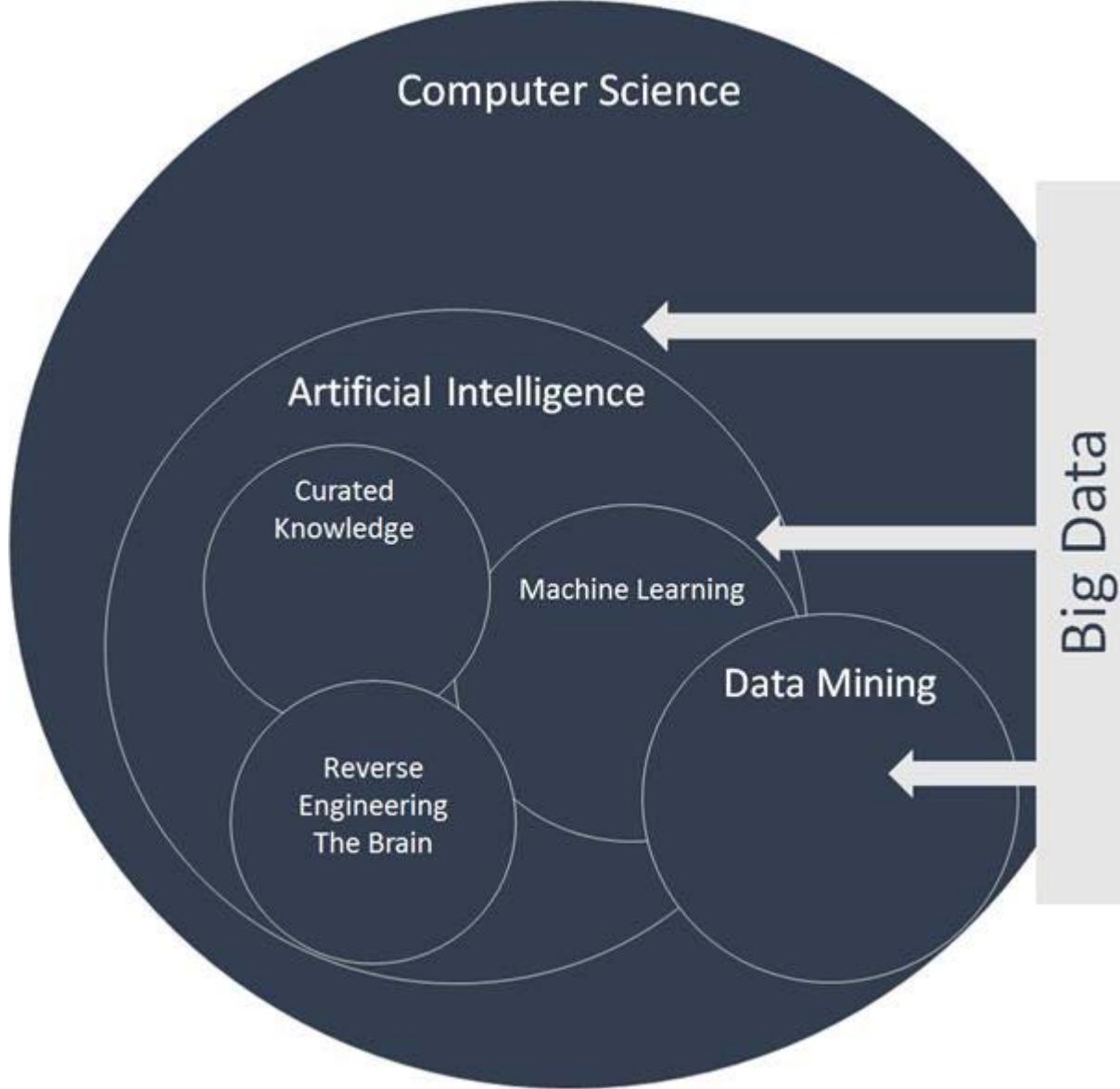
Ключевые особенности извлечения данных

- Данные могут быть неточными, неполными, гетерогенными, непрямыми и в то же время иметь большие объемы;
- Алгоритмы анализа данных сами по себе могут иметь возможность обучаться через прецеденты;
- Процессы обработки "сырых" данных в информацию требуют нетривиальной автоматизации.

