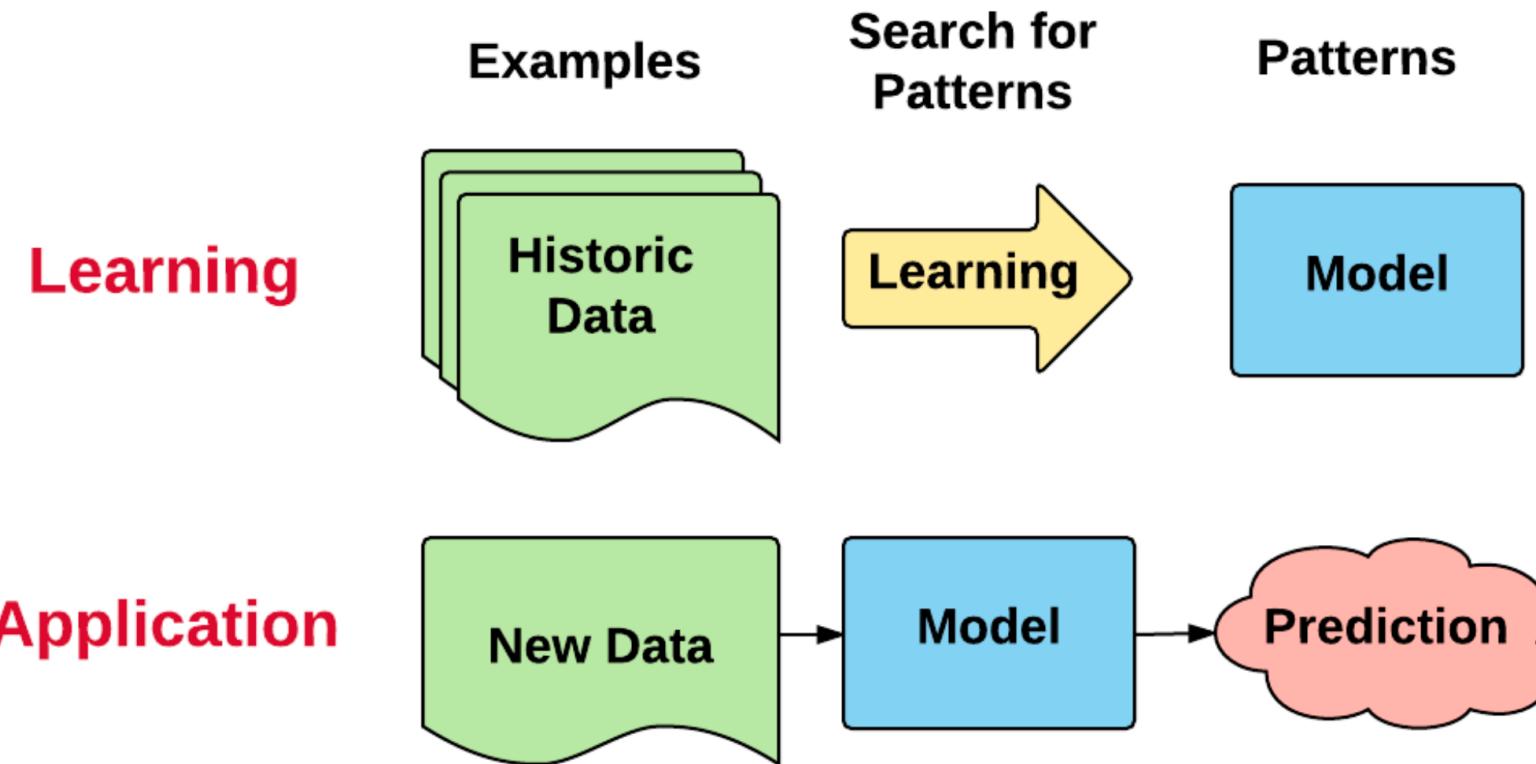

Data Analysis

Prof E. Burnaev
Skoltech

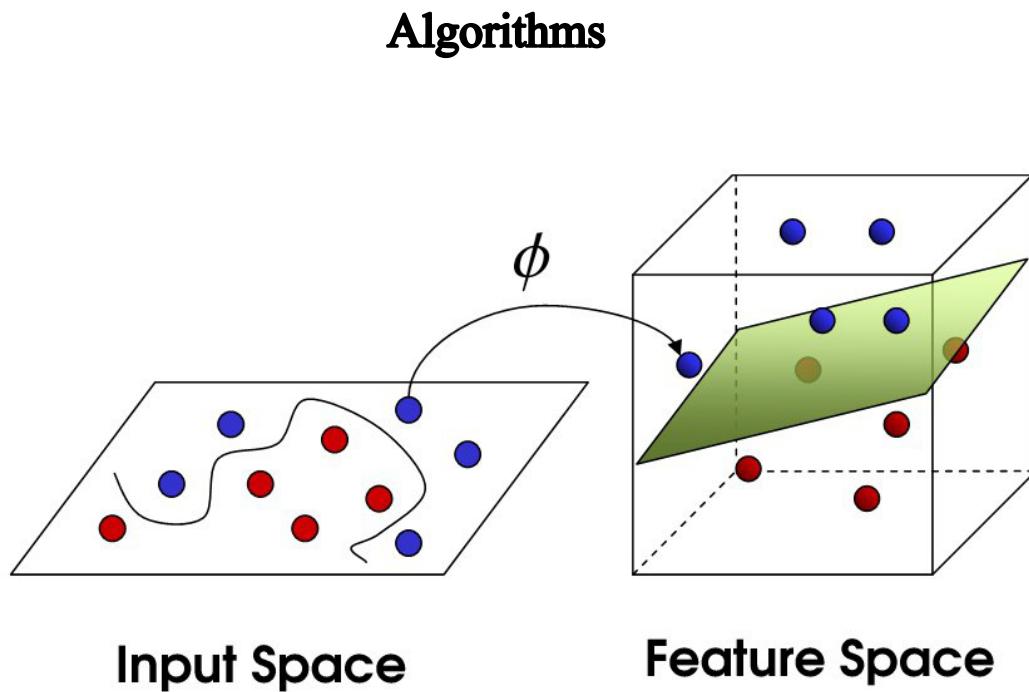
What is behind the term “Predictive Modeling based on data”?



Machine Learning is a set of tools to construct models based on data

Machine learning

Machine learning — set of algorithms for **pattern mining** in data



Structure of AI

- **ARTIFICIAL INTELLIGENCE**

AI is the broadest term, applying to any technique that enables computers to mimic human intelligence, using logic, if-then rules, decision trees, and machine learning (including deep learning)

- **MACHINE LEARNING**

The subset of AI that includes abstruse statistical techniques that enable machines to improve at tasks with experience. The category includes deep learning

- **DEEP LEARNING**

The subset of machine learning composed of algorithms that permit software to train itself to perform tasks, like speech and image recognition, by exposing multilayered neural networks to vast amounts of data

Evolution of AI

Neural networks, genetic algorithms

First AI problems
seminar

Simple AI systems

Expert systems
development

Machine learning
development

Deep learning



1956 yr.

'50-'70 yrs.

'80 yr.

'90 yr.

Beginning 21c.

