Artificial general intelligence introduction

Antal Péter, Bolgár Bence

Számítógépes orvosbiológiai munkacsoport

Mesterséges Intelligencia kutatócsoport

BME, VIK, Méréstechnika és információs rendszerek tanszék

MŰEGYETEM 1782

Méréstechnika és Információs Rendszerek Tanszék

COMBINE Computational Biomedicine Workgroup

Course info

- Course site
  - https://portal.vik.bme.hu/kepzes/targyak/VIMIAV22/
  - http://www.mit.bme.hu/oktatas/targyak/VIMIAV22/
- Lecturer
  - Bence Bolgár, bolgar@mit.bme.hu
  - Péter Antal, antal@mit.bme.hu
- Schedule
  - Wednesday, 12.30-14.00, QBF09
- Contact hour
  - Tuesday, 10.00-11.00, IE.423 (building I, wing E, 2nd floor)
- Books
  - Russell, S.J. and Norvig, P., 2002. Artificial intelligence: a modern approach (AIMA)
  - Magyar fordítás: MI Almanach
    - http://project.mit.bme.hu/mi_almanach/
- Slides
  - At course site

Academic calendar

- Last day of classes: 13 December (Friday)
- Repeat week (resits and late submission of home assignments): 16-20 December (Monday-Friday)
- Exams begin: 2 January (Thursday)
- Duration of examination period 20 working days
- Last day of examination period: 29 January (Wednesday)

Requirements

- Grading:
  - Two homeworks 50-50%.
    - Review, essay, programming..

Goal of the course: nature of intelligence
(from physics through computation to society and beyond)

- History of AI
- History of cognitive science
- Computational models
- Computational linguistics
- Bayesian AI
- Logic: truth and proof
- Machine learning, machine teaching
- Heuristics, creativity, scientific discovery
- Neuroscience-inspired AI (neurobiology, deep neural networks)
- Explainable AI, Provably beneficial AI, existential threat of AI

Personal research interest in AGI

- 1992<: neural nets, human learning (teaching)
- 1993: visual/diagrammatic reasoning
- 1995: non-symbolic(analogic)-vs-symbolic reasoning
- 1996: cellular automatons+emergence
- 1997<, causality
- 1997<, neural-causal hybrid systems, decision support systems,..
- Why I.: belief systems, causality research, biases in human beliefs
- Why II.: reward-system/ reward deficit syndromes in human
    - 2009< depression research
- Why III.: hormonal-neural circuitry of teaching
    - ~2010: genetic variations/evolution of oxytocin, dogs

Overview of today lecture

- What is artificial intelligence?
- What is intelligence?
  - Myths, misconceptions, analogies, models..
- Factors of intelligence: data, computing power, efficient learning
- Moore’s law
- The knowledge era
- The data age
- Examples of intelligent solutions

The Hitchhiker's Guide to the Galaxy: Phases of civilizations

1. **'How can we eat?'**
2. **'Why do we eat?'**
3. **'Where shall we have lunch?”**

1. **'How can we define intelligence?'**
2. **'Why do we need intelligence?'**
3. **'Where shall we have intelligent solutions?”**

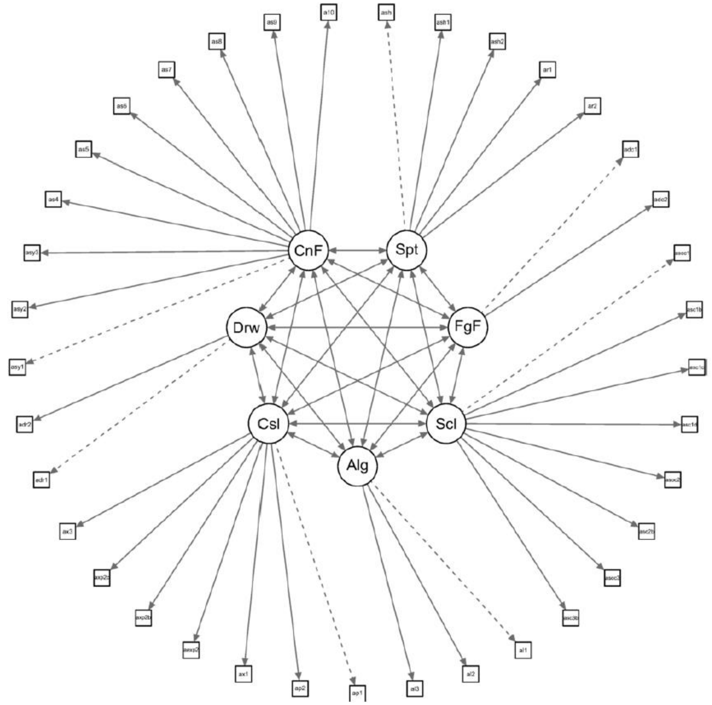
What is intelligence?

- What is X?
  - X= fire, light, life, intelligence, creativity, consciousness, ethics
- What is intelligence?
  - Animal intelligence
  - There are smart people ;-)
  - Computation: complexity theories (there are hard problems)
- Application areas
  - Expert systems
  - Data mining (text-mining)
  - Game playing
  - Self-driving car, advanced driver assistance..
  - ...

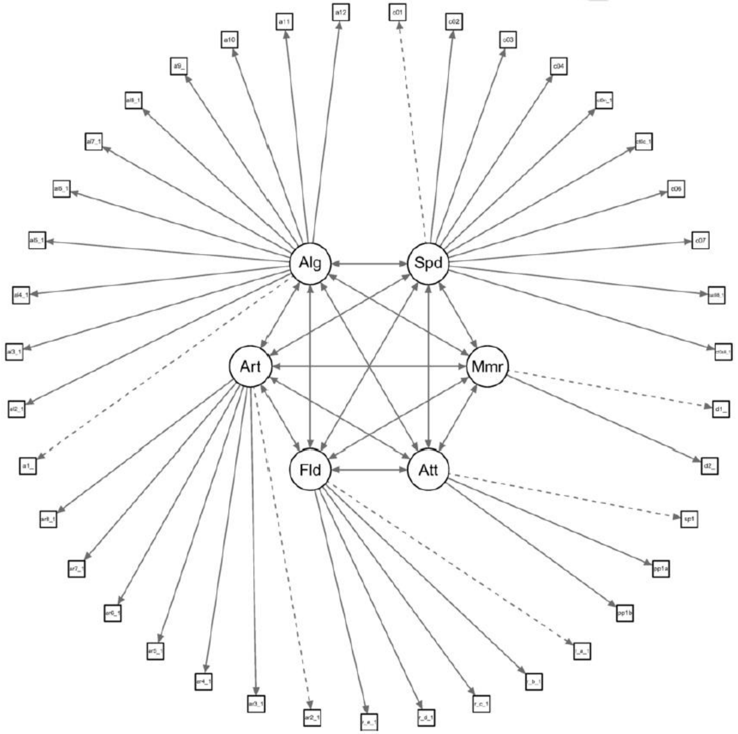
Animal intelligence

- **A Pigeon Solves the Classic Box-and-Banana Problem**
  - https://www.youtube.com/watch?v=mDntbGRPeEU
- **Are Crows the Ultimate Problem Solvers? | Inside the Animal Mind**
  - **https://www.youtube.com/watch?v=cbSu2PXOTOc**
- **Causal understanding of water displacement by a crow**
  - https://www.youtube.com/watch?v=ZerUbHmuY04

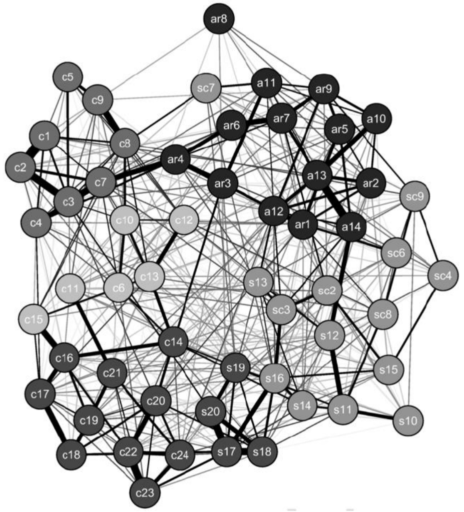
Dimensions of human intelligence



Spt = spatial ability; FgF = figural fluency; Scl = social reasoning; Alg = algebraic reasoning;
Csl = causal reasoning; Drw = drawing ability; CnF = conceptual fluency.