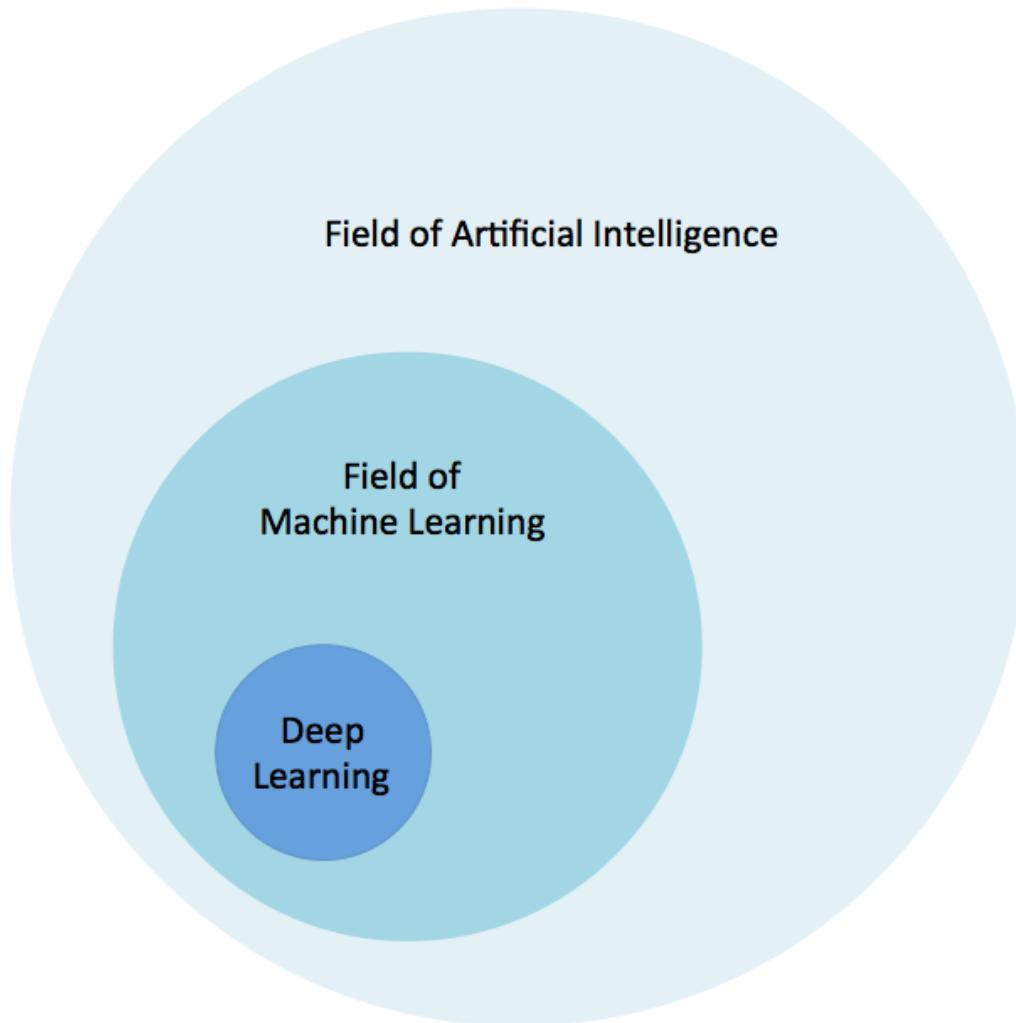
Ключевые особенности извлечения данных

По прошествии последних десятилетий, значительные попытки в области извлечения данных сфокусировались на создании специальных алгоритмов, способных к выполнению схожих задач во в течение линейного или даже логарифмического времени без серьезных потерь в точности.

Ключевые особенности извлечения данных



Ключевые особенности извлечения данных



Ключевые особенности извлечения данных

Компьютерная программа обучается на основе опыта E по отношению к некоторому классу задач T и меры качества R , если качество решения задач из T , измеренное на основе R , улучшается с приобретением опыта E .

(T. Mitchell)

Ключевые особенности извлечения данных

Выделяют два типа обучения:

- Обучение по прецедентам
- Дедуктивное обучение

Ключевые особенности извлечения данных

Цель обучения по прецедентам заключается в обобщении или получении знания о "законе природы"

Ключевые особенности извлечения данных

Aa

ABCDEFGHIJKLMNOPQRSTUVWXYZ 01234
abcdefghijklmnopqrstuvwxyz 56789

Ключевые особенности извлечения данных

SPHINX OF BLACK QUARTZ JUDGE MY VOW

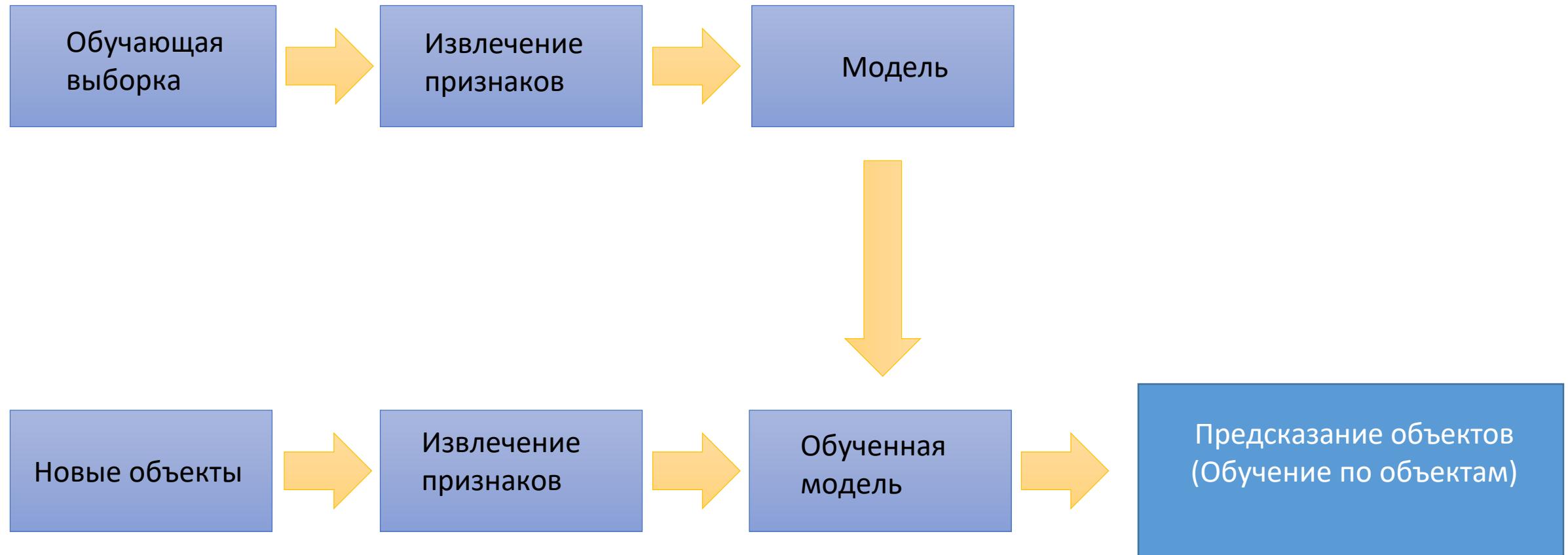
Sphinx of black quartz judge my vow

A B C D E F G H a b c d e f g h

0 1 2 3 4 5 6 7 8 9 / @ # \$ %

• глаза и оправа лучше. Но
ищем у глазок замысл
и они имеют особенности
личных предметов и вакансий
быть обновлена и куплен
запросов. Помимо обычном
то законы природы и

Ключевые особенности извлечения данных



Определение задачи ML

Machine learning (машинное обучение):

- Что есть объект X (каковы признаки) и что есть ответ Y ?
- Как построить модель M ?
- Каким образом мы можем выполнить аппроксимацию (приближение) Y от X с M ?

Определение задачи ML

Типы входных данных:

- Изображение
- Текст
- Звук
- Геоданные
- Временные ряды
- Техническая диагностика, спецификации

Обобщение

Ошибка обобщения (также известные как ошибка вне-образца) является мерой того, насколько точно алгоритм способен предсказать исход для значений ранее невидимых данных.

