Lecture 01

Introduction to Machine learning

Kirill Svyatov
Ulyanovsk State Technical University
k.svyatov@ulstu.ru

Agenda

Brief history of technological revolutions

Future technologies

Definitions

Formulation of the problem

Definitions

ML applications

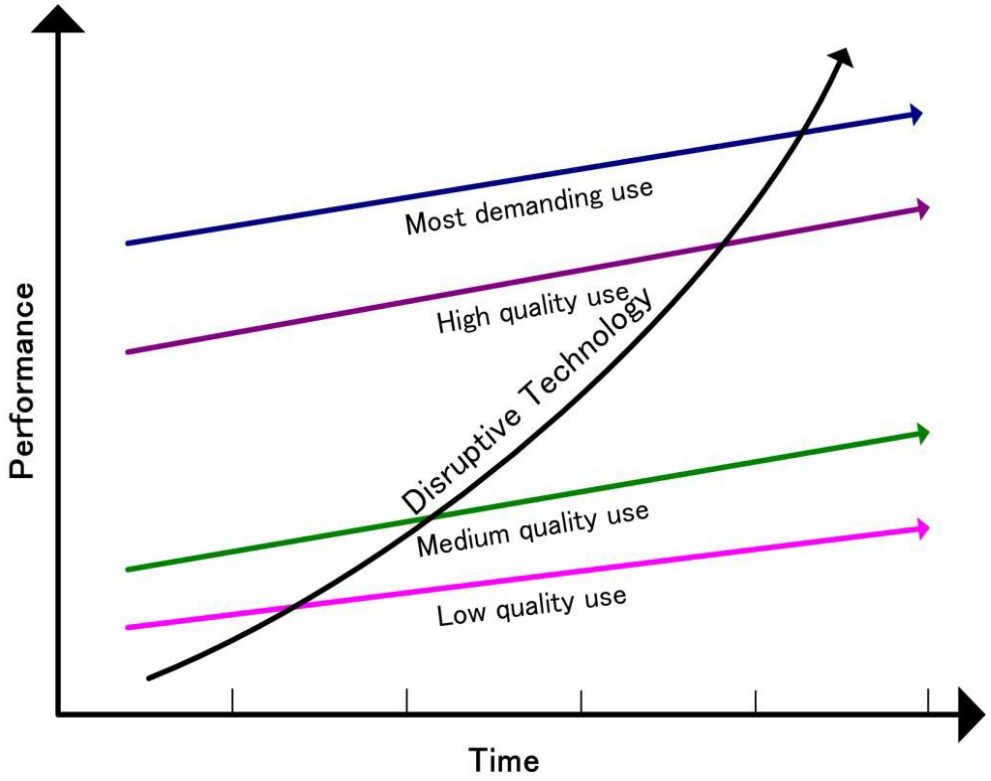
Methods of ML

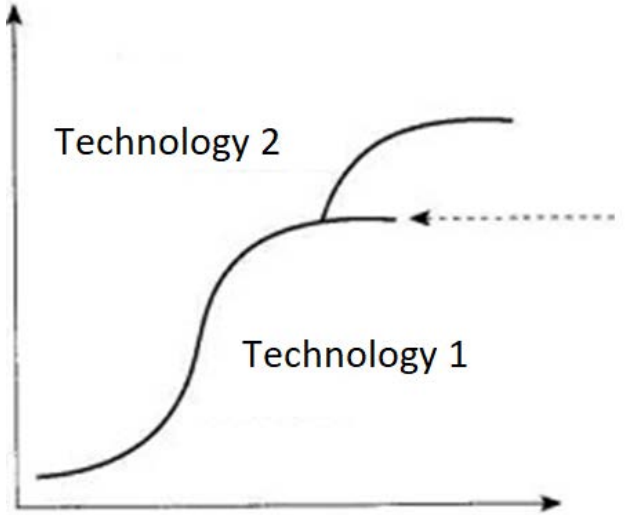
Industrial revolutions

	«I» IR XVII century	II IR XVIII – XIX	III IR XIX – XX	IV IR XXI
Technologies	▸ wood ▸ peat burn ▸ wind energy	▸ Iron, steel ▸ coal burn ▸ steam ▸ surgery, anesthesia ▸ agriculture machines	▸ plastic ▸ oil, electricity ▸ cars, airplains ▸ satellites ▸ antibiotics, early diagnosis ▸ mineral fertilizers	▸ 3D printing & materials ▸ thin films ▸ clean energy ▸ robots ▸ small satellites ▸ organic agriculture
	 	 	  	 
	Netherlands	*England*	*USA*	

[1] http://www.tv2.tomsk.ru/sites/www.tv2.tomsk.ru/files/mobil.jpg

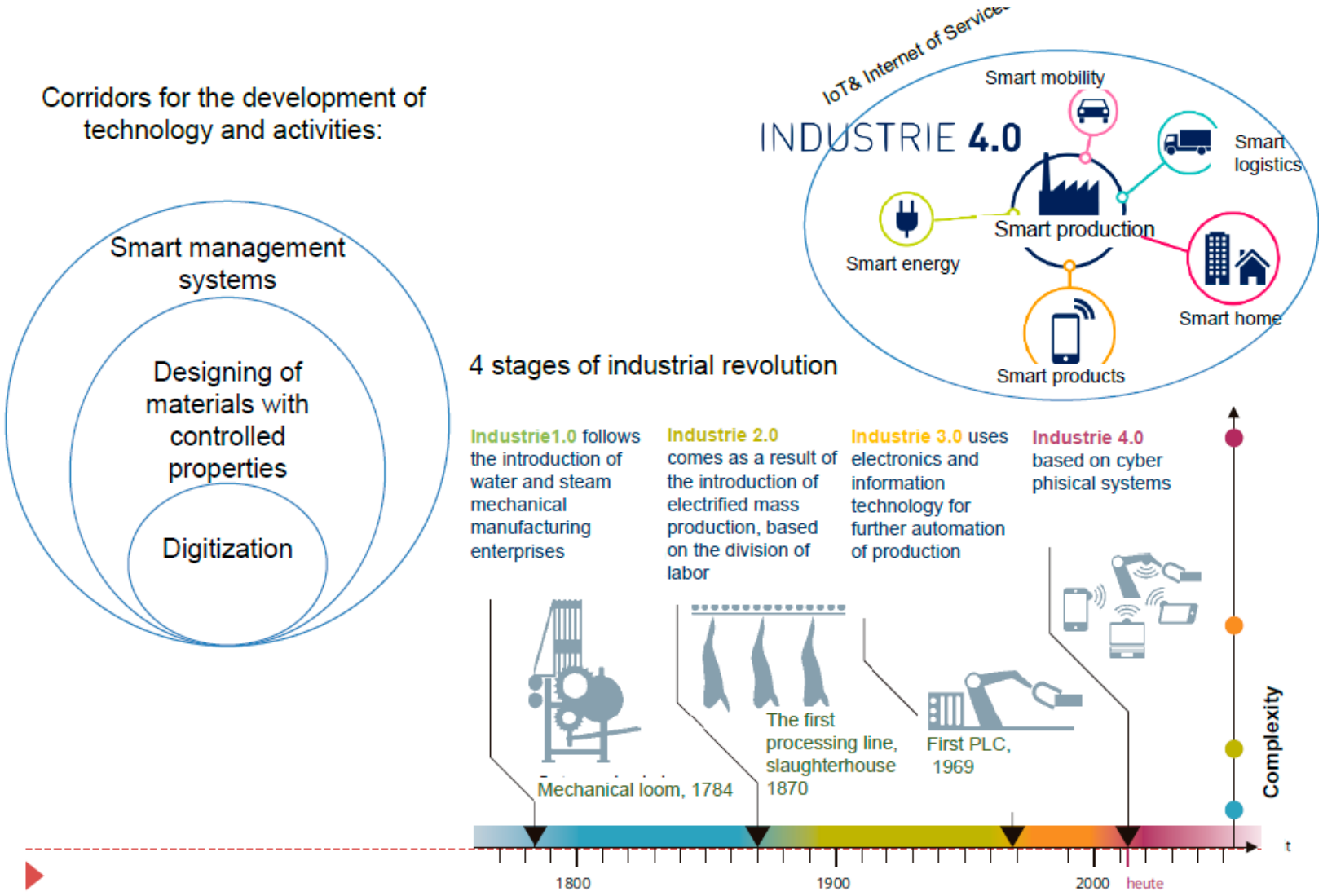
Disruptive technologies





Clayton Cristensen

Industry 4.0


Corridors for the development of technology and activities:
Smart management systems
Designing of materials with controlled properties
Digitization
IoT& Internet of Services
INDUSTRIE 4.0
Smart mobility
Smart logistics
Smart production
Smart energy
Smart home
Smart products
4 stages of industrial revolution
Industrie1.0 follows the introduction of water and steam mechanical manufacturing enterprises
Industrie 2.0 comes as a result of the introduction of electrified mass production, based on the division of labor
Industrie 3.0 uses electronics and information technology for further automation of production
Industrie 4.0 based on cyber phisical systems
Mechanical loom, 1784
The first processing line, slaughterhouse 1870
First PLC, 1969
Complexity
t
1800
1900
2000
heute