Spd = speed; Mmr = memory; Att = attention; Fld = fluid reasoning; Art = arithmetic reasoning; Alg = algebraic reasoning.

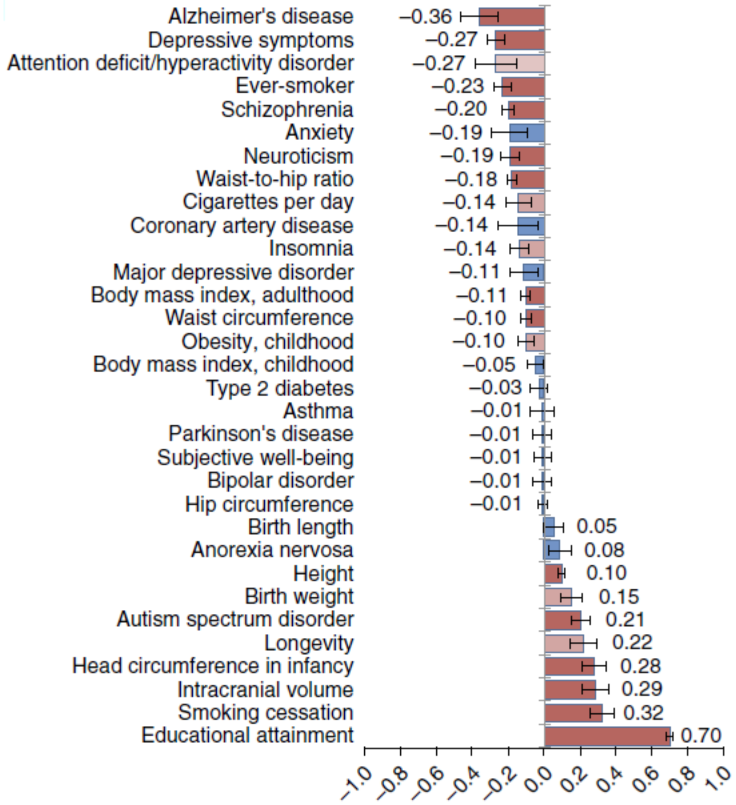


Each cluster represents a dimension: cluster 1 = arithmetic reasoning; cluster 2 = concepts n.1; cluster 3 = concepts n.2; cluster 4 = sentence completion; cluster 5 = concepts n.3

Golino, H.F. and Demetriou, A., 2017. Estimating the dimensionality of intelligence like data using Exploratory Graph Analysis. *Intelligence*.

Genetic factors and traits of intelligence

rsID	Annotation
rs2490272	*FOXO3* intronic
rs9320913	Intergenic
rs10236197	*PDE1C* intronic
rs2251499	Intergenic
rs36093924	*CYP2D7* ncRNA_intr
rs7646501	Intergenic
rs4728302	*EXOC4* intronic
rs10191758	*ARHGAP15* intronic
rs12744310	Intergenic
rs66495454	*NEGR1* upstream
rs113315451	*CSE1L* intronic
rs12928404	*ATXN2L* intronic
rs41352752	*MEF2C* intronic
rs13010010	*LINC01104* ncRNA_intr
rs16954078	*SKAP1* intronic
rs11138902	*APBA1* intronic
rs6746731	*ZNF638* intronic
rs6779302	Intergenic



Sniekers, Suzanne, et al. "Genome-wide association meta-analysis of 78,308 individuals identifies new loci and genes influencing human intelligence." *Nature Genetics* (2017).

What is AI?

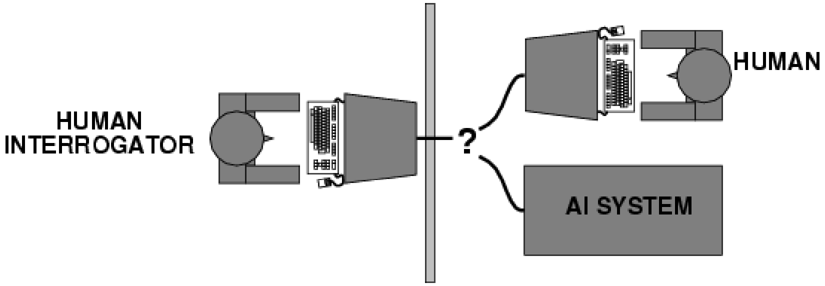
AI approaches can be grouped as follows:

Thinking humanly	Thinking rationally
Acting humanly	Acting rationally

AIMA

Acting humanly: Turing Test

- Turing (1950) "Computing machinery and intelligence":
- "Can machines think?" → "Can machines behave intelligently?"
- Operational test for intelligent behavior: the Imitation Game



- Predicted that by 2000, a machine might have a 30% chance of fooling a lay person for 5 minutes
- Anticipated all major arguments against AI in following 50 years
- Suggested major components of AI: knowledge, reasoning, language understanding, learning

AIMA

Thinking humanly: cognitive modeling

- 1960s "cognitive revolution": information-processing psychology
-
- Requires scientific theories of internal activities of the brain
-
- -- How to validate? Requires

    1) Predicting and testing behavior of human subjects (top-down)

    or 2) Direct identification from neurological data (bottom-up)

- Both approaches (roughly, Cognitive Science and Cognitive Neuroscience) are now distinct from AI, but➔
- Hassabis, Demis, et al. "Neuroscience-inspired artificial intelligence." *Neuron* 95.2 (2017): 245-258.

AIMA

Thinking rationally: "laws of thought"

- Aristotle: what are correct arguments/thought processes?
-
- Several Greek schools developed various forms of *logic*: *notation* and *rules of derivation* for thoughts; may or may not have proceeded to the idea of mechanization
-
- Direct line through mathematics and philosophy to modern AI
-
- Problems:
  1. Not all intelligent behavior is mediated by logical deliberation
  2. What is the purpose of thinking? What thoughts should I have?

  ➔ **(Symbolic) reasoning is mainly for collaborative thinking**!

AIMA

Acting rationally: rational agent

- Rational behavior: doing the right thing
-
- The right thing: that which is expected to maximize goal achievement, given the available information
-
- Doesn't necessarily involve thinking – e.g., blinking reflex – but thinking should be in the service of rational action
-

AIMA

Rational agents

- An agent is an entity that perceives and acts
-
- This course is about designing rational agents
-
- Abstractly, an agent is a function from percept histories to actions:
-

$$[f: \mathcal{P}^{\star} \rightarrow \mathcal{A}]$$

- For any given class of environments and tasks, we seek the agent (or class of agents) with the best performance
-
- Caveat: computational limitations make perfect rationality unachievable