В задачах обучения, цель состоит в том, чтобы разработать функцию , которая предсказывает выходные значения на основе некоторых входных данных . Ожидаемая ошибка конкретной функции по всем возможным значениям и является: $f(x)yxI[f_n]f_nxy$

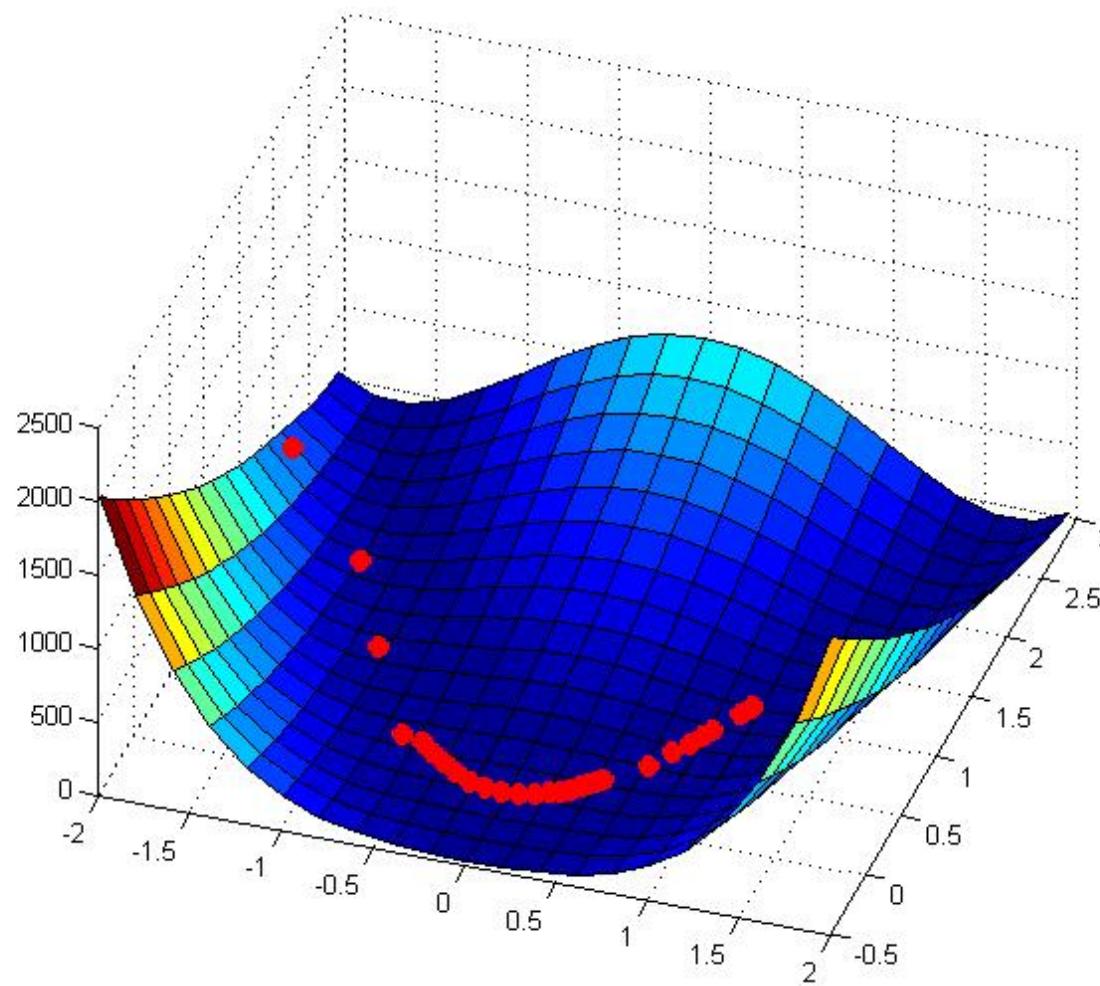
$$I[f_n] = \int_{X \times Y} V(f_n(x), y) \rho(x, y) dx dy,$$

где обозначает функцию потерь и является неизвестным совместным распределением вероятностей для и . V
 $\rho(x, y)xy$

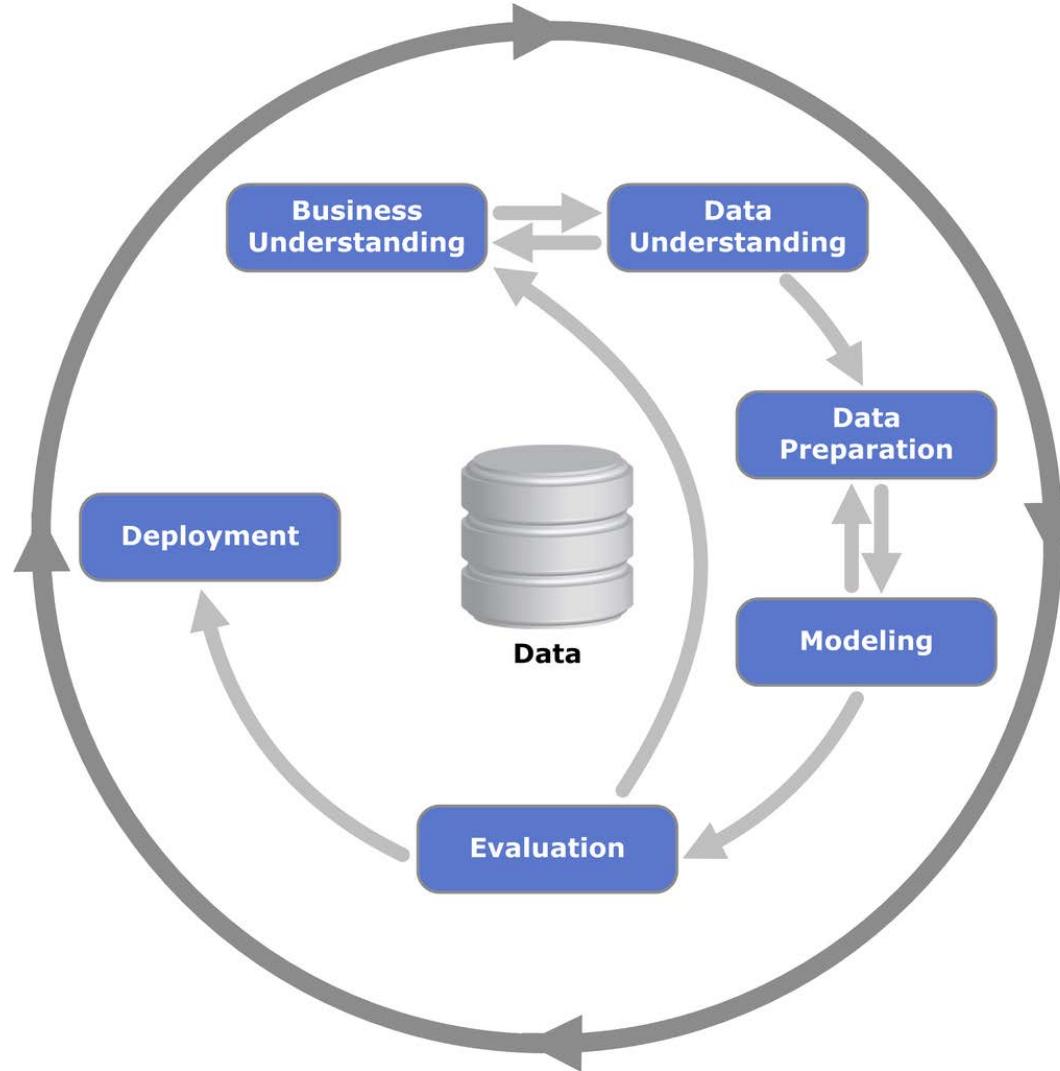
Не зная совместное распределение вероятностей, невозможно вычислить . Вместо этого, мы можем вычислить эмпирическую ошибку на выборочных данных. С учетом точек данных, эмпирическая ошибка: $I[f]n$