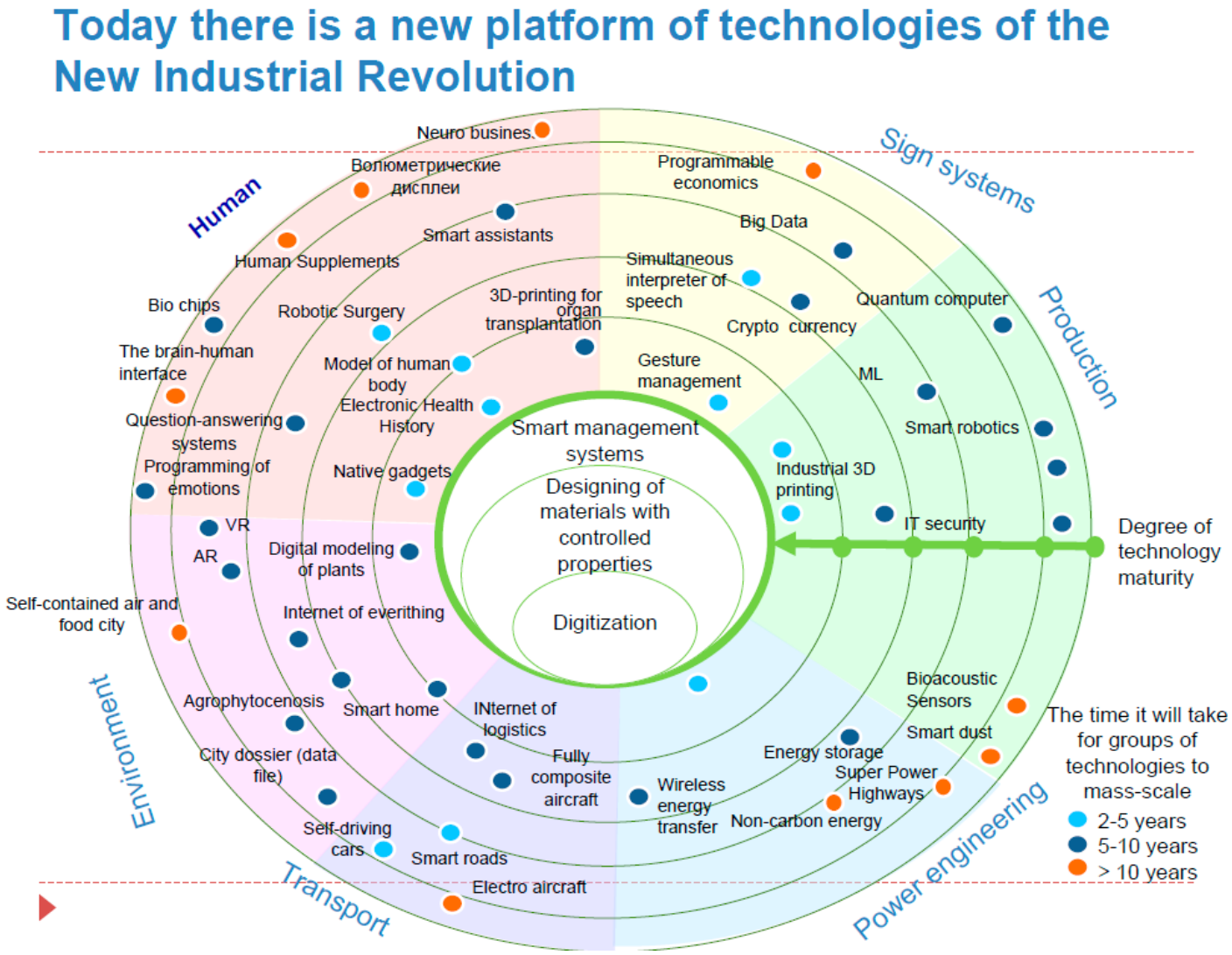
Labor productivity

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

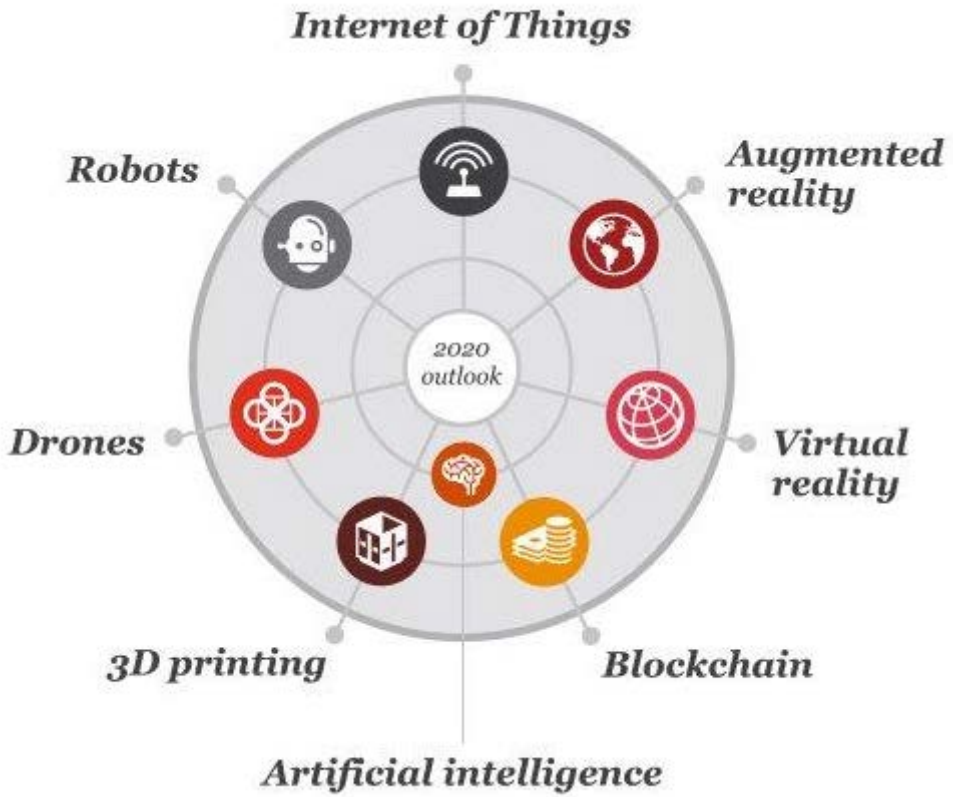
Steam engine	Early robotics	IT	Robots, AI, ML
1850-1910	1993-2007	1995-2005	2015-2065
Growth in labor productivity per year			
0,3%	0,4%	0,6%	0,8 - 1,4%

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

Modern technologies



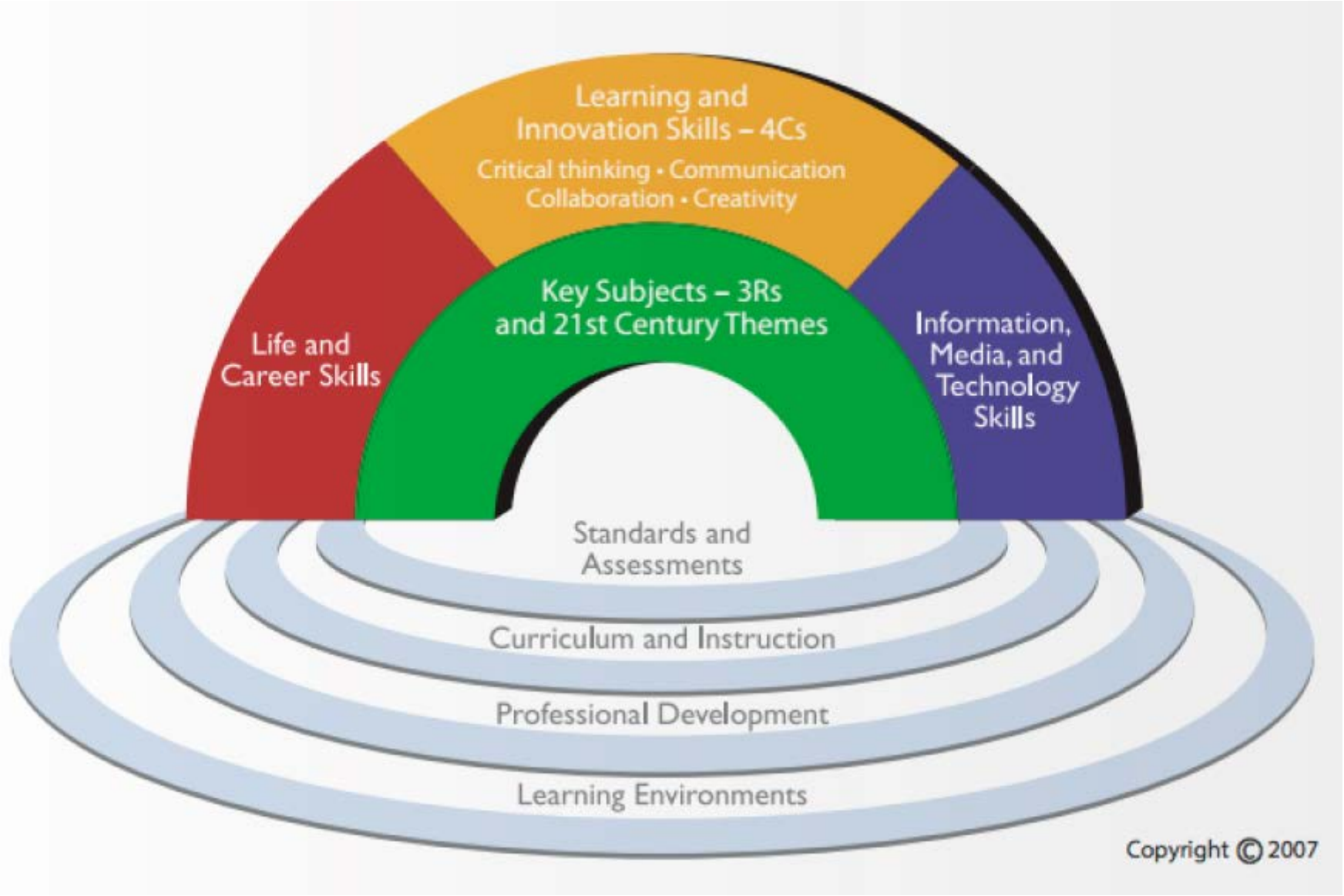
The essential 8 technologies

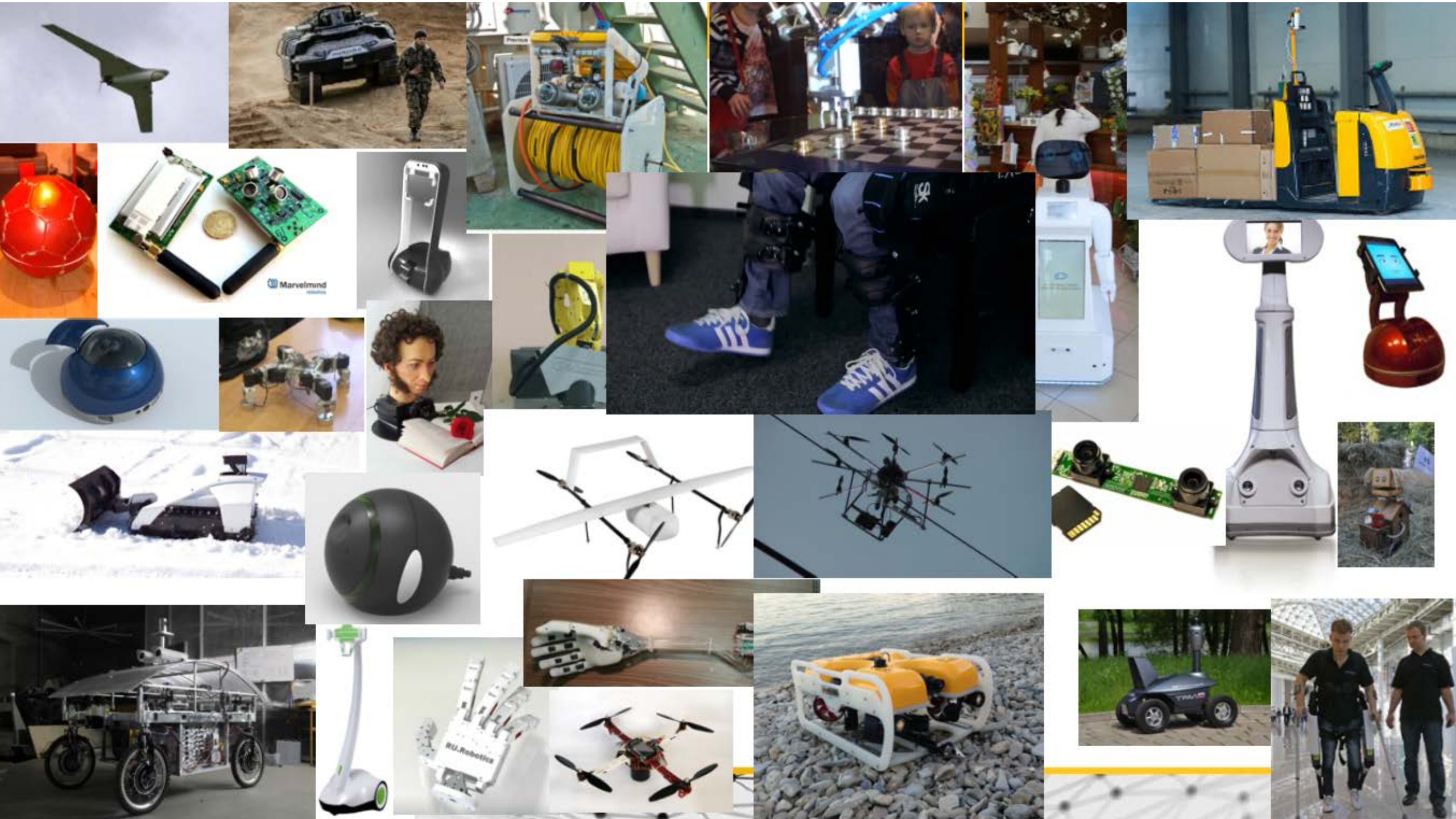


https://pwc.blogs.com/ceoinsights/2016/08/a-guide-to-the-essential-eight-emerging-technologies.html