  → design best program for given machine resources

AIMA

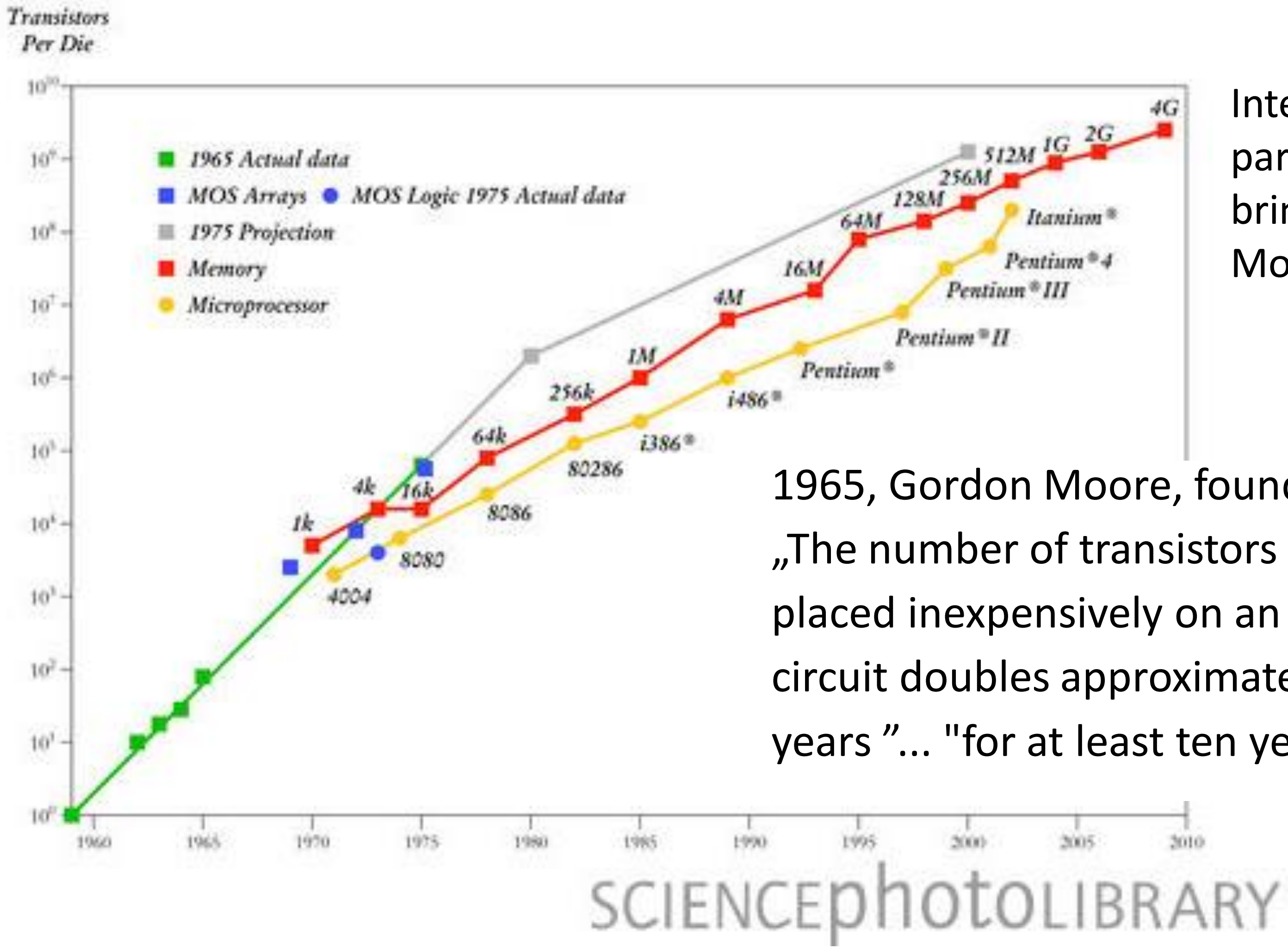
Why do we need AI?

- Understanding human cognition
- Supporting and complementing human expertise
- No choice: data & knowledge exceeded the scope of human cognition
- Instead of human experts, it is
  - slightly cheaper ;-), scalable, multiplicable,..
- **Curiosity + optimism!**

Factors behind intelligence explosion

- Computation
  - Moore’s law
- Data
  - Big data age
- Knowledge
  - Publications, knowledge bases,…
- Technologies
  - Artificial intelligence? Language understanding?
  - Machine learning? Deep learning?

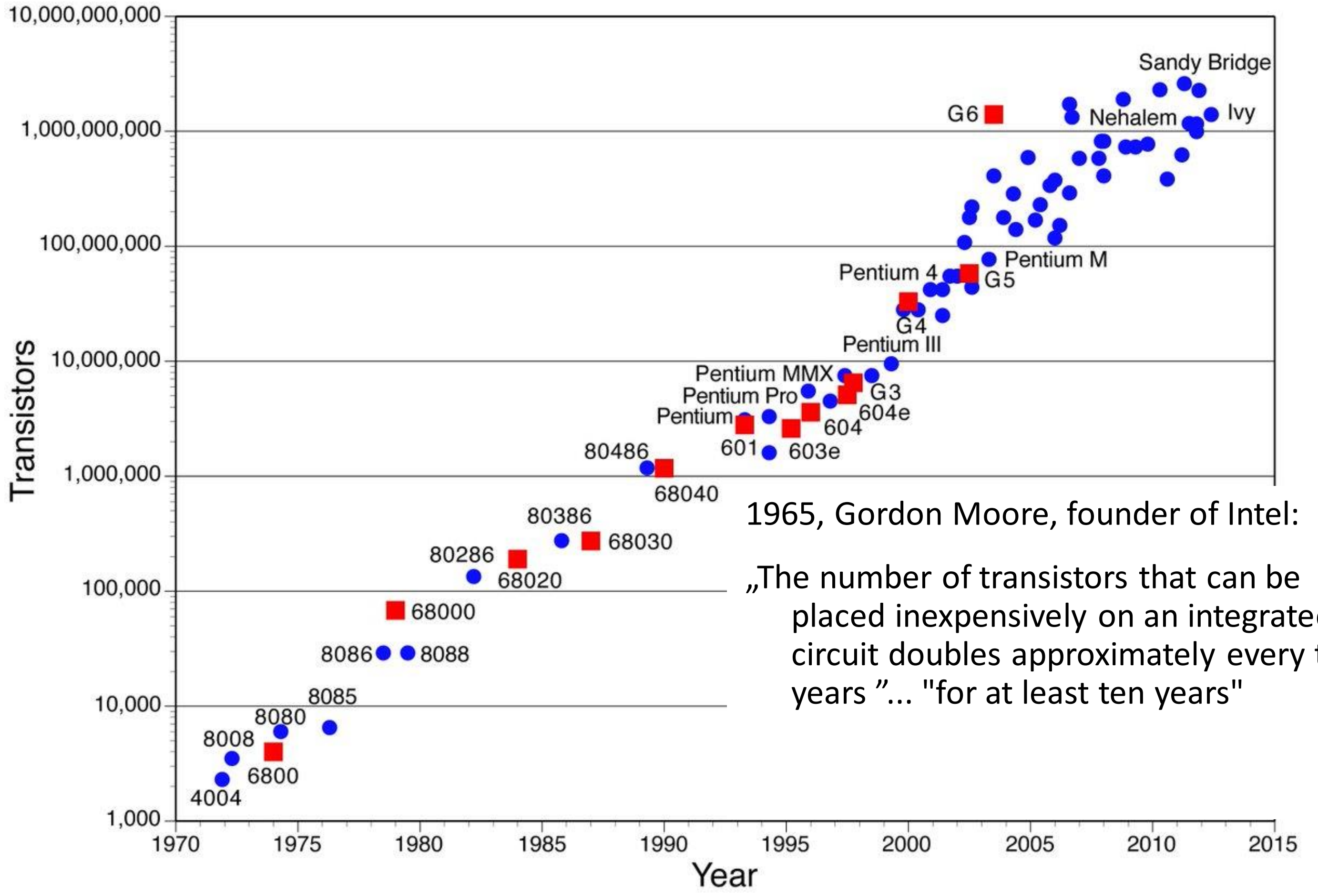
Computing power: Moore’s Law



Integration and parallelization wont bring us further. End of Moor’s law?