$$I_S[f_n] = \frac{1}{n} \sum_{i=1}^n V(f_n(x_i), y_i)$$

Метод наискорейшего спуска



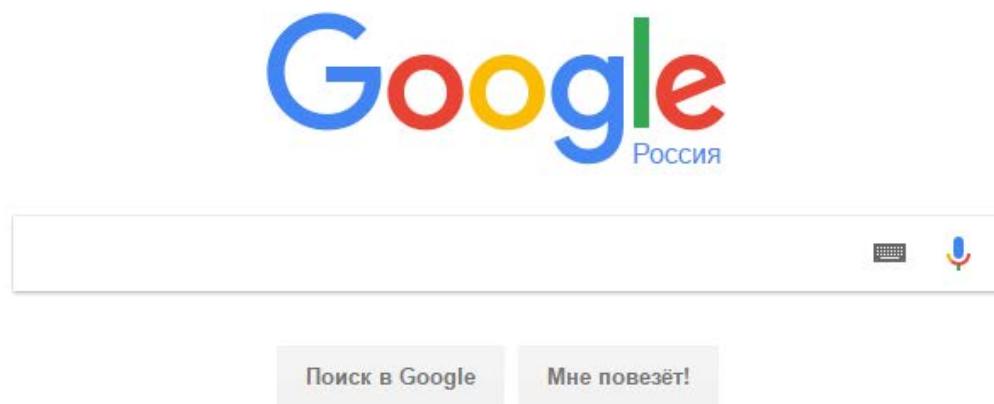
Методология анализа данных CRISP-DM



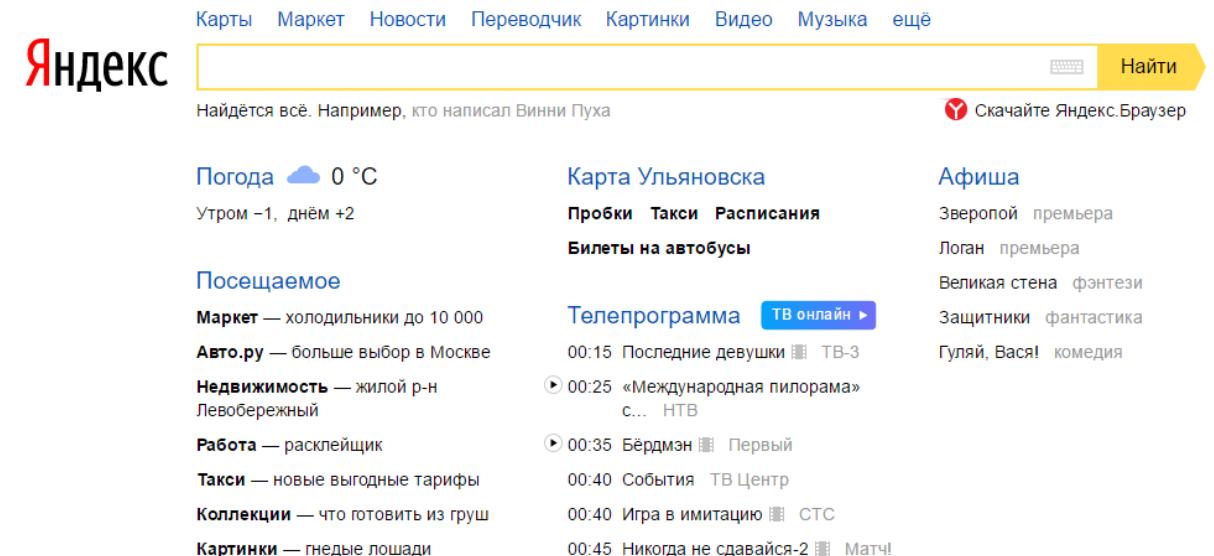
Сфера применения ML

Automation control	Bioinformatics	Stock technical analysis	Image generation	Speech generation
Text generation	Categorizing documents	Credit scoring	Medical diagnosis	Detection of fraud
Spam detection	Learning ranking in information search	Searching potential customers	Predicting customer care	Decision-making
Predicting time series	Gesture recognition	Image recognition	Speech recognition	Recognition of handwriting
Recognition of physical activity	Technical diagnostics	Financial supervision	Chemoinformatics	

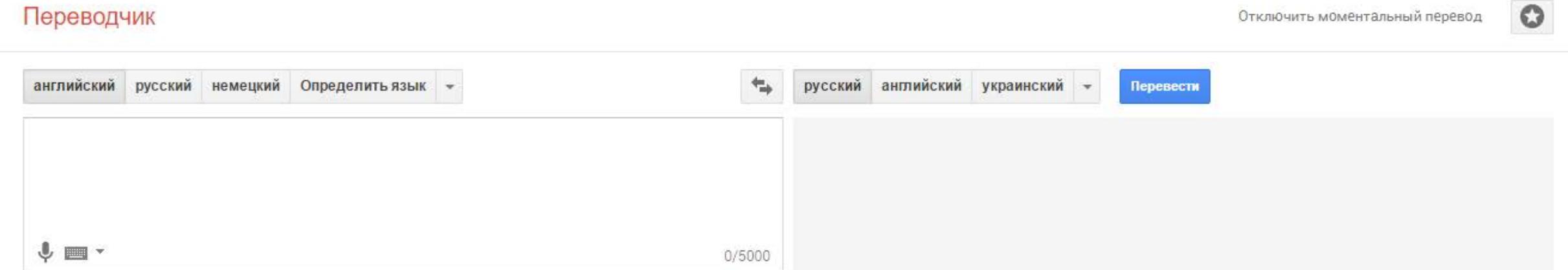
ML приложение



The screenshot shows the Google Russia search homepage. It features the Google logo with 'Россия' below it. A search bar is centered with a microphone icon and a keyboard icon. Below the search bar are two buttons: 'Поиск в Google' (Search in Google) and 'Мне повезёт!' (I'm feeling lucky). The background is white with some light gray UI elements.