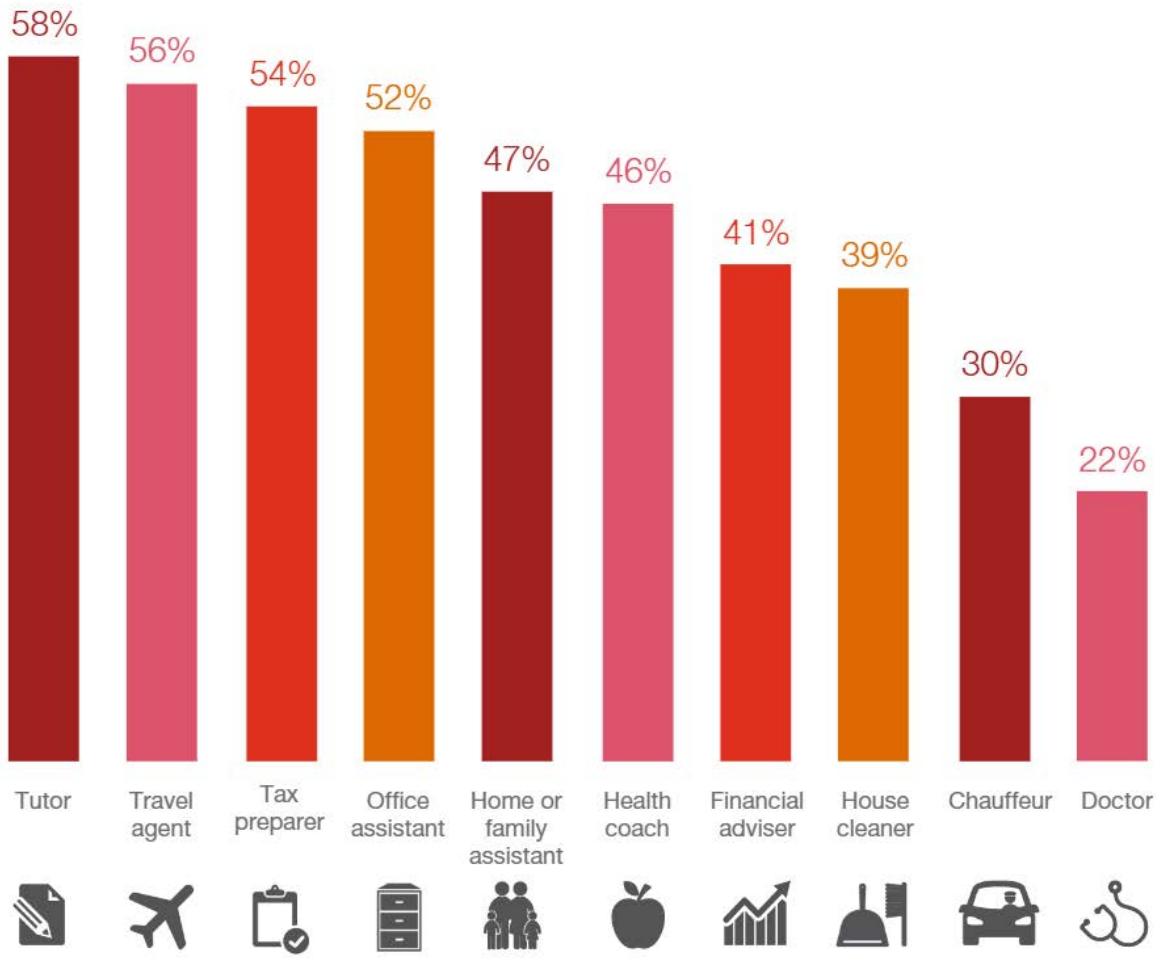
Problems of transformation

Important skills



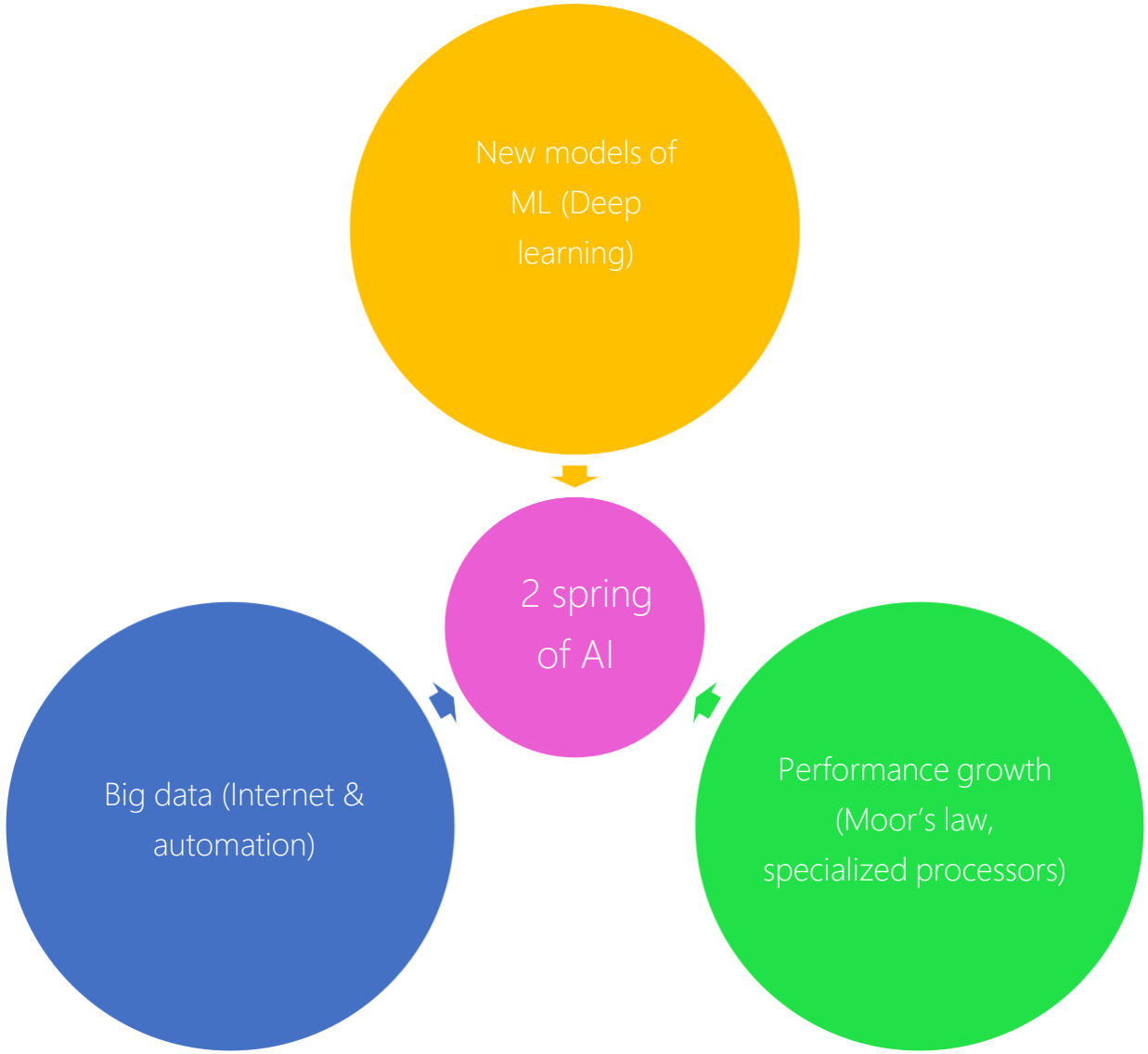

Marvelmind
RU.Robotics

"In the next five years, I can see AI replacing humans as a..."



http://pwcartificialintelligence.com/

AI winter & spring



1940—1960: first neural networks

1970—1990: «winter of AI»