1965, Gordon Moore, founder of Intel: „The number of transistors that can be placed inexpensively on an integrated circuit doubles approximately every two years ”... "for at least ten years"

Computing power: Moore's Law



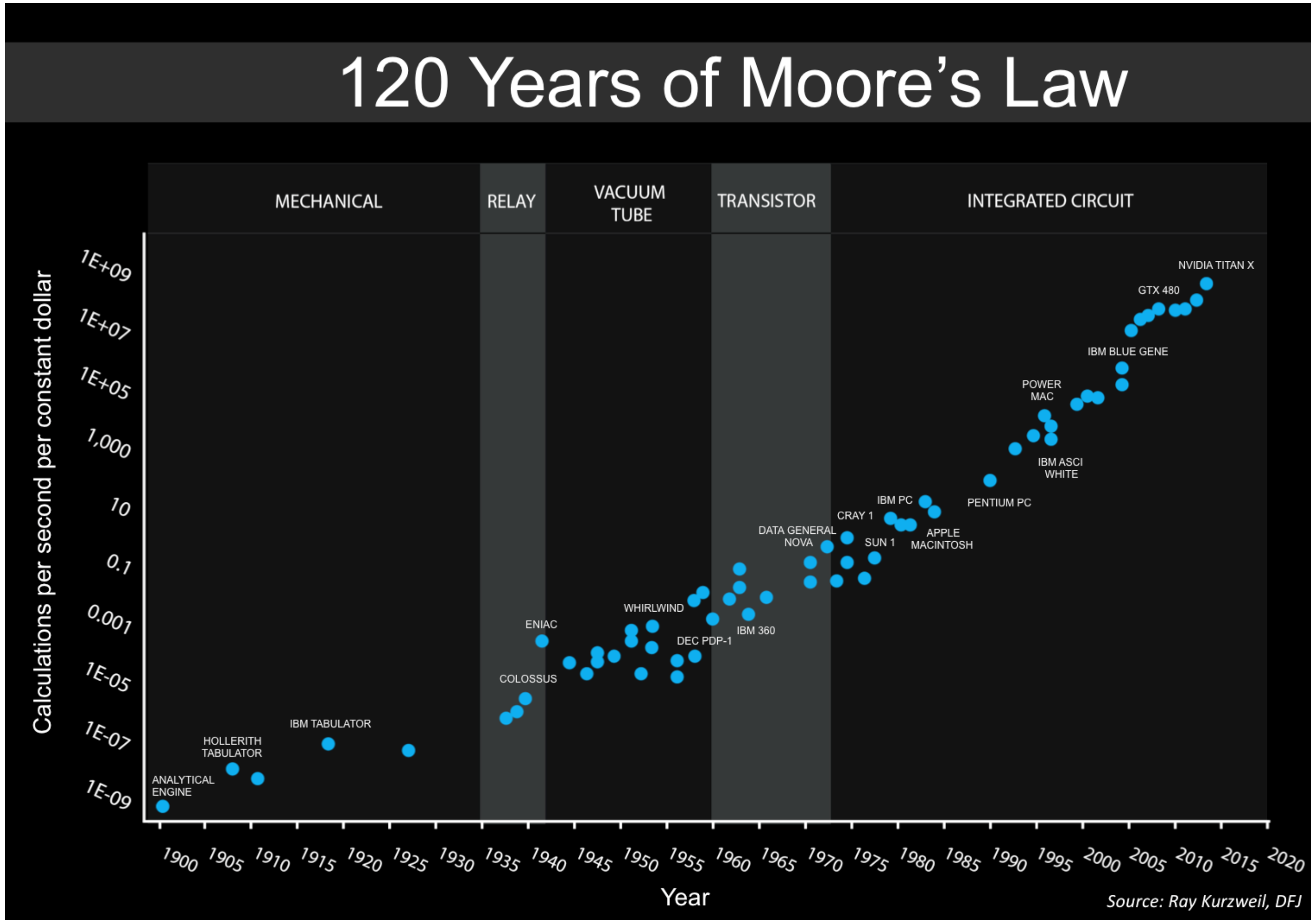
1965, Gordon Moore, founder of Intel:

„The number of transistors that can be placed inexpensively on an integrated circuit doubles approximately every two years ”... "for at least ten years"

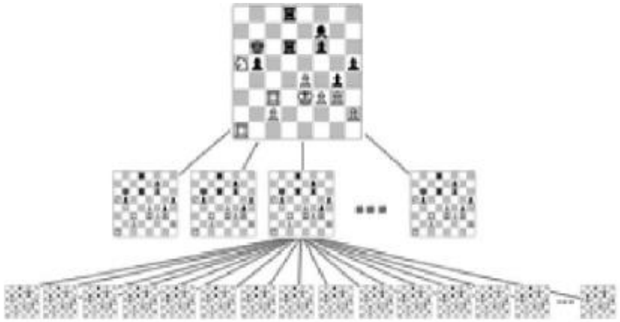
- 10 µm – 1971
- 6 µm – 1974
- 3 µm – 1977
- 1.5 µm – 1982
- 1 µm – 1985
- 800 nm – 1989
- 600 nm – 1994
- 350 nm – 1995
- 250 nm – 1997
- 180 nm – 1999
- 130 nm – 2001
- 90 nm – 2004
- 65 nm – 2006
- 45 nm – 2008
- 32 nm – 2010
- 22 nm – 2012
- **14 nm – 2014**
- **10 nm – 2017**
- 7 nm – *~2019*
- **5 nm** – *~2021*

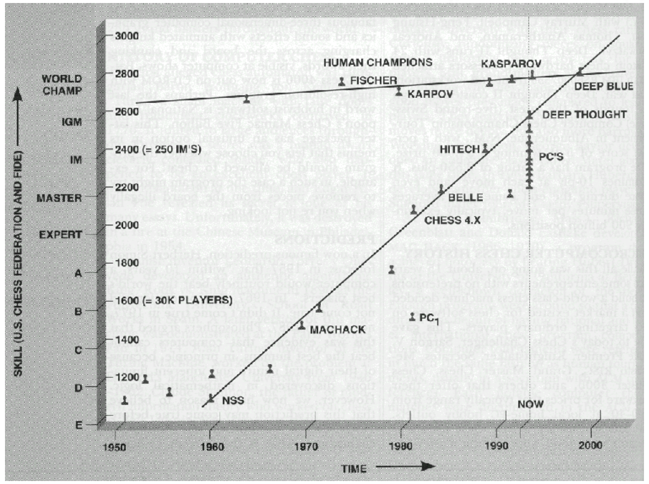
2012: single atom transistor (~0.1n, 1A)

Moore’s law: calculation/$



Computing power and search: performance in chess

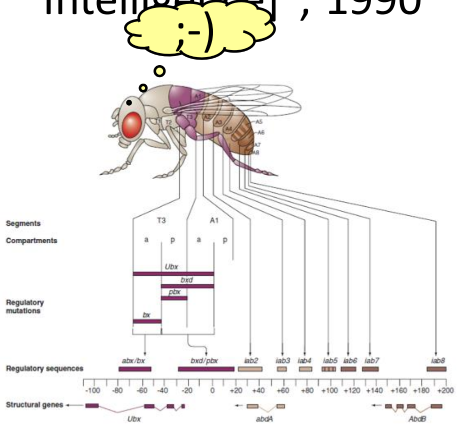




#	Név	Élőpont
1	**SugaR XPrO 1.2 64-bit 4CPU**	**3415**
2	**Komodo 11.2 64-bit 4CPU**	**3402**
3	**Houdini 5.01 64-bit 4CPU**	**3382**
	IBM Deep Blue (1997)	-