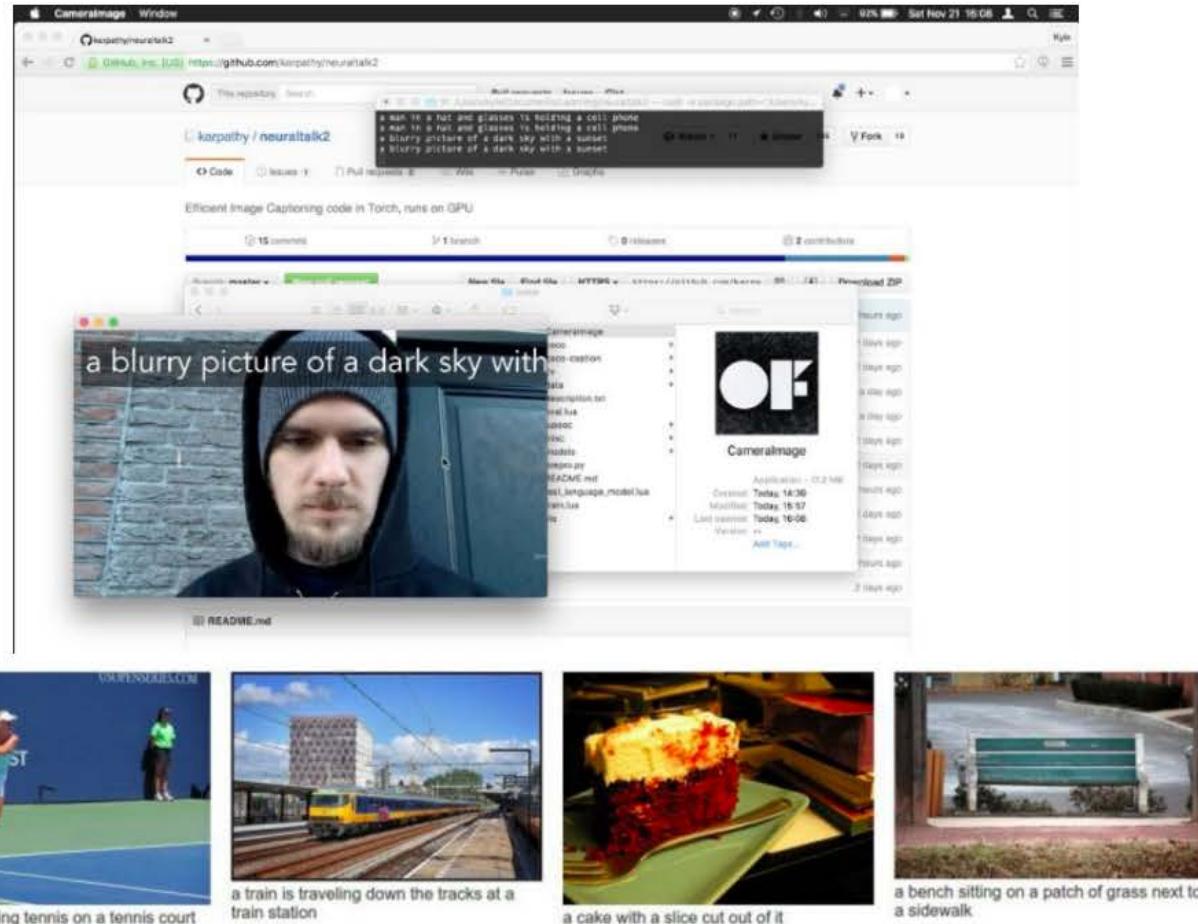
The screenshot shows the Yandex search results page. At the top, there's a navigation bar with links to Карты (Maps), Маркет (Market), Новости (News), Переводчик (Translator), Картички (Cartoon Images), Видео (Video), Музыка (Music), and 'ещё' (more). The main search bar contains the query 'Найдётся всё. Например, кто написал Винни Пуха'. Below the search bar, there's a weather widget showing 'Погода' (Weather) with a cloud icon and '0 °C'. The page displays a list of search results under sections like 'Посещаемое' (Popular) and 'Телепрограмма' (TV Program). On the right side, there are sidebar sections for 'Карта Ульяновска' (Map of Ulyanovsk), 'Афиша' (Events), and a list of TV shows with their air times and channels.



The screenshot shows the Yandex Translator interface. At the top, there's a header with the word 'Переводчик' (Translator) on the left and a 'Отключить моментальный перевод' (Disable instant translation) link on the right. Below the header, there are language selection dropdowns for the source language ('английский', 'русский', 'немецкий') and the target language ('русский', 'английский', 'украинский'). A blue 'Перевести' (Translate) button is positioned between the dropdowns. The main area consists of two large input fields for text entry. The left field has a microphone and keyboard icon with a dropdown arrow, and the right field shows '0/5000' characters. There are also small icons for a double-headed arrow and a star at the bottom of each field.