1980—2000: algorythms for the second AI spring

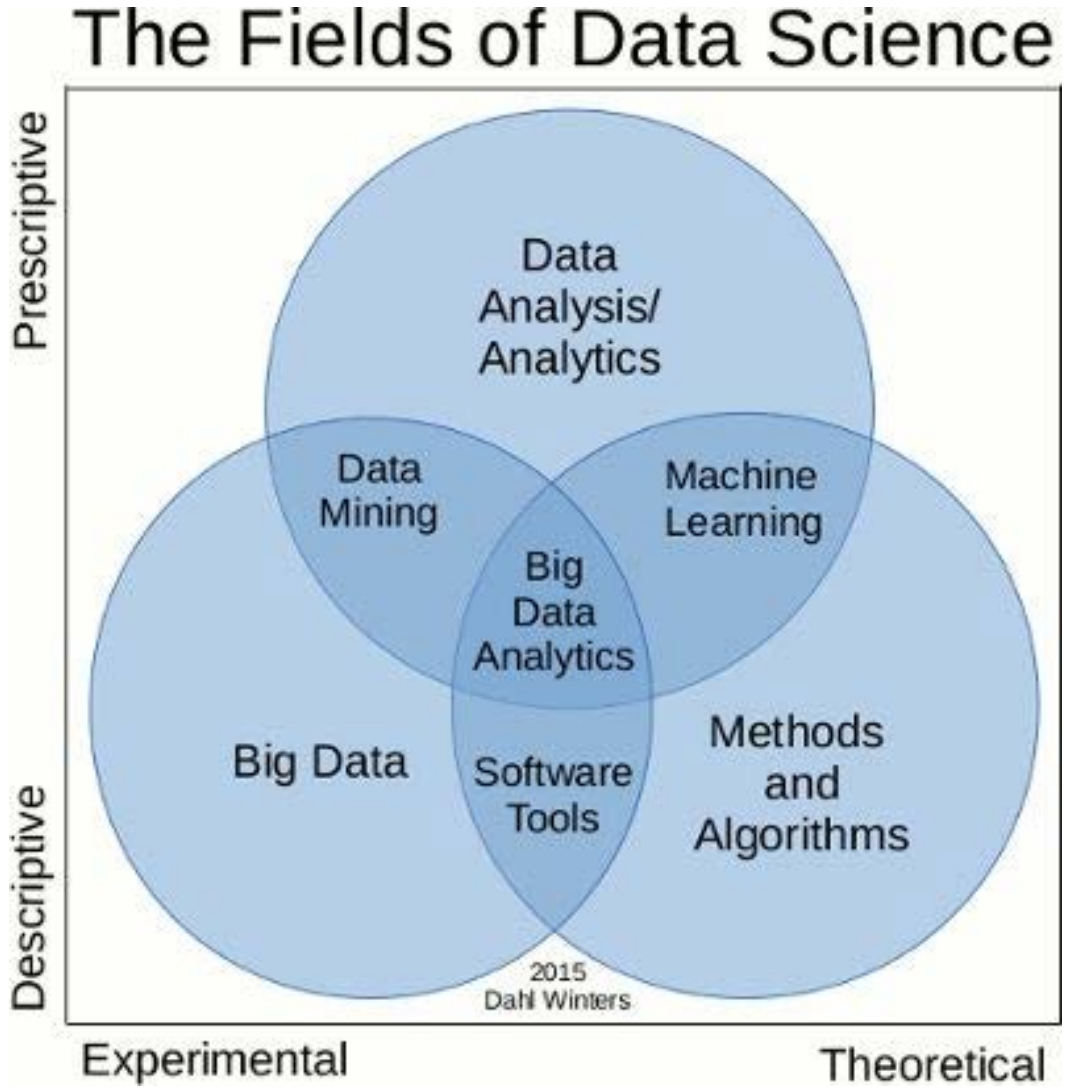
2010+ — spectacular examples of AI

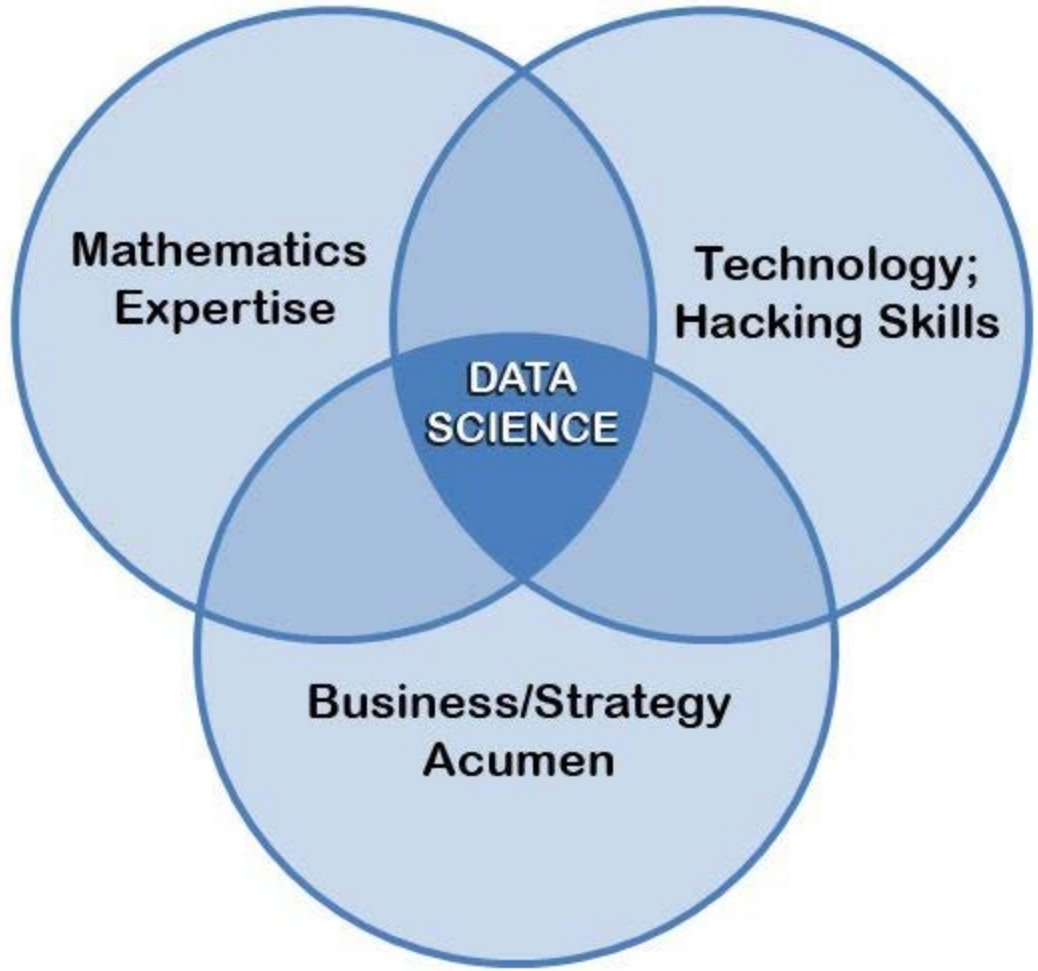
2016+ — investment and media boom

ML definition

Machine Learning:

- Machine learning is a topic of an artificial intelligence, a mathematical discipline that uses mathematical statistics, numerical optimization methods, probability theory, discrete analysis, to extract knowledge from data.
- Field of study that gives computers the ability to learn without being explicitly programmed. (Arthur Samuel, 1959)


The Fields of Data Science
Prescriptive
Descriptive
Experimental
Theoretical
Data Analysis/ Analytics
Data Mining
Machine Learning
Big Data Analytics
Big Data
Software Tools
Methods and Algorithms
2015
Dahl Winters


Mathematics
Expertise
Technology;
Hacking Skills
DATA
SCIENCE
Business/Strategy
Acumen

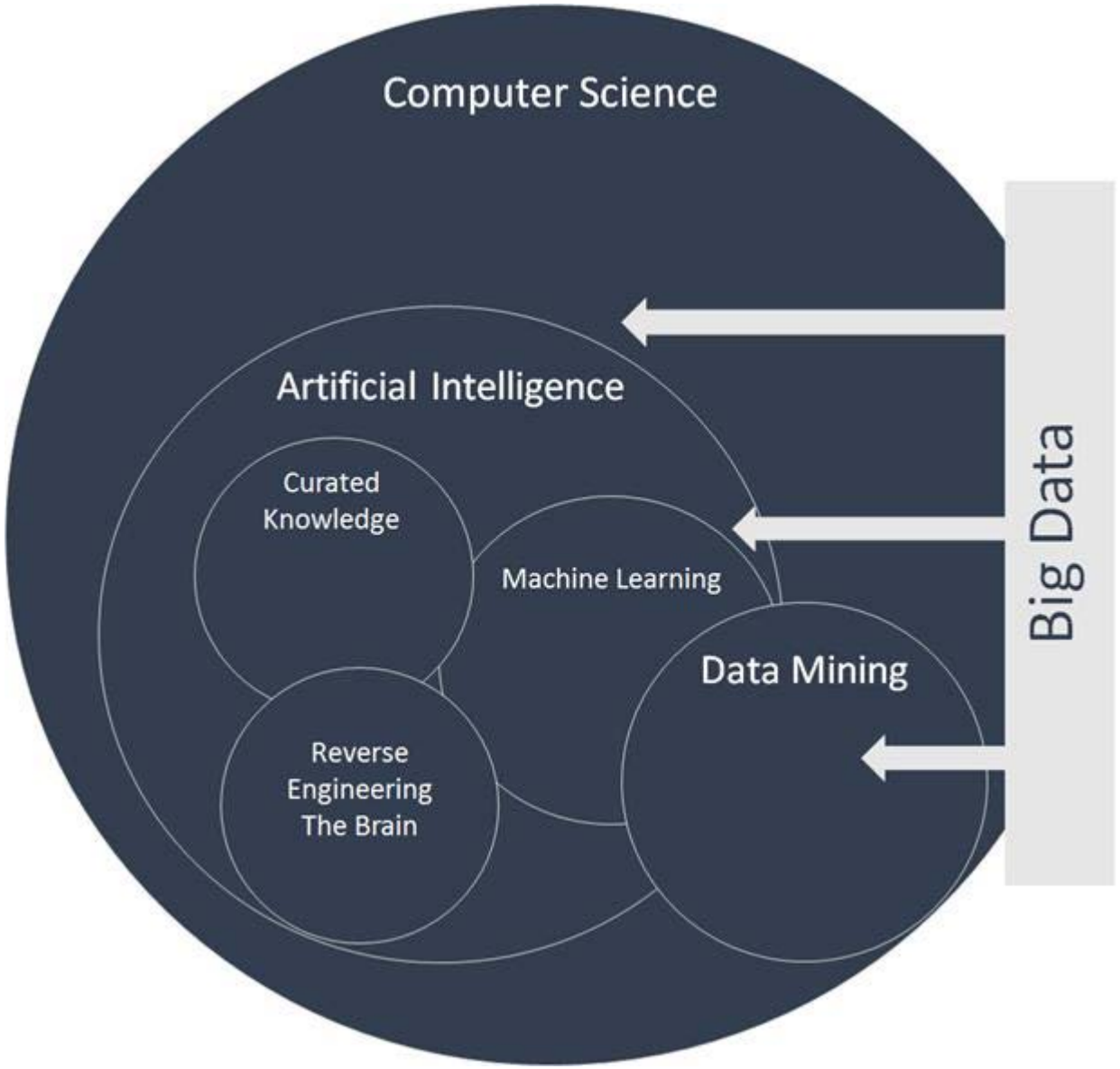
Data mining key features

- Data can be inaccurate, incomplete, heterogeneous, indirect, and at the same time have huge volumes;
- Data analysis algorithms themselves may have the ability to learn from precedents;
- Processes of raw data processing into information require non-trivial automation.

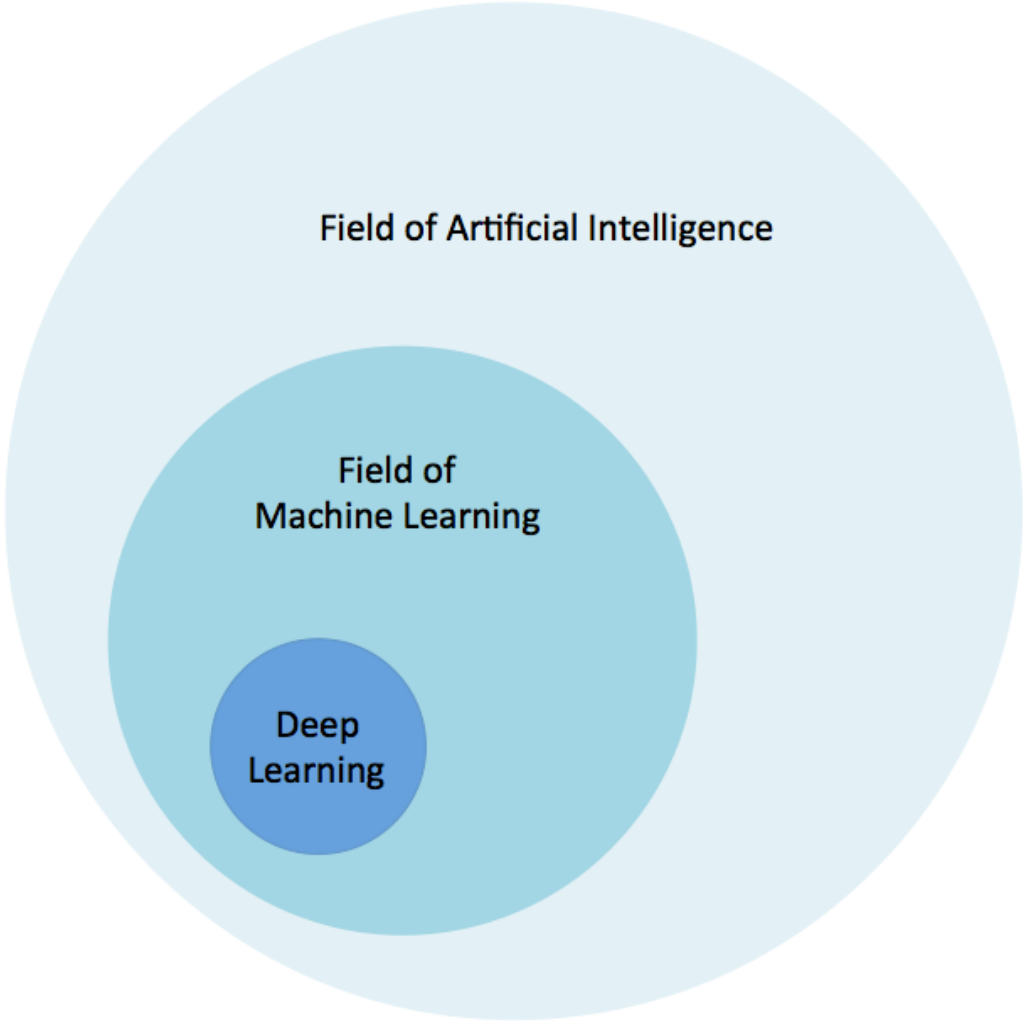
Data mining key features

Over the past decades, significant efforts in the field of Data Mining have focused on the creation of specialized algorithms capable of performing the same tasks during linear or even logarithmic time without significant loss of accuracy.

Data mining key features



Data mining key features



Data mining key features

A computer program is said to learn from experience E with respect to some class of tasks T and performance measure P, if its performance at tasks in T, as measured by P, improves with experience E.