http://www.computerchess.org.uk/ccrl/4040/

J.McCarthy: "Chess as the Drosophila of AI. [Artificial Intelligence]", 1990

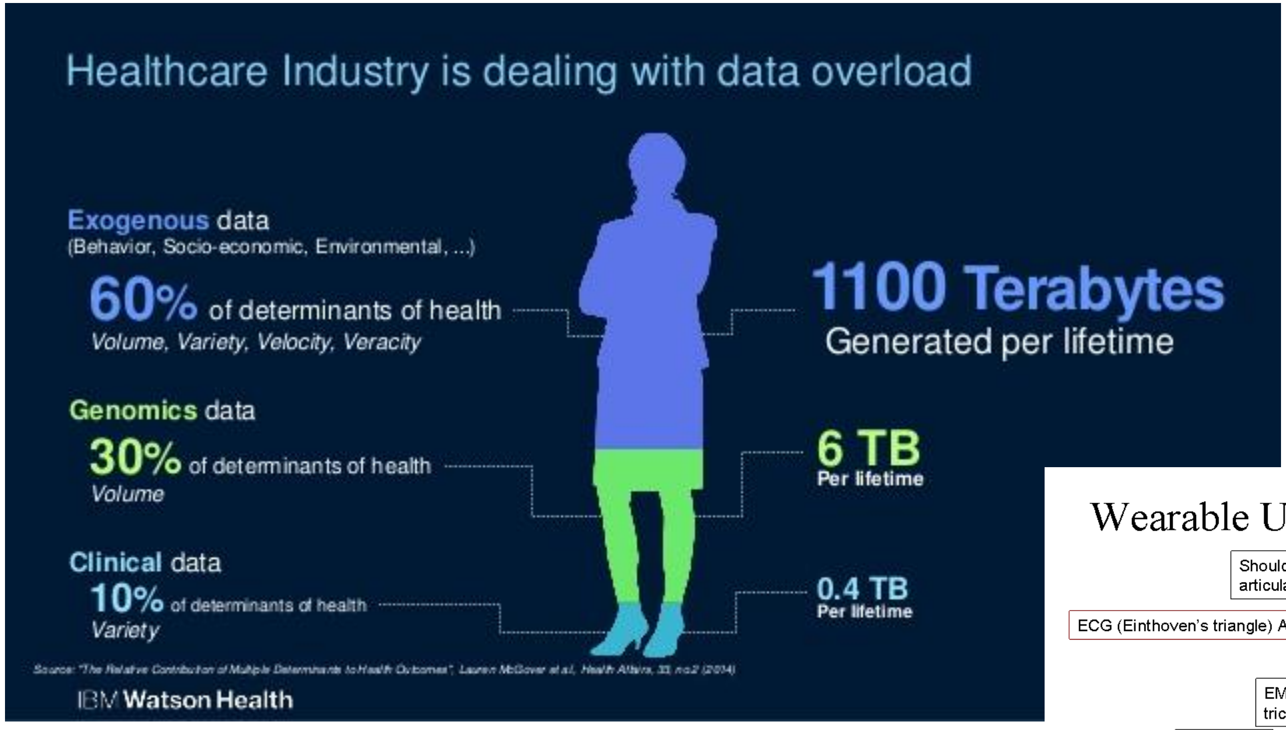


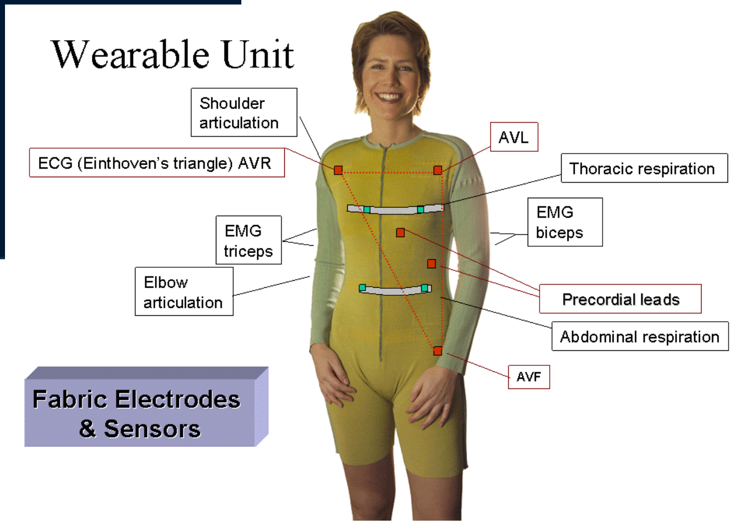
Chess as the Drosophila of AI

- Chase&Simon: Perception in chess, 1973
- Chi: Knowledge structures and memory development, 1978
- Schneider: Chess expertise and memory for chess positions, 1993
- ...
- Simons: How experts recall chess positions, 2012

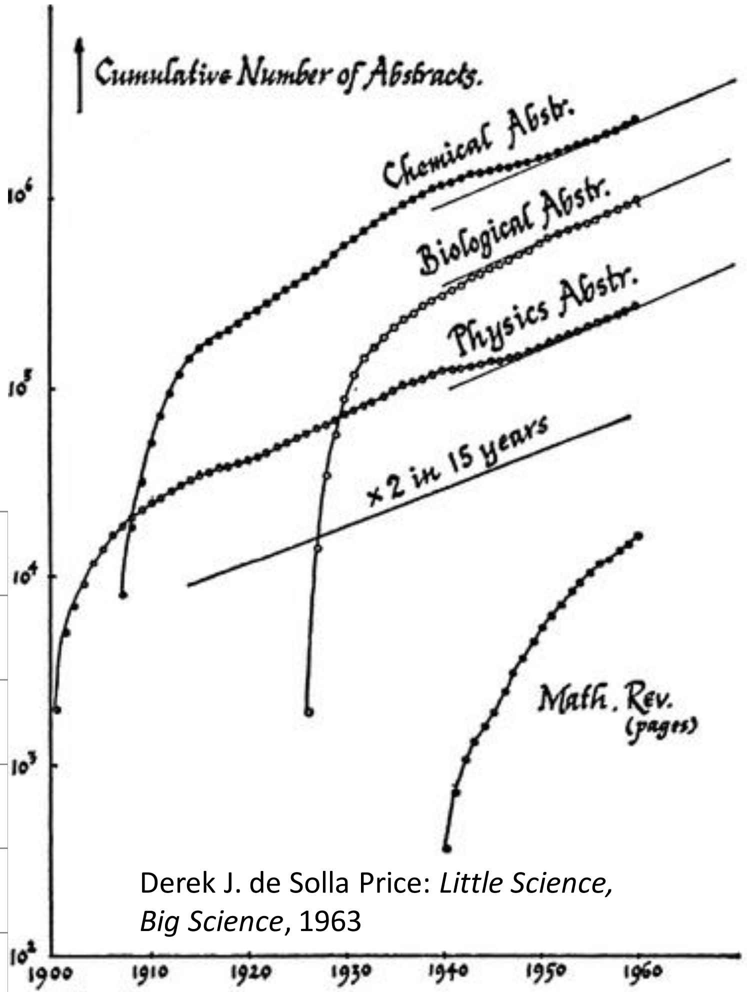
- Mérő László: Észjárások, 1990
  - Kezdő, haladó, mesterjelölt, nagymester