ML приложение

NeuralTalk



<https://github.com/karpathy/neuraltalk2>

<https://vimeo.com/14649200>

ML приложение



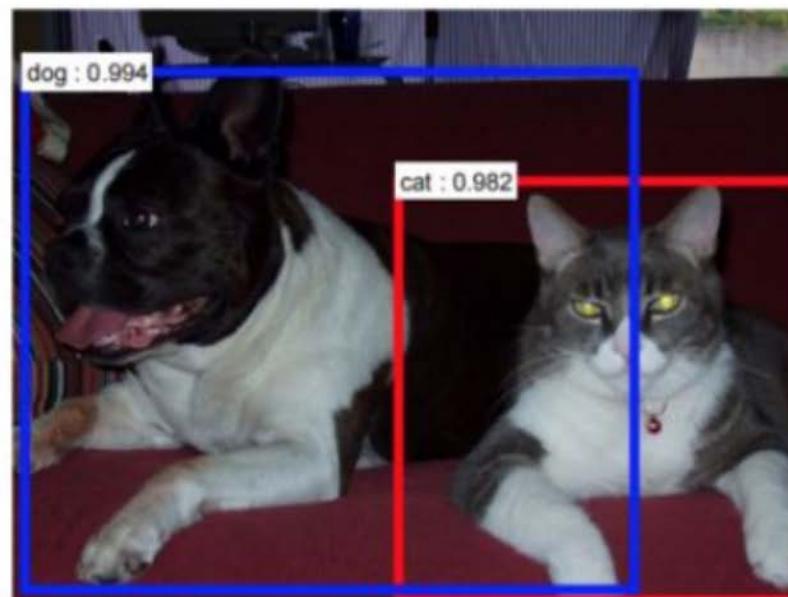
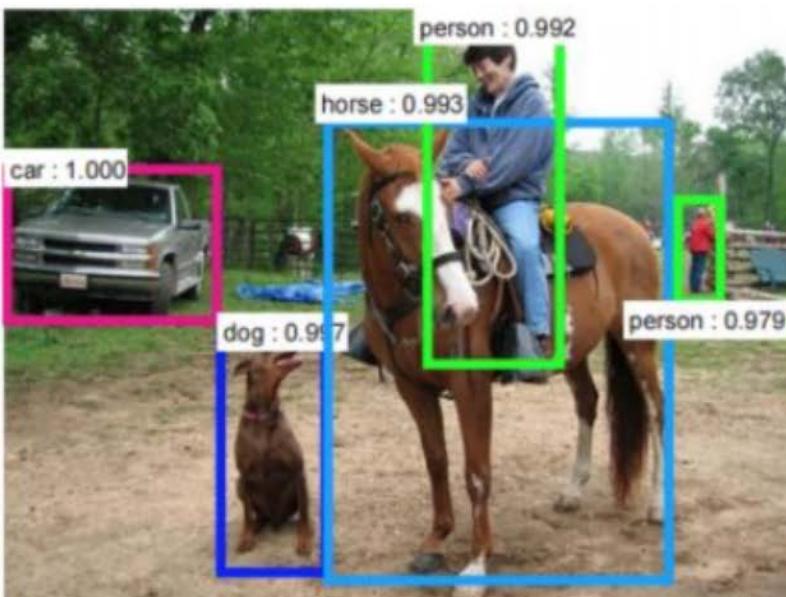
ML приложение



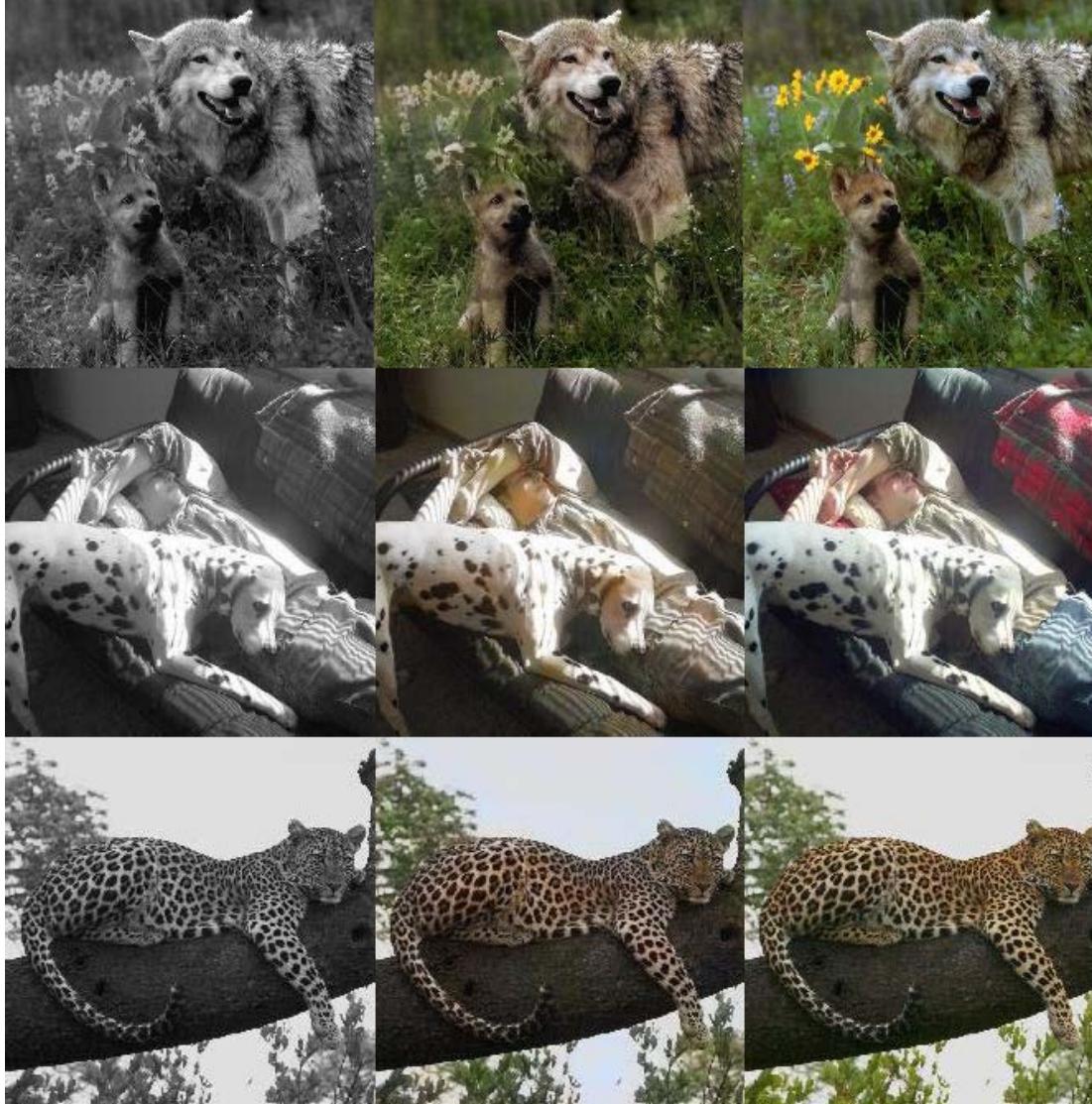
ML приложение. Порождающие модели.