(T.Mitchell)

Data mining key features

There are two types of learning:

- Precedent training
- Deductive training

Data mining key features

The purpose of precedent training is to generalize or gain knowledge about the "law of nature."

Data mining key features

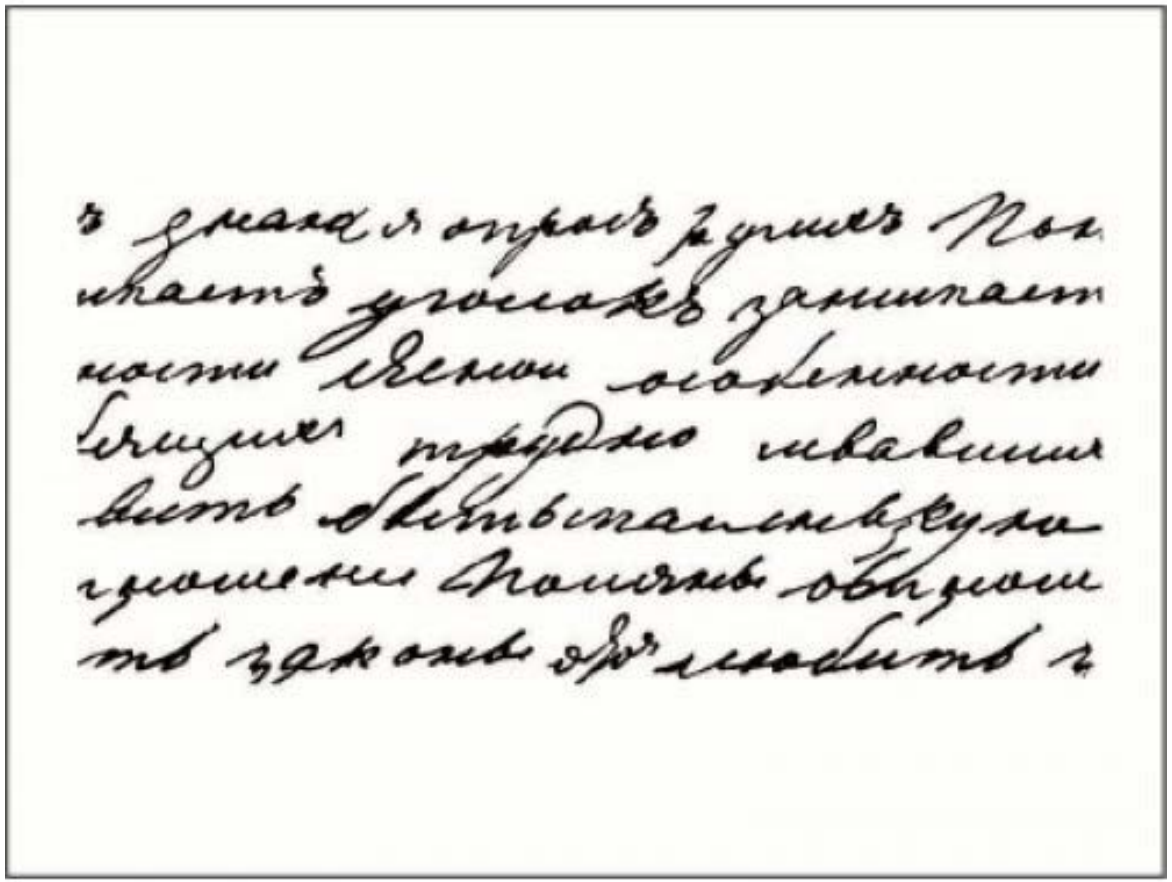
Data mining key features

SPHINX OF BLACK QUARTZ JUDGE MY VOW

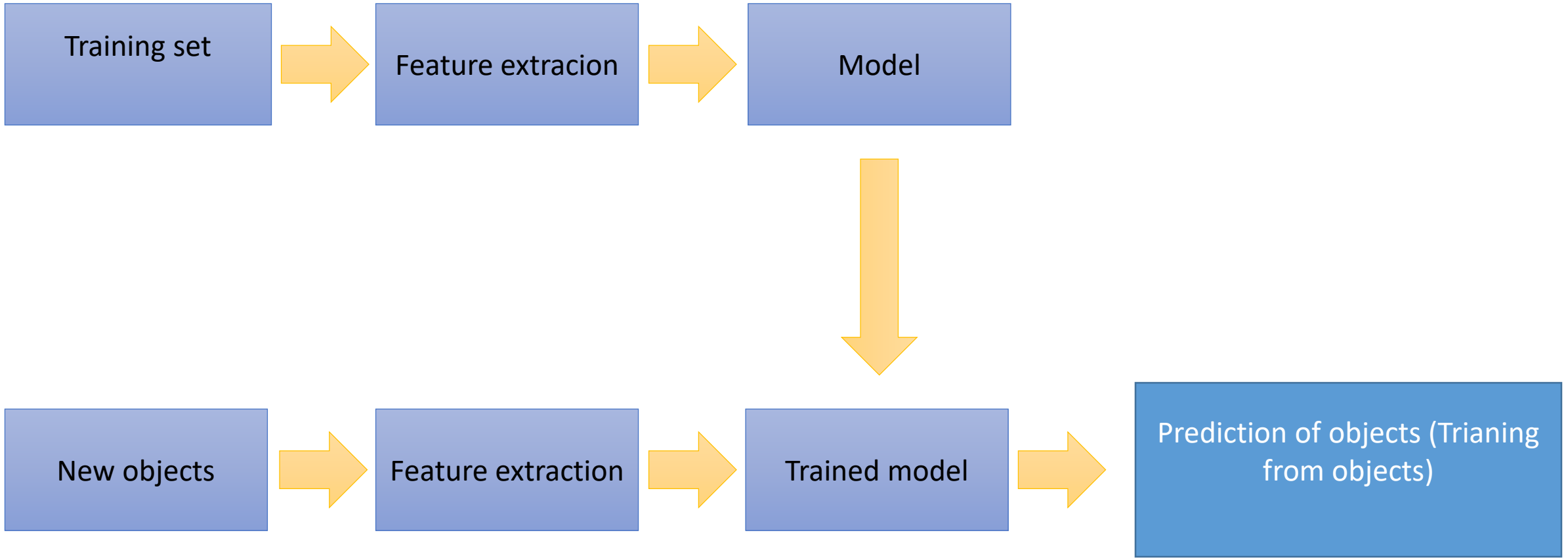
Sphinx of black quartz judge my vow

ABCDEFGHabcdefhg

0123456789/!@#$%



Data mining key features



ML task definition

Machine learning:

- What is an object X (what features) and what is an answer Y?
- How to build a model M?
- How can we make an approximation of Y from X with M?

ML task definition

Input data types:

- Image
- Text
- Sound
- GEO data
- Timeseries
- Datasheet

Generalization

Generalization error (also known as the **out-of-sample error**[1]) is a measure of how accurately an algorithm is able to predict outcome values for previously unseen data

In a learning problem, the goal is to develop a function $f(x)$ that predicts output values y based on some input data x. The expected error, $I[f_n]$ of a particular function f_n over all possible values of x and y is:

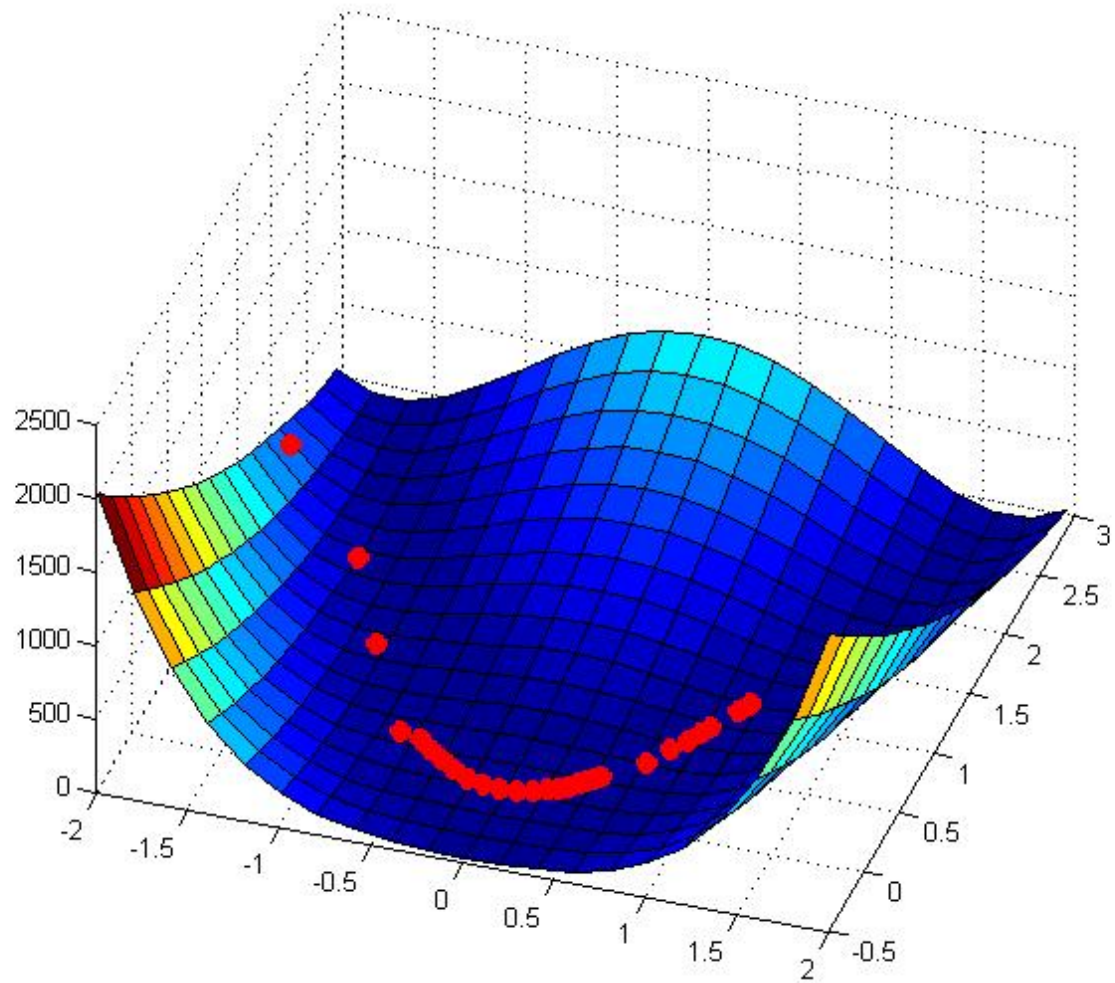
$$I[f_n] = \int_{X \times Y} V(f_n(x), y) \rho(x, y) dx dy,$$

where V denotes a loss function and $\rho(x, y)$ is the unknown joint probability distribution for x and y.