Data: Big data in life sciences





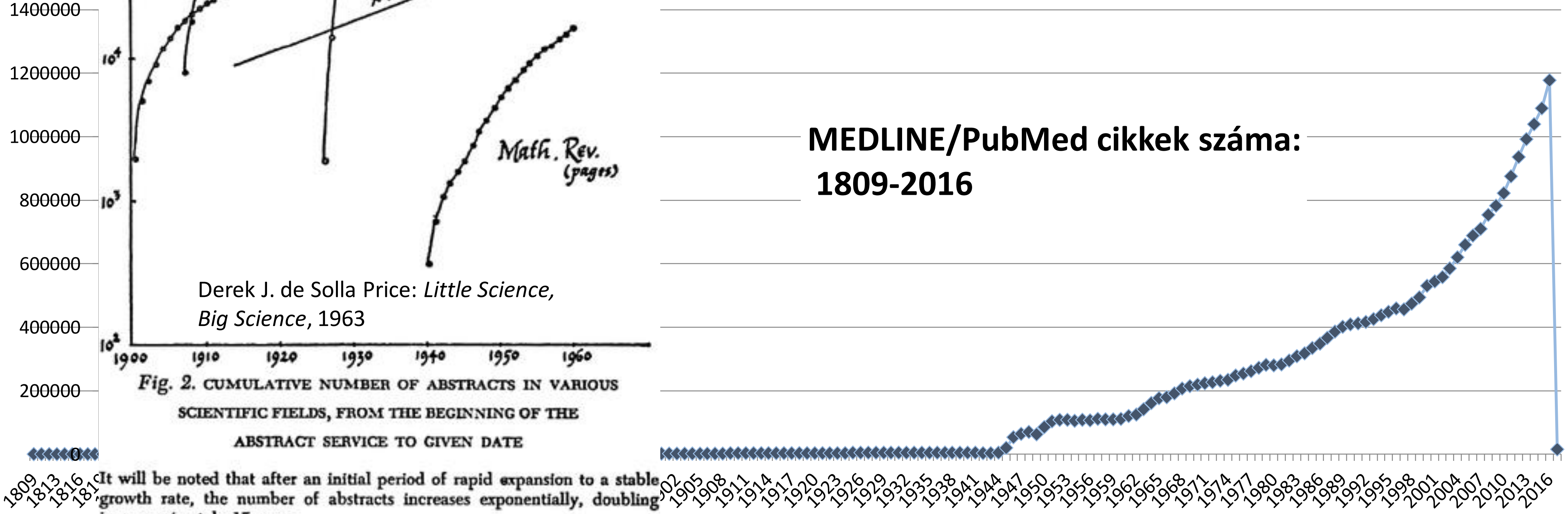
Írottan tárolt intelligencia: cikkek



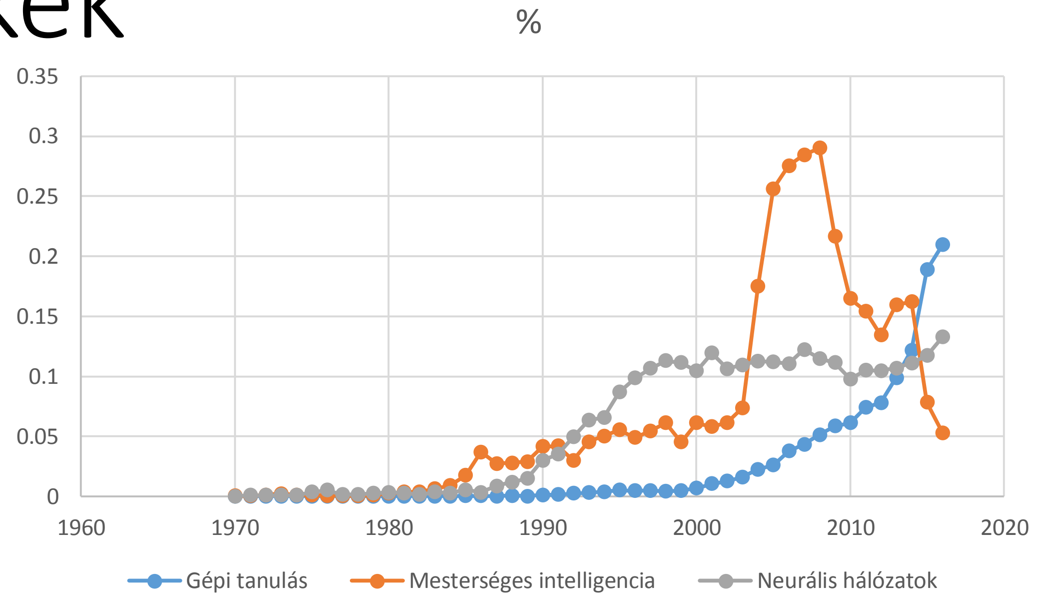
Derek J. de Solla Price: *Little Science, Big Science*, 1963

Fig. 2. CUMULATIVE NUMBER OF ABSTRACTS IN VARIOUS SCIENTIFIC FIELDS, FROM THE BEGINNING OF THE ABSTRACT SERVICE TO GIVEN DATE

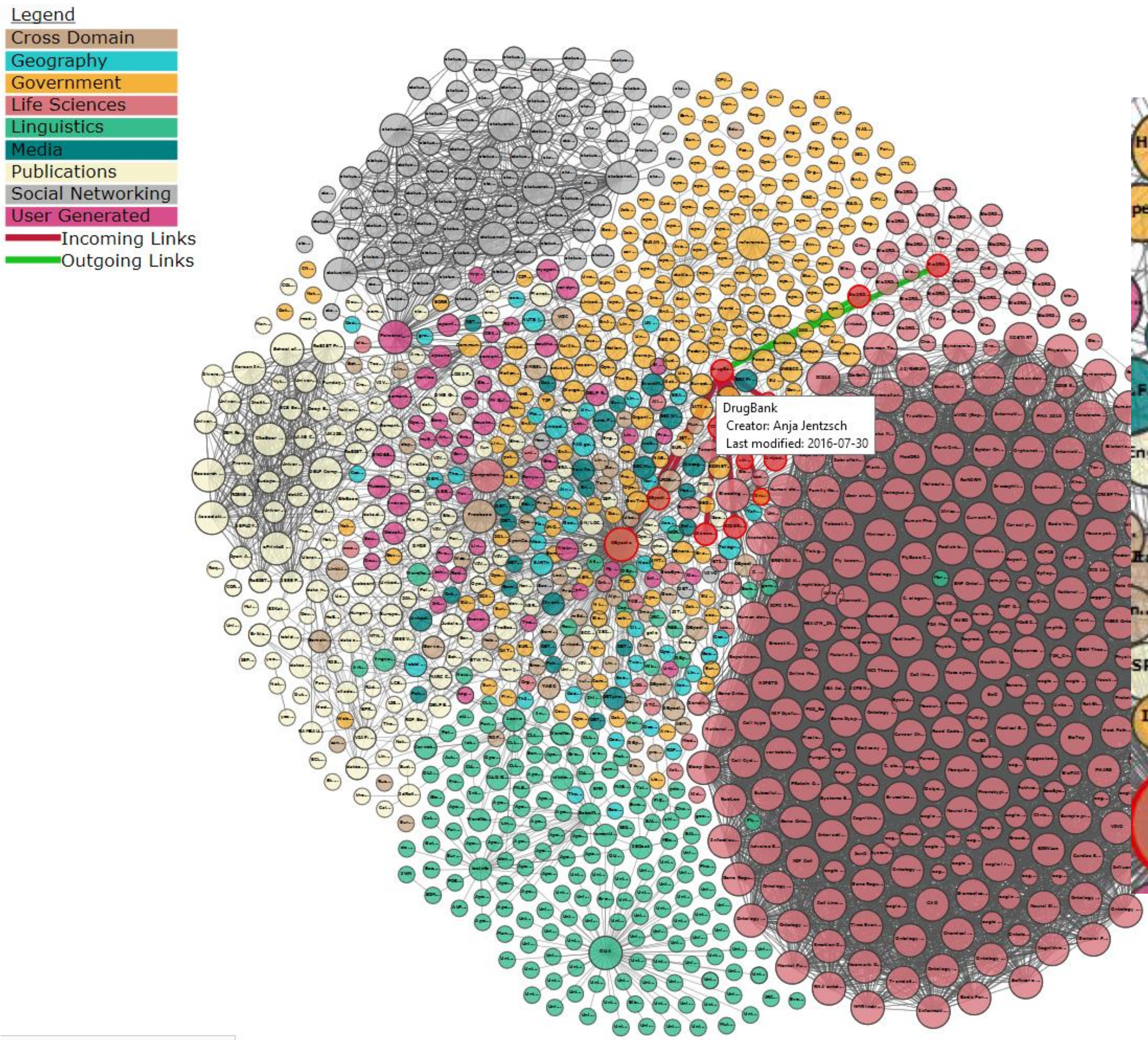
It will be noted that after an initial period of rapid expansion to a stable growth rate, the number of abstracts increases exponentially, doubling in approximately 15 years.