ML приложение



ML приложение. Раскрашивание.



ML приложения. Маркировка.



"girl in pink dress is jumping **in** air."



"black and white dog **jumps** over bar."



"young girl **in** pink shirt is swinging on swing."



"man in blue wetsuit is surfing on wave."



"little girl is eating piece of cake."



"baseball player is throwing ball in game."

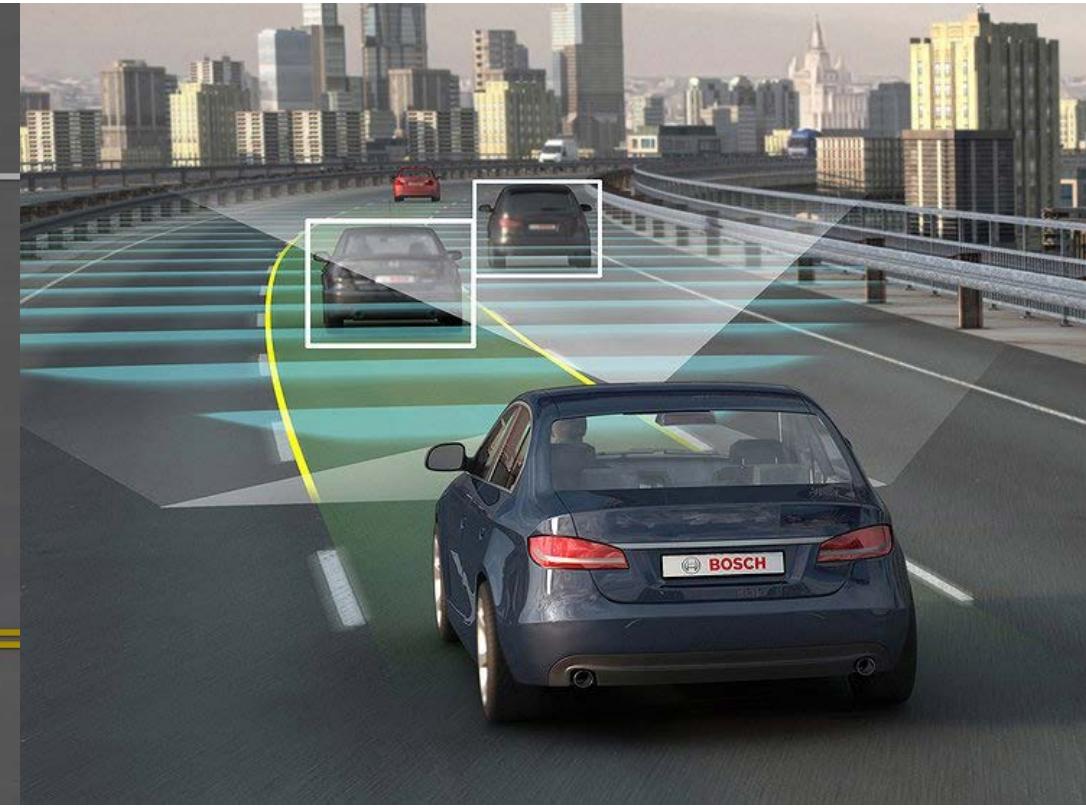
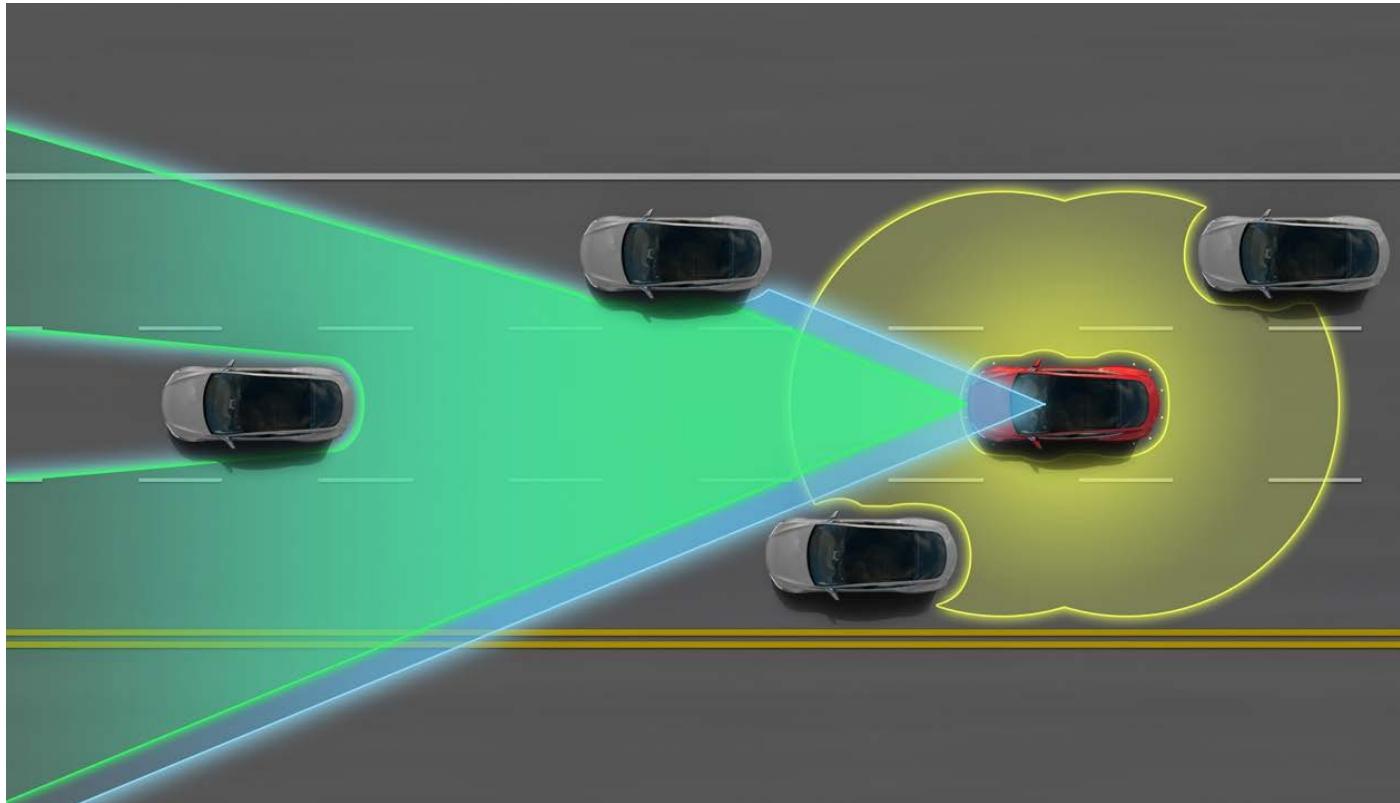


"woman is holding bunch of bananas."