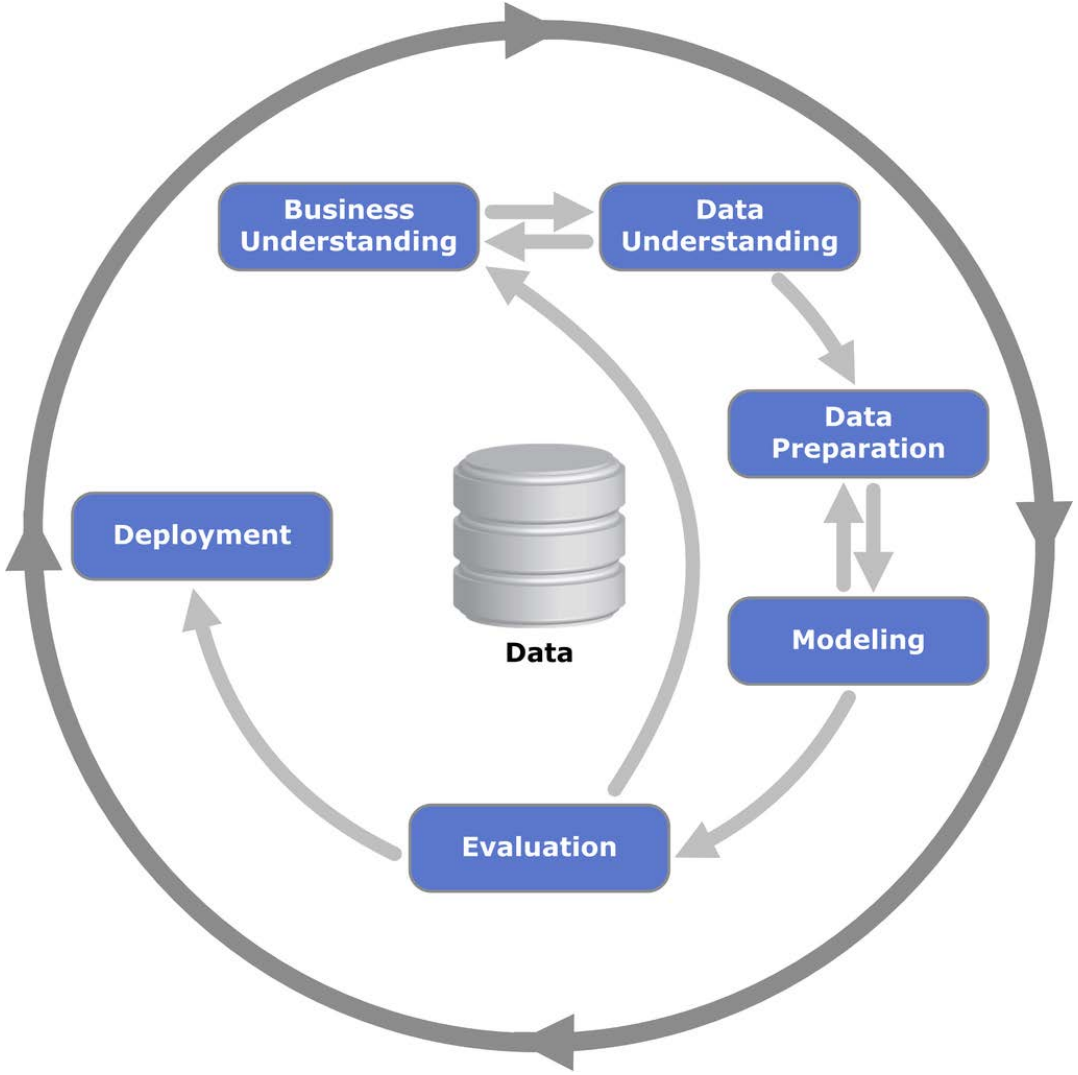
Without knowing the joint probability distribution, it is impossible to compute $I[f]$. Instead, we can compute the empirical error on sample data. Given n data points, the empirical error is:

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

Gradient descent



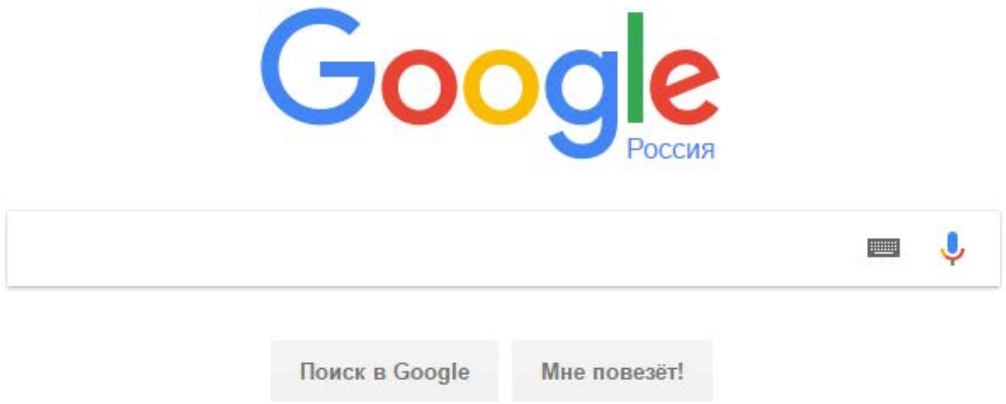
CRISP-DM methodology

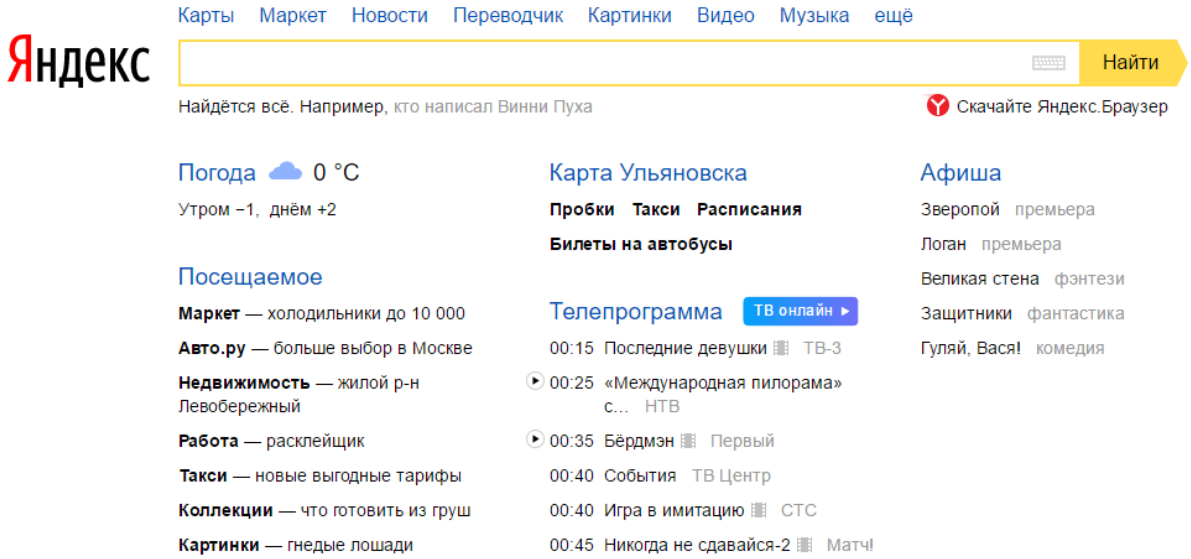


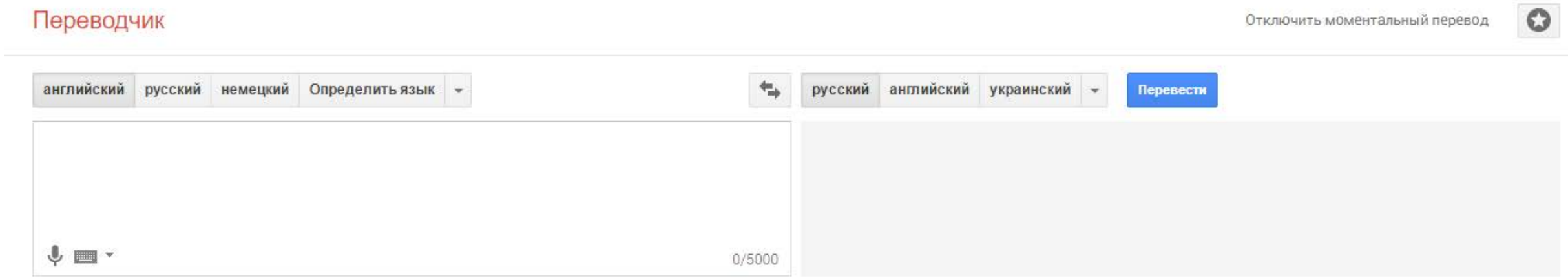
Appliances of ML

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	Pecognition of handwriting
Recognition of physical activity	Technical diagnostics	Financial supervision	Chemoinformatics	