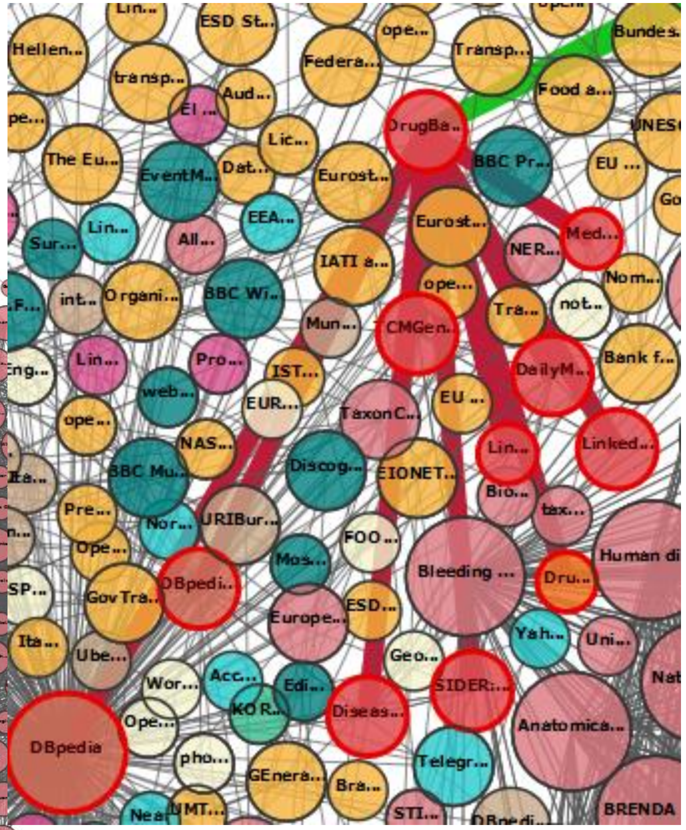


MEDLINE/PubMed cikkek száma: 1809-2016



Knowledge: Linked open data



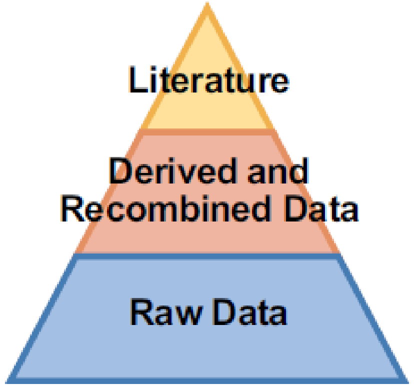


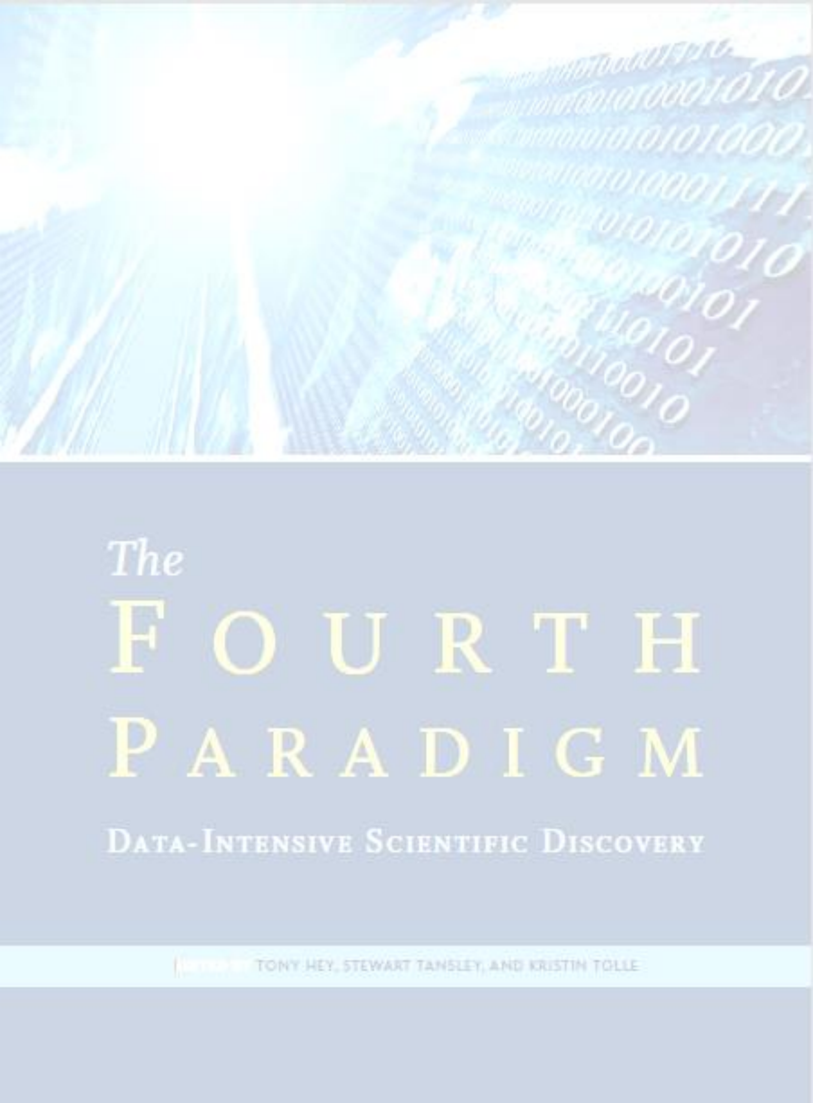
Linking Open Data cloud diagram 2017, by Andrejs Abele, John P. McCrae, Paul Buitelaar, Anja Jentzsch and Richard Cyganiak. http://lod-cloud.net/

E-science, data-intensive science

All Scientific Data Online

- Many disciplines overlap and use data from other sciences
- Internet can unify all literature and data
- Go from literature to computation to data back to literature
- Information at your fingertips for everyone-everywhere
- Increase Scientific Information Velocity
- Huge increase in Science Productivity





Methods: new learning methods

ARTICLE

doi:10.1038/nature16961

Mastering the game of Go with deep neural networks and tree search

David Silver[1]*, Aja Huang[1]*, Chris J. Maddison[1], Arthur Guez[1], Laurent Sifre[1], George van den Driessche[1], Julian Schrittwieser[1], Ioannis Antonoglou[1], Veda Panneershelvam[1], Marc Lanctot[1], Sander Dieleman[1], Dominik Grewe[1], John Nham[2], Nal Kalchbrenner[1], Ilya Sutskever[2], Timothy Lillicrap[1], Madeleine Leach[1], Koray Kavukcuoglu[1], Thore Graepel[1] & Demis Hassabis[1]

LETTER

doi:10.1038/nature14236

Human-level control through deep reinforcement learning

Volodymyr Mnih[1]*, Koray Kavukcuoglu[1]*, David Silver[1]*, Andrei A. Rusu[1], Joel Veness[1], Marc G. Bellemare[1], Alex Graves[1], Martin Riedmiller[1], Andreas K. Fidjeland[1], Georg Ostrovski[1], Stig Petersen[1], Charles Beattie[1], Amir Sadik[1], Ioannis Antonoglou[1], Helen King[1], Dharshan Kumaran[1], Daan Wierstra[1], Shane Legg[1] & Demis Hassabis[1]

AI prehistory

- Philosophy — Logic, methods of reasoning, mind as physical system foundations of learning, language, rationality
- Mathematics — Formal representation and proof algorithms, computation, (un)decidability, (in)tractability, probability
- Economics — utility, decision theory
- Neuroscience — physical substrate for mental activity
- Psychology — phenomena of perception and motor control, experimental techniques
- Computer engineering — building fast computers
- Control theory — design systems that maximize an objective function over time
- Linguistics — knowledge representation, grammar

https://en.wikipedia.org/wiki/History_of_artificial_intelligence
https://en.wikipedia.org/wiki/Timeline_of_artificial_intelligence

AIMA

The physical symbol system hypothesis

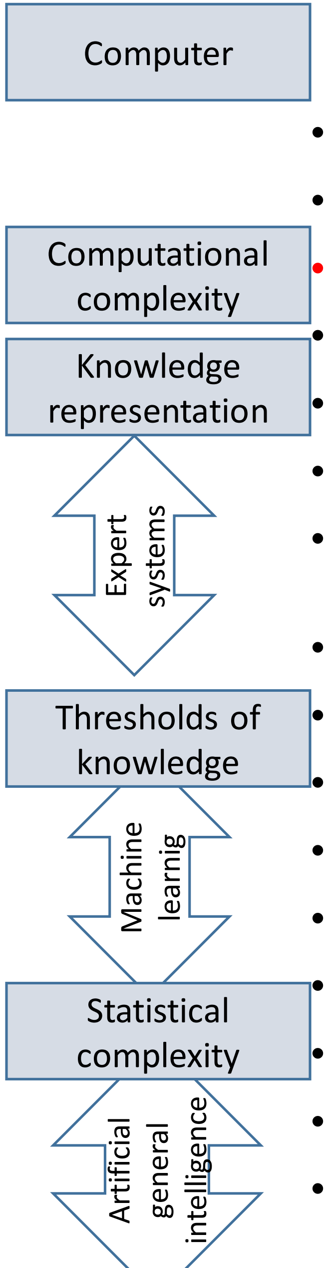
- A.Newel&H.A.Simon (1976): „A physical symbol system (PSS) has
  - the necessary and
  - sufficient
  - means for general intelligent action."

- „GOFAI”: good old-fashioned AI
  - PSS + search
  - **General Problem Solver (GPS)**

Newell, A., Shaw, J.C. and Simon, H.A., 1959, June. Report on a general problem solving program. In *IFIP congress* (Vol. 256, p. 64).
Simon, H.A. and Newell, A., 1962. Computer simulation of human thinking and problem solving. *Monographs of the Society for Research in Child Development*, pp.137-150.

AIMA

Milestones and phases in AI/ML



- ~1930: Zuse, Neumann, Turing..: „instruction is data”:
  - Laws of nature can be represented, „executed”/simulated with modifications, learnt
  - Knowledge analogously: representation, execution, adaptation and learning
- 1943 McCulloch & Pitts: Boolean circuit model of brain
- 1950 Turing's "Computing Machinery and Intelligence"
- 1956 Dartmouth meeting: the term "Artificial Intelligence”
- 1950s Early AI programs (e.g. Newell & Simon's Logic Theorist)
- **The physical symbol system hypothesis: search**
- 1965 Robinson's complete algorithm for logical reasoning
- 1966—73 AI discovers computational complexity
  Neural network research almost disappears
- 1969—79 Early development of knowledge-based systems
- **The knowledge system hypothesis: knowledge is power**
- 1986- Neural networks return to popularity
- 1988- Probabilistic expert systems
- 1990- Kernel methods
- **The „big data” hypothesis: let the data speak**
- 1995-2006/2009/2012 Emergence of machine learning
- **The artificial general intelligence hypothesis:**
- ~2010-... Deep learning, GPUs, reinforcement learning, machine teaching, explainable AI, provably beneficial AI,...

AIMA

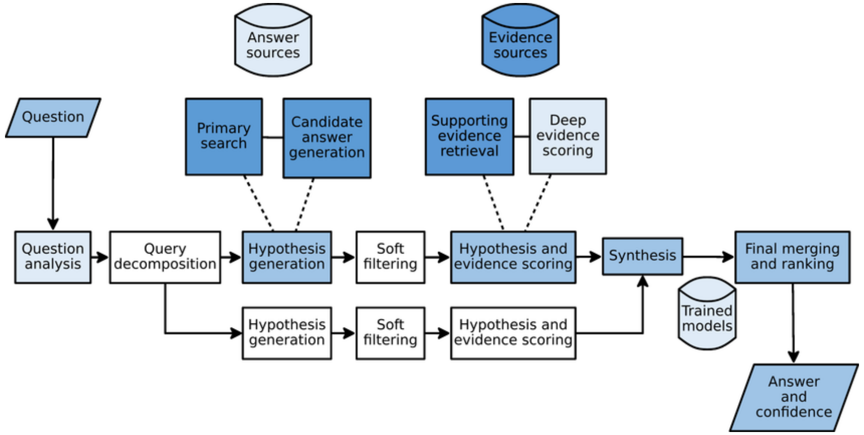
State of the art: ☺

- Deep Blue defeated the reigning world chess champion Garry Kasparov in 1997
- Proved a mathematical conjecture (Robbins conjecture) unsolved for decades
- No hands across America (driving autonomously 98% of the time from Pittsburgh to San Diego)
- During the 1991 Gulf War, US forces deployed an AI logistics planning and scheduling program that involved up to 50,000 vehicles, cargo, and people
- NASA's on-board autonomous planning program controlled the scheduling of operations for a spacecraft
- `Proverb` solves crossword puzzles better than most humans
- Google search/car/face recognition/…

AIMA

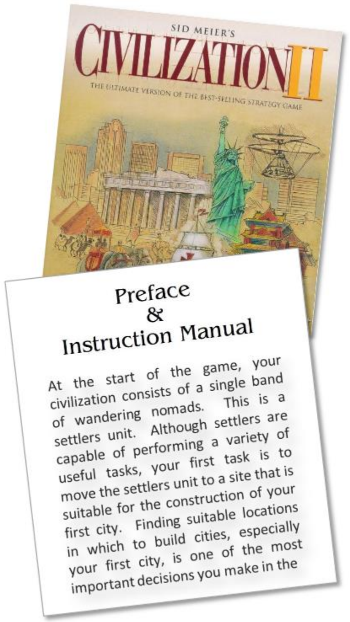
IBM Watson (2011): Jeopardy

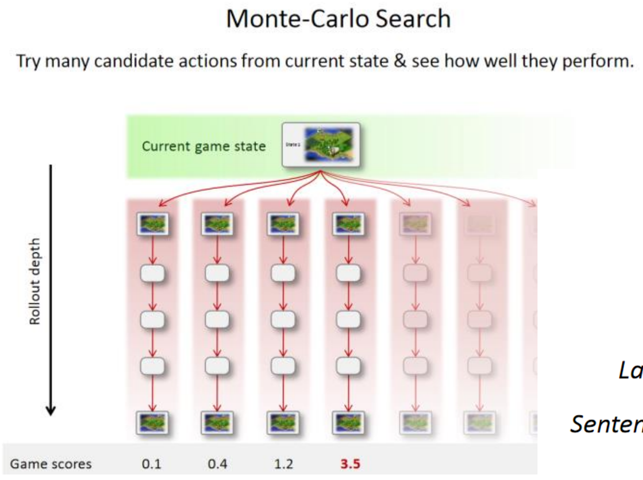
- IBM Grand Challenge
  - 1997: **Deep Blue** wins human champion G. Kasparov.
  - 1999-2006<: **Blue Gene**, protein prediction
  - 2011: **Watson**
    - Natural language processing
    - inference
    - Game theory



Machines playing Civilization

- Teaching + Learning: learning from manual and from practice