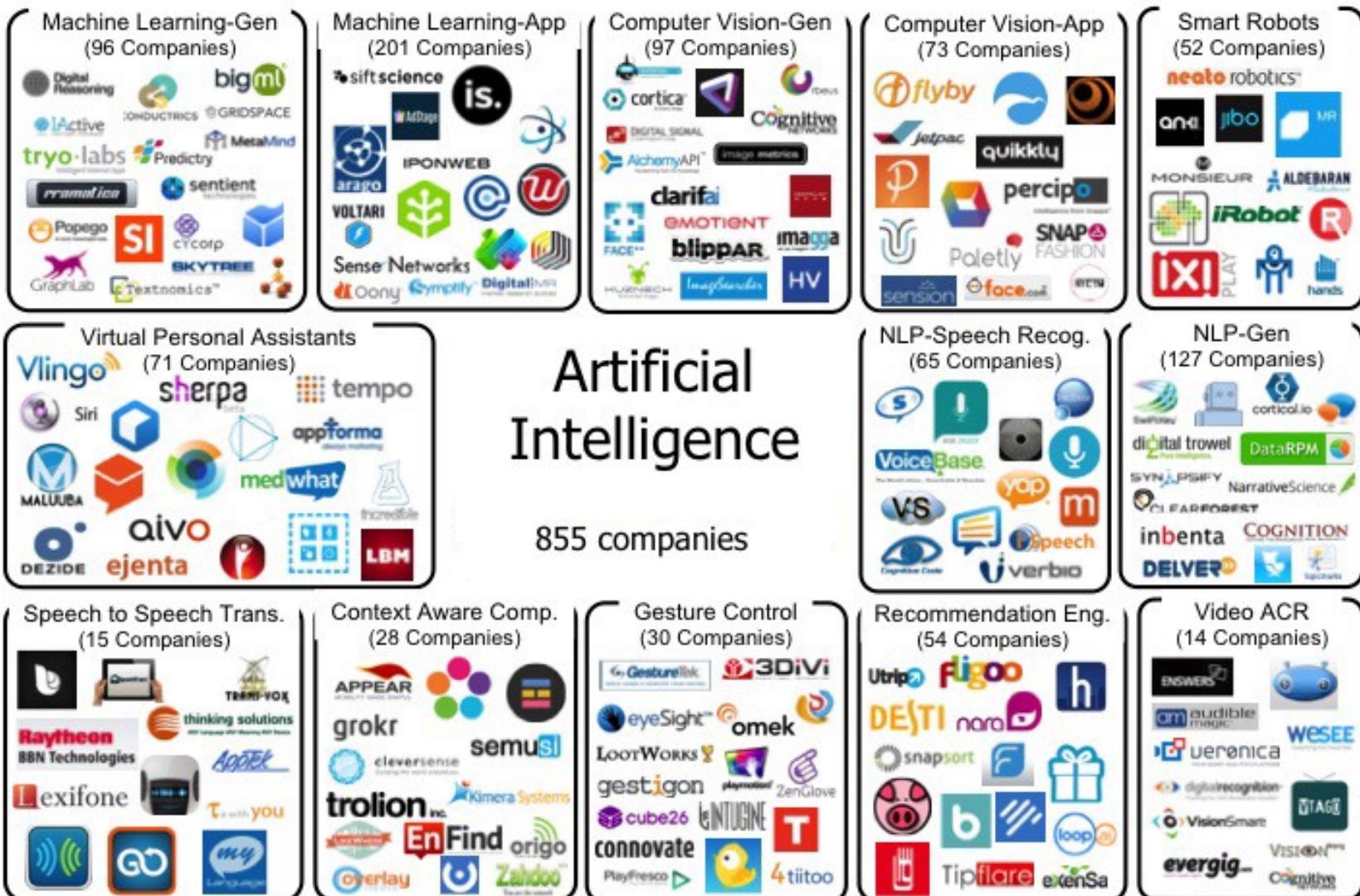


"black cat is sitting on top of suitcase."

ML приложение.



ML приложение



Artificial Intelligence

855 companies

Типы задач ML

1. Обучение с "учителем":
 - Классификация
 - Регрессия
 - Обучение ранжированию
 - Прогнозирование
2. Обучение без "учителя":
 - Кластеризация
 - Обучение на ассоциативных правилах (Association rules learning)
 - Обнаружение выбросов (outliers detection)
3. Обучение с подкреплением
4. Метаобучение (Meta-learning or learning-to-learn)
5. Формирование инвестиционного портфеля (Portfolio selection)
6. Коллаборативная фильтрация
7. ...

Подходы и методы

- Регрессия
- Байесовский вывод
- Деревья решений
- Нейронные сети
- Метод К-ближайших соседей
- Метод главных компонент (Principal component analysis)
- Метод опорных векторов (Support vector machines)
- Генетические алгоритмы

Инструменты

1. Python, R, Matlab, ...
2. IPython Notebook, PyCharm
3. Tensor Flow, Keras, Theano, Scikit-Learn
4. CNTK
5. Torch
6. Coffe
7. Apach Spark
8. Azure

Ссылки

- [ML course by Andrew Ng](<https://www.coursera.org/learn/machine-learning>)
- [ML course by Dmitry Efimov](https://github.com/diefimov/MTH594_MachineLearning)
- [ML course by OpenData Science](<https://github.com/Yorko/mlcourse.ai>)
- [MIT Deep learning](<https://github.com/lexfridman/mit-deep-learning>)
- <https://github.com/qati/DeepLearningCourse>
- https://github.com/roebius/deeplearning_keras2
- <https://github.com/enggen/Deep-Learning-Coursera>
- <https://github.com/fchollet/deep-learning-with-python-notebooks>