ML application

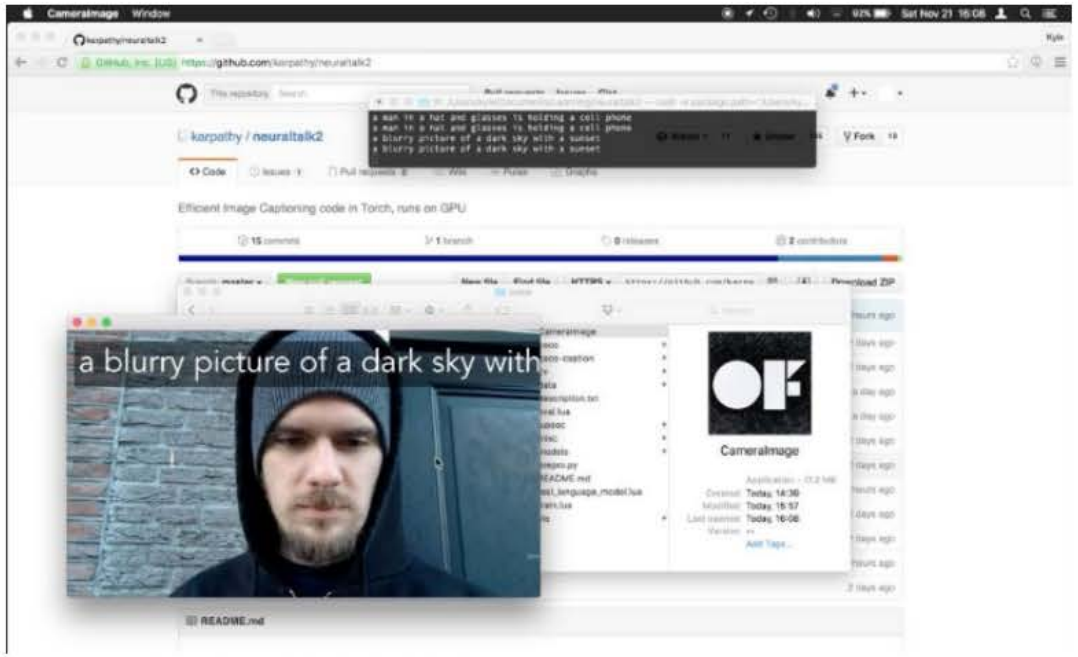






ML application

NeuralTalk




a man is playing tennis on a tennis court


a train is traveling down the tracks at a train station


a cake with a slice cut out of it


a bench sitting on a patch of grass next to a sidewalk

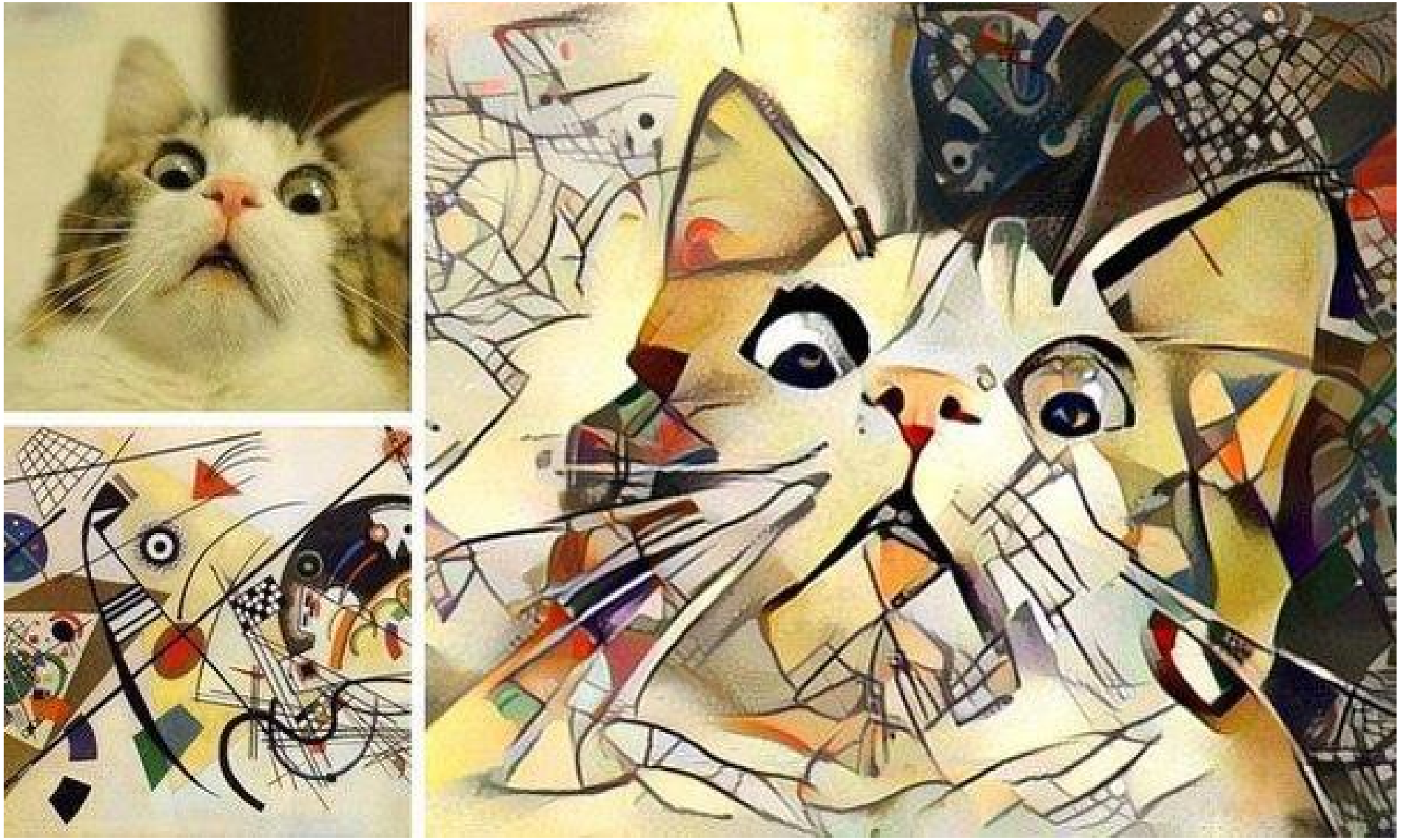
https://github.com/karpathy/neuraltalk2

https://vimeo.com/14649200

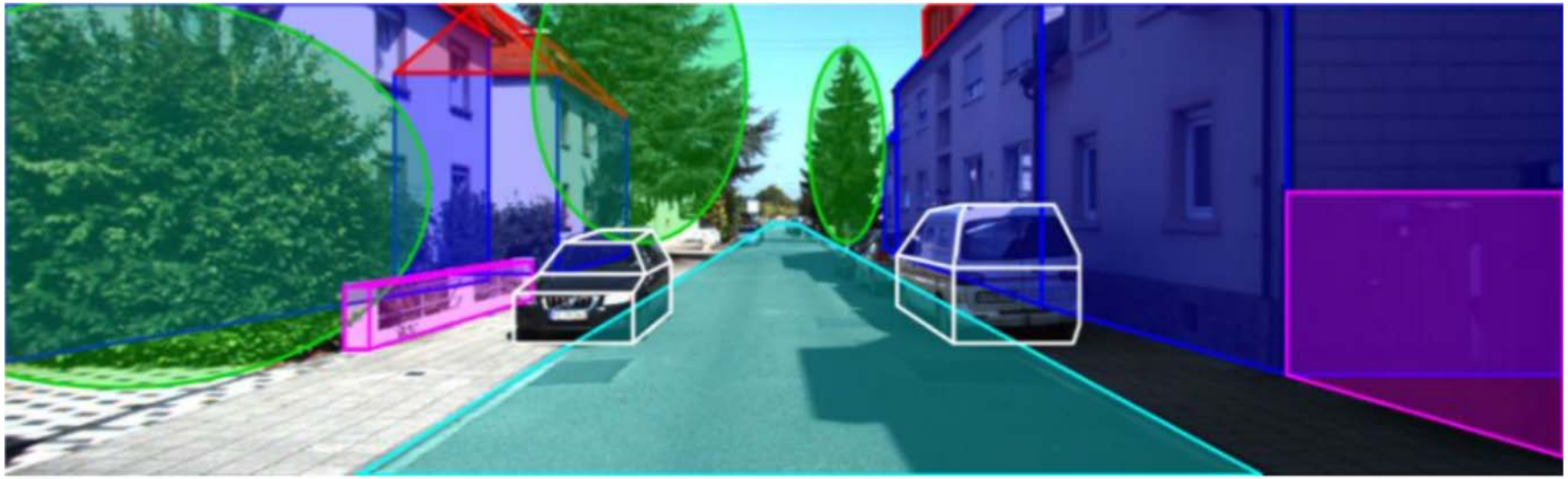
ML application

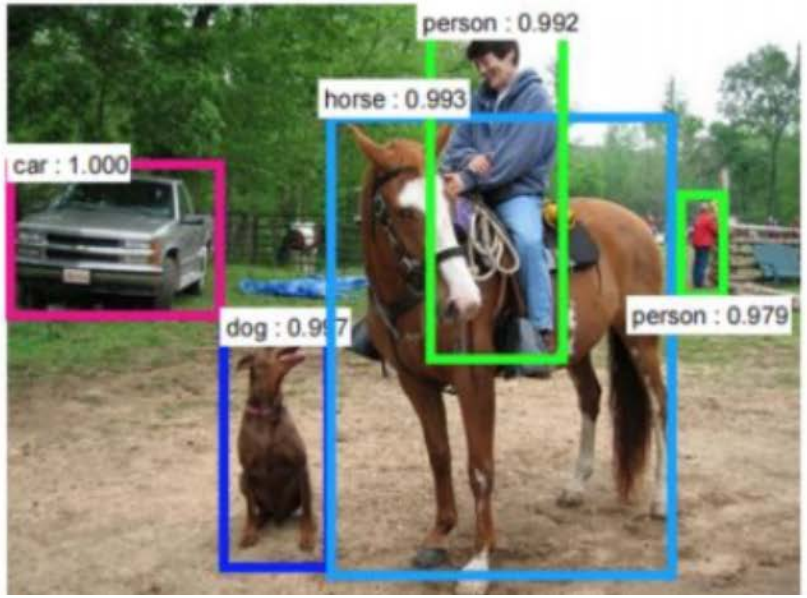
ML application

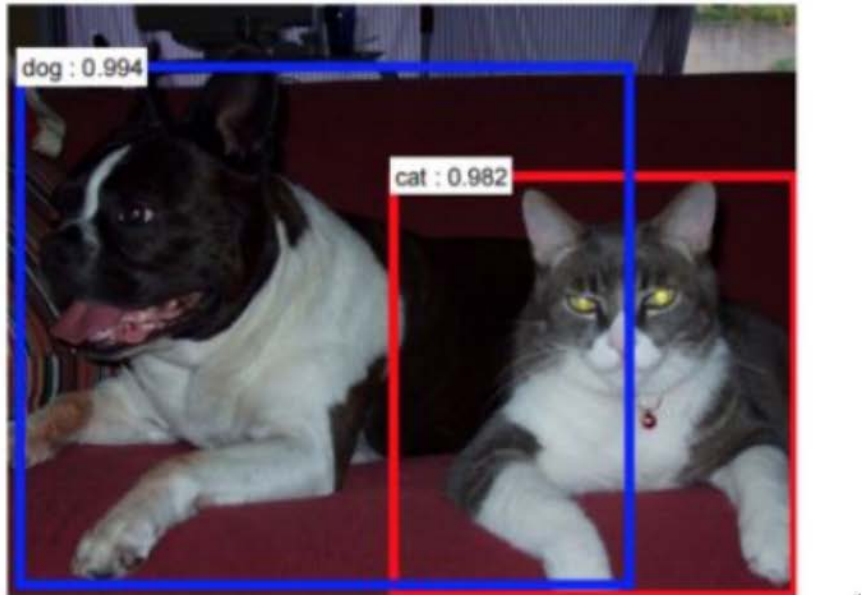
ML application. Generative models



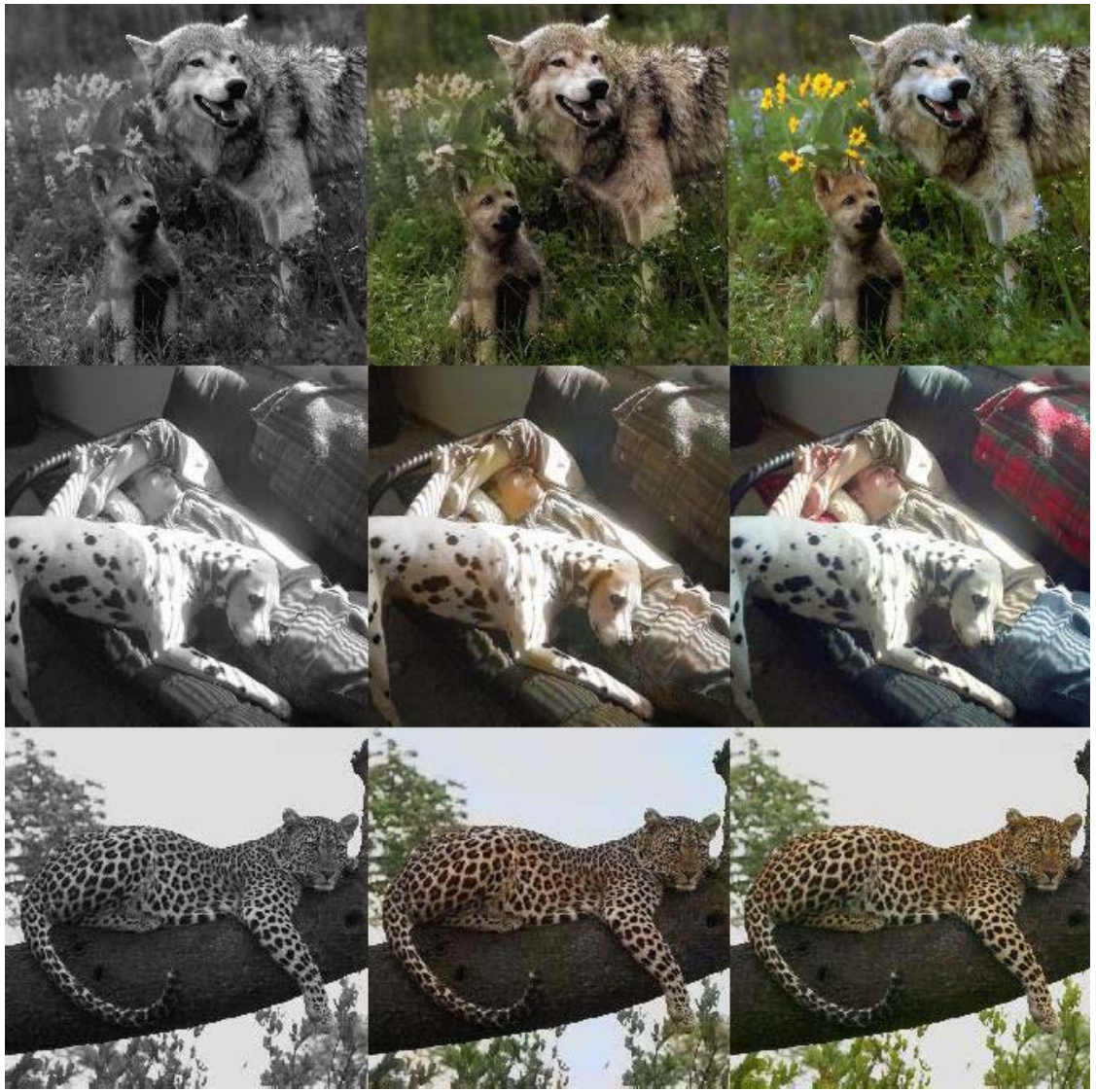
ML application







ML application. Colorization



ML applications. Labeling

"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."