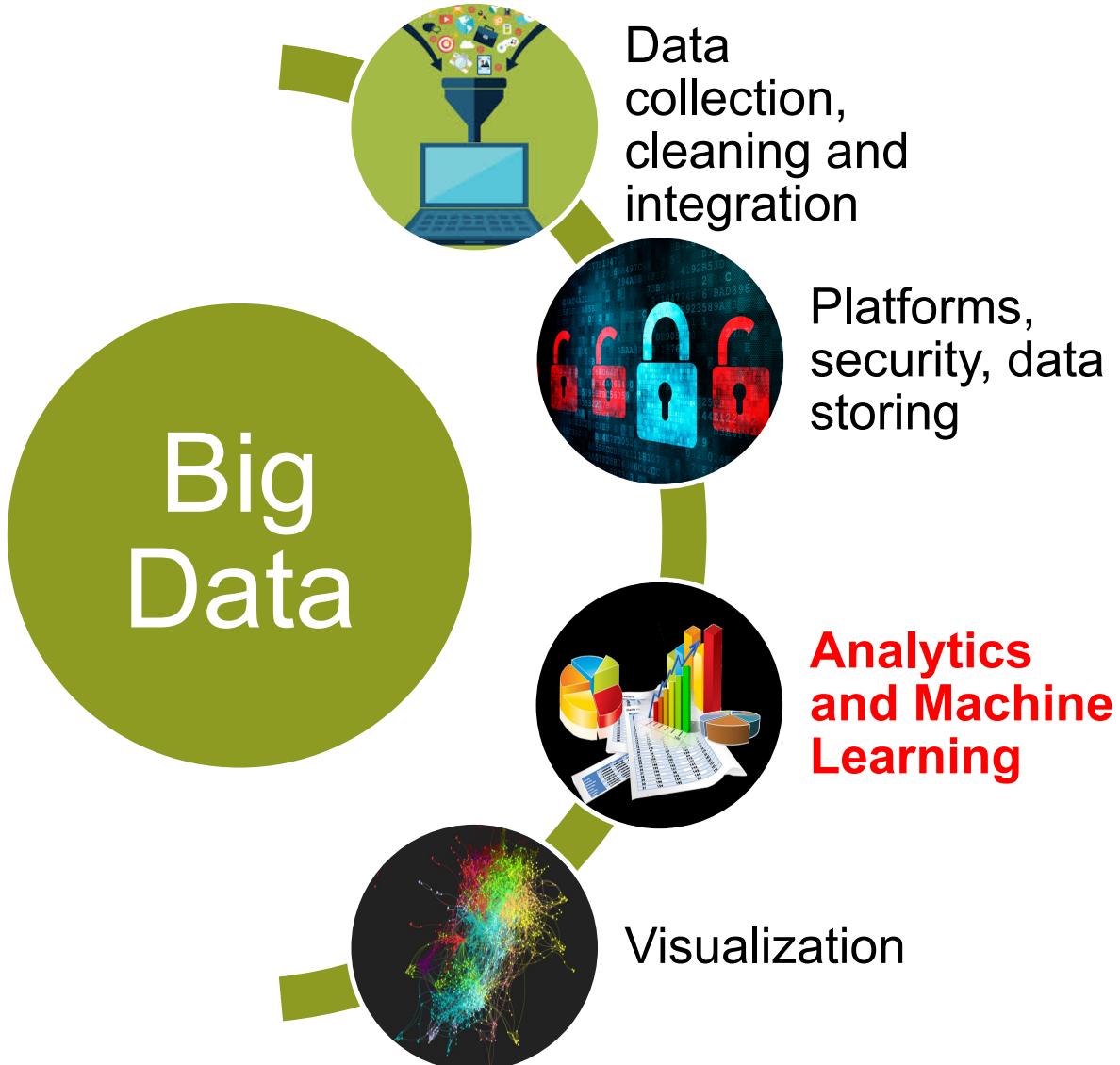
Robot-psychotherapist ELIZA

- ➔ Named with honor of Eliza Doolittle (Show, Pygmalion)
- ➔ Match words with templates
 - «I have the **headache**»
 - «Why are saying you have the **headache?**»
- ➔ In awkward situations answers “I understand”

Detection of failures in technical systems

Human-level precision
of image recognition

Buzzwords: what is Big Data?



→ Big data benefits

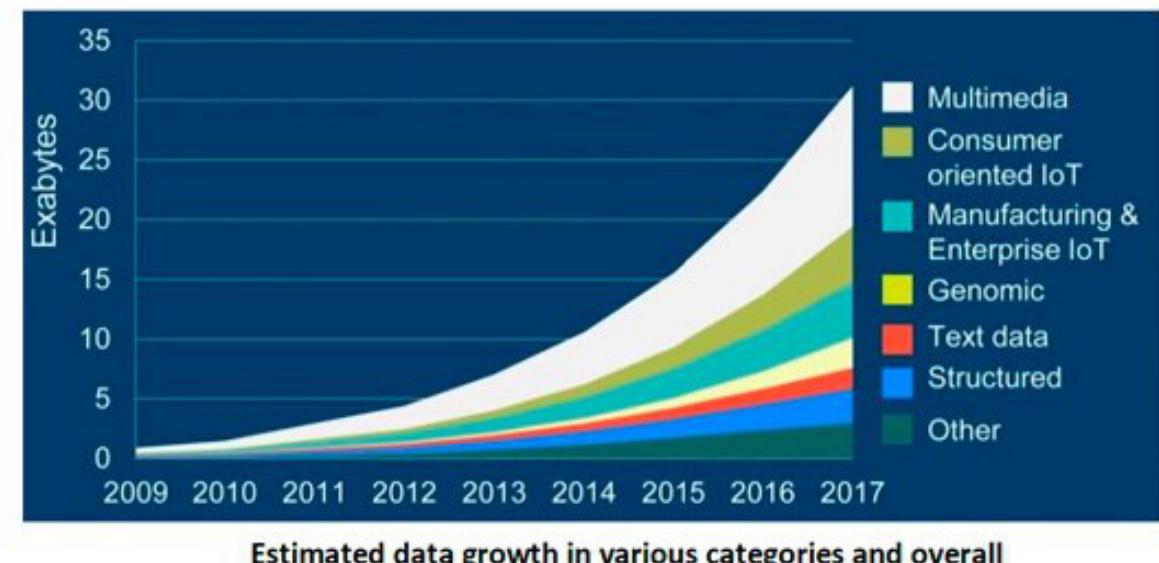
- More data – better accuracy
- The quantity turns into quality

→ Growth of computing powers

- HPC
- GPU

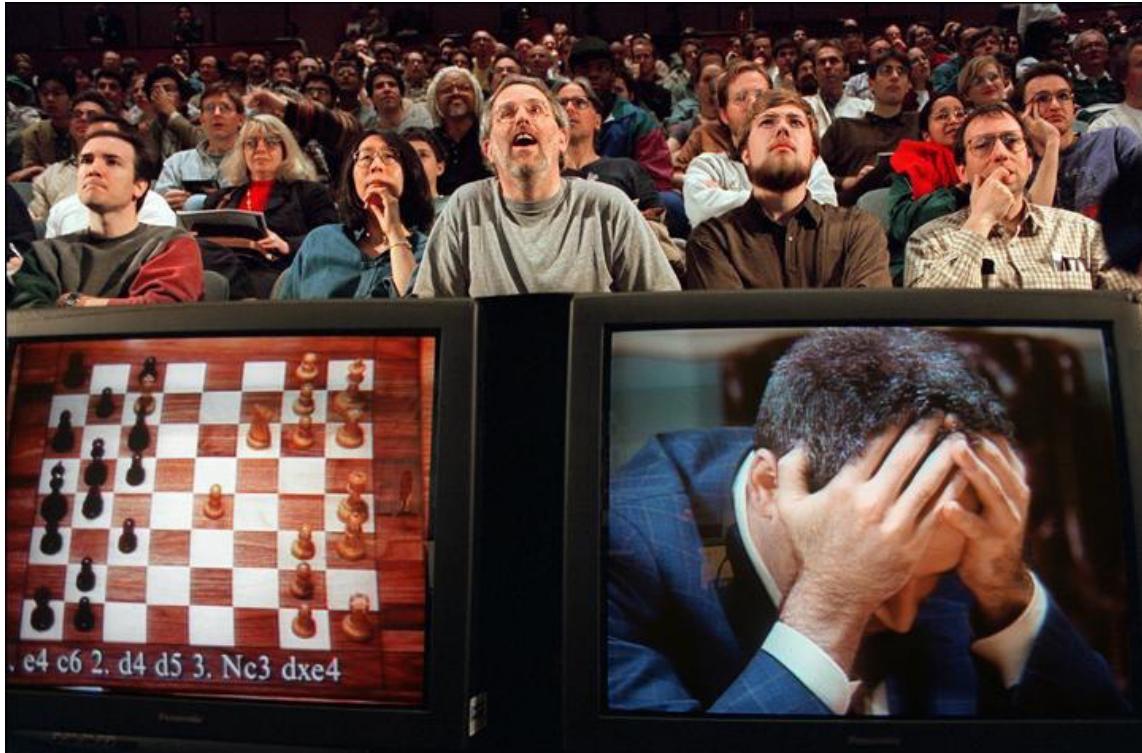
→ Problems

- Calculation-greedy
- Huge storage systems



Deep Blue

- Choose the best move in chess game
- Listing of moves, knows best openings and endgames
- **1997 yr.** — victory over Kasparov



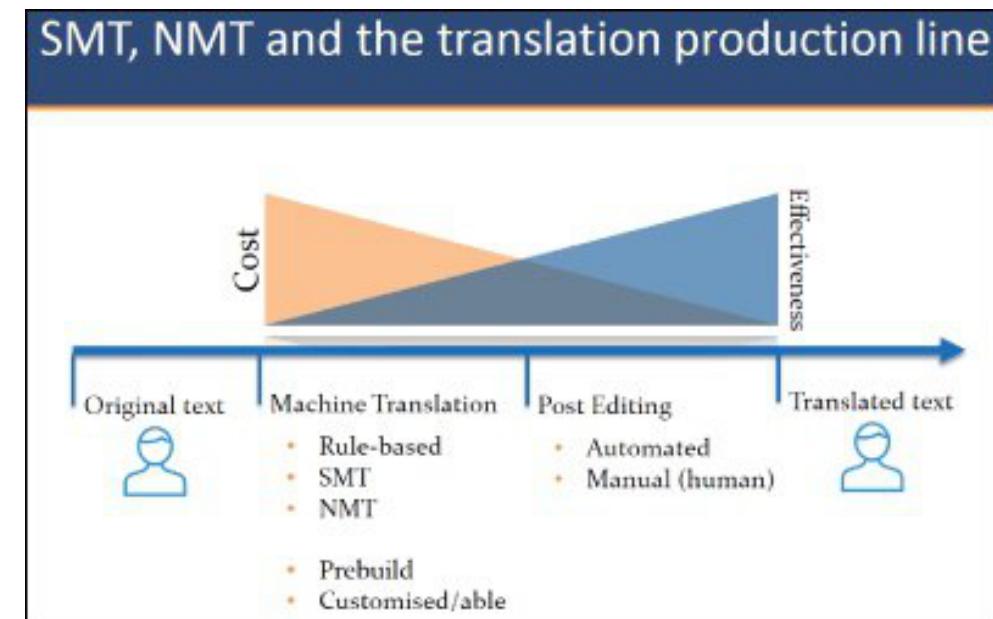
IBM Watson

- **Question-answering system**, machine learning based system
- Extracting answers from big structured (WordNet) and unstructured (Wikipedia) databases
- **February 2011** — winning in «Jeopardy!» game



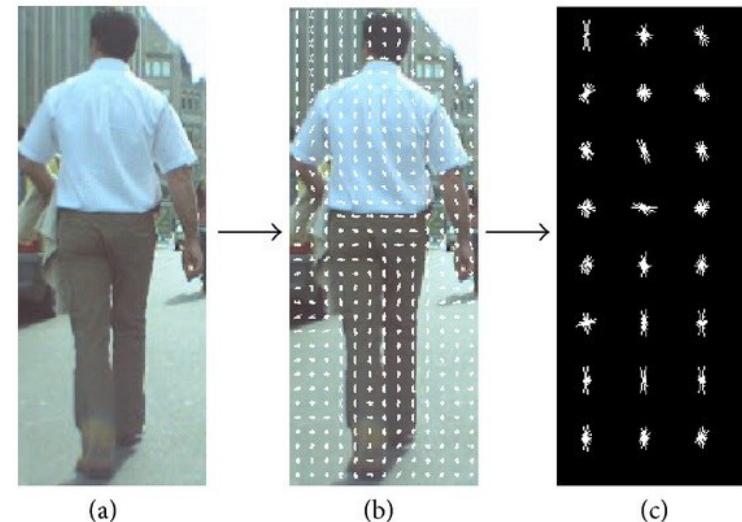
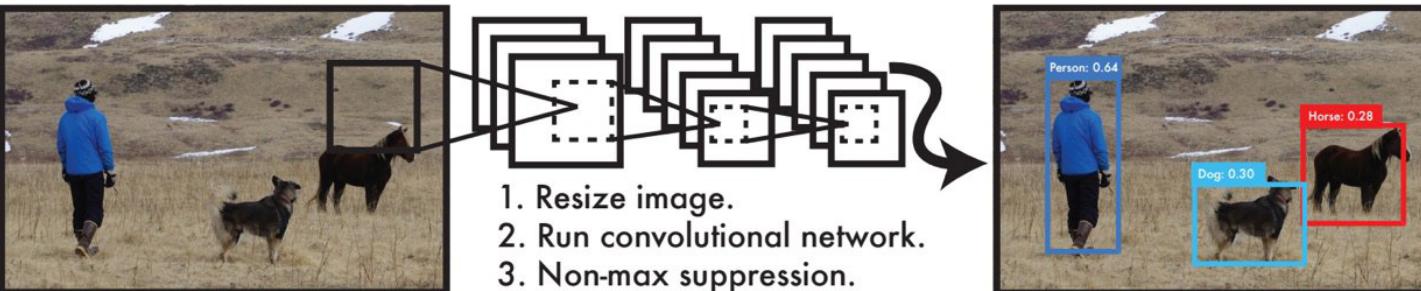
AI and translation

- **1950 – 1980 yrs.** — rule-based models
 - Different proposals of rule-based algorithms
 - Unfortunately, it needs too many particular examples
- **1990 – 2000 yrs.** — statistical machine-based translation
 - Model is trained on bilingual corpus
 - Includes a lot of heuristics
- **2016 yr.** — neural-based translation
 - End-to-end models



AI and object detection

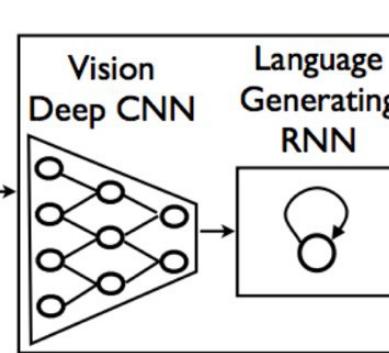
- **1970 – 1990 yrs.** — model based on similarity measure
 - Investigation of human image recognition as algorithmic process
 - Too sensitive for environment conditionals (lighting, texture, etc.)
- **2000s** — key-points detection based methods
 - Improved accuracy compared to previous approaches
 - From image to bag-of-features approach
- **since ~2012** — deep learning
 - Convolutional neural networks
 - Algorithm takes as output raw images
 - Better than human precision



Deep learning

Deep learning — machine learning methods with

- auto-extraction features by
- composition of linear mappings and nonlinearities



A group of people shopping at an outdoor market.
There are many vegetables at the fruit stand.

Deep learning is not silver bullet

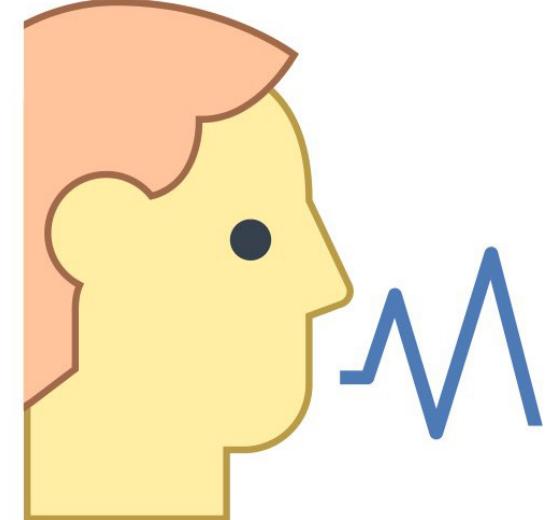
Deep Learning fits to several cases:



Economic growth has slowed down in recent years .

Das Wirtschaftswachstum hat sich in den letzten Jahren verlangsamt .
Economic growth has slowed down in recent years .

La croissance économique s' est ralentie ces dernières années .



Images

- What is on the image?
(Recognition)
- Where is the object?
(Detection)

Text

- Summarize text
- Translation

Sound

- Speech recognition
- Sentiment analysis

Image captioning

Describes without errors



A person riding a motorcycle on a dirt road.

Describes with minor errors



Two dogs play in the grass.

Somewhat related to the image



A skateboarder does a trick on a ramp.



A group of young people playing a game of frisbee.



Two hockey players are fighting over the puck.



A little girl in a pink hat is blowing bubbles.

AlphaGo

→ Neural network, playing the Go game

At the Go game there is 10^{171} combinations of possible positions (at the chess 10^{43} , atoms 10^{80})

→ March'16 — victory over the world champion

→ Learned with both: history of human games and self-playing