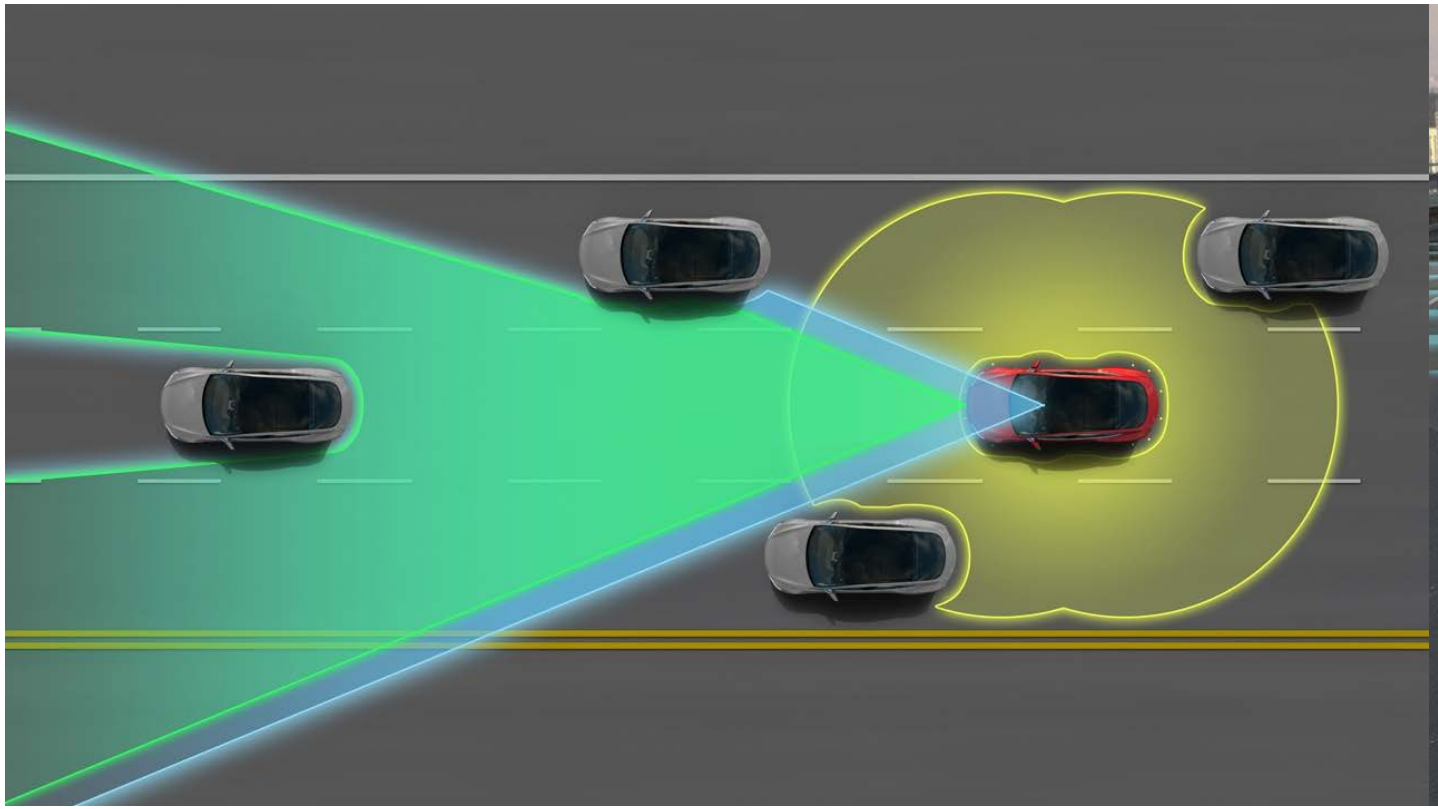
"little girl is eating piece of cake."

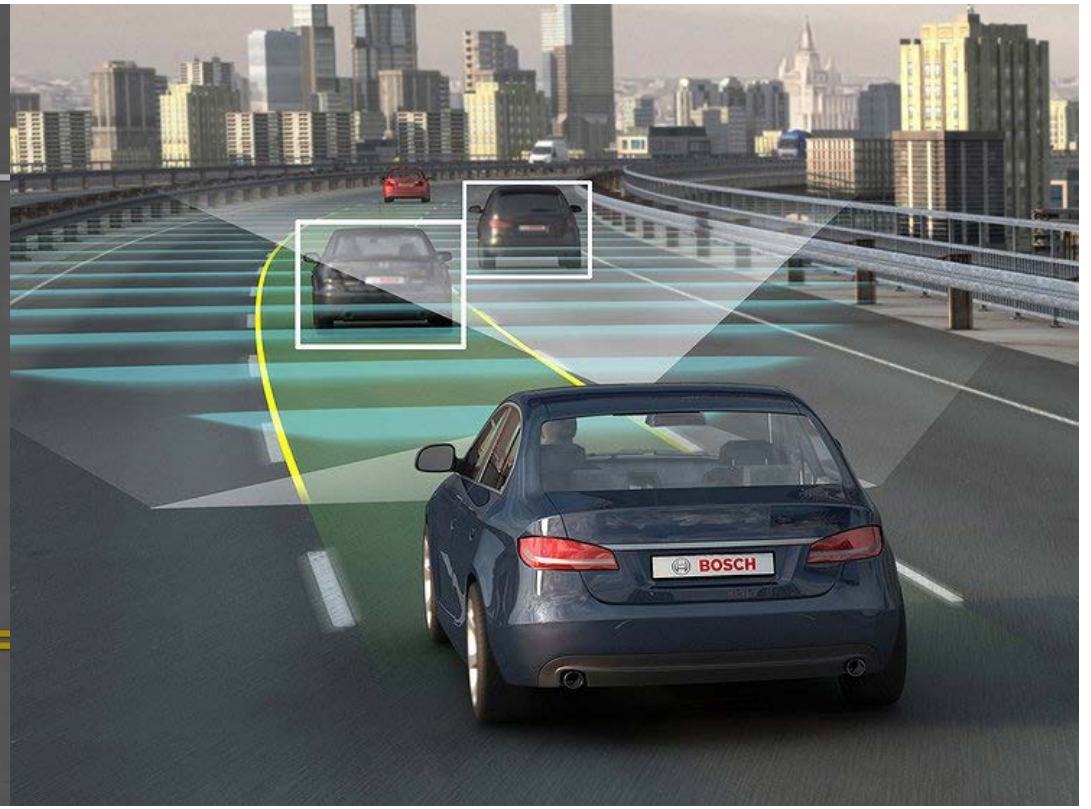
"baseball player is throwing ball in game."

"woman is holding bunch of bananas."

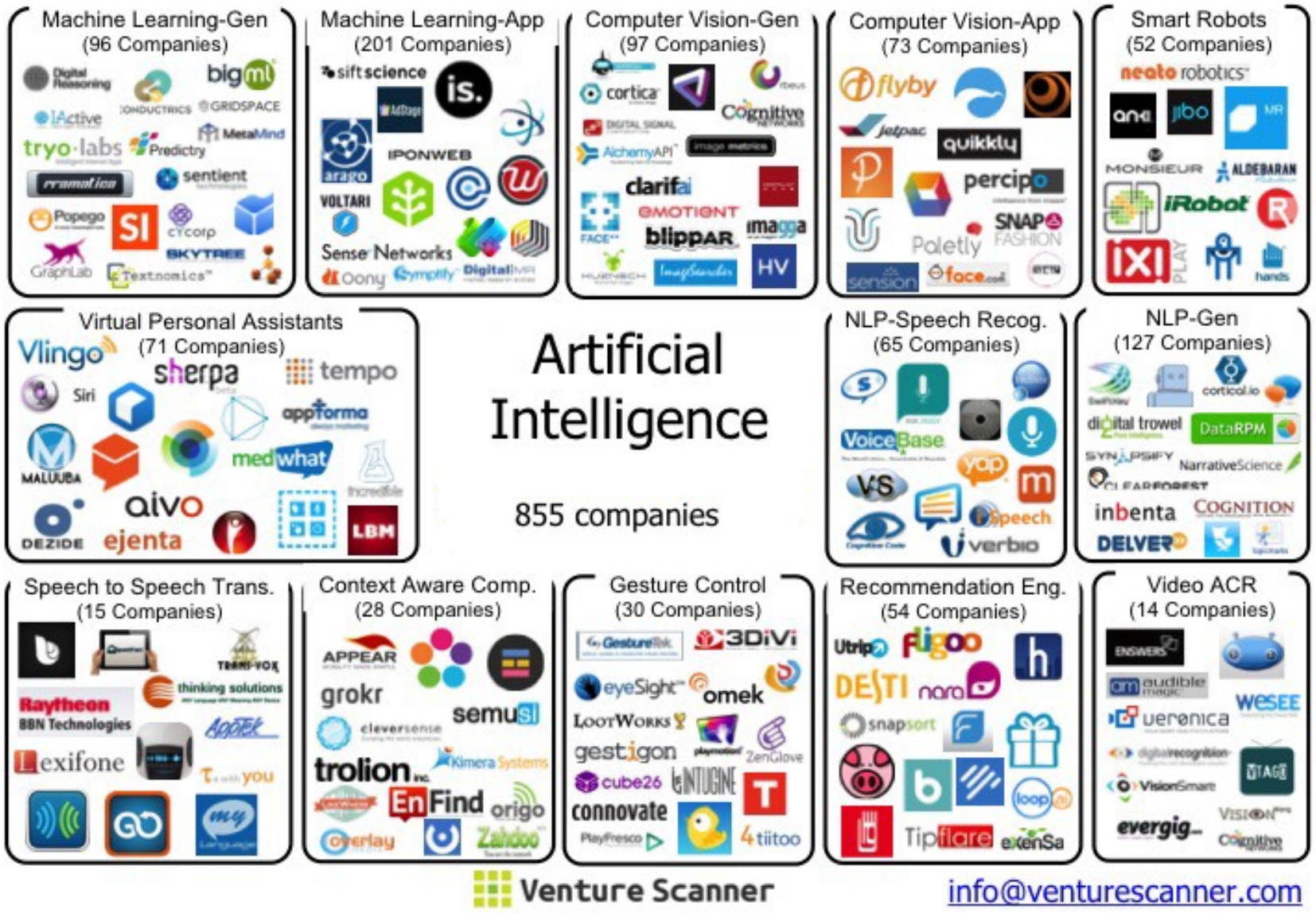
"black cat is sitting on top of suitcase."

ML application





ML application



ML tasks types

1. Supervised learning:
    - Classification
    - Regression
    - Learning to rank
    - forecasting
2. Uunsupervised learning:
    - Clustering
    - Association rules learning
    - Outliers detection
3. Reinforcement learning
4. Meta-learning or learning-to-learn
5. Portfolio selection
6. Collaborative filtering
7. …

Approaches and methods

- Regression
- Байесовский вывод
- Decision trees
- Neural networks
- K-nearest neighbors
- Principal component analysis
- Support vector machines
- Genetic algorythms

Tools

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

Links

- [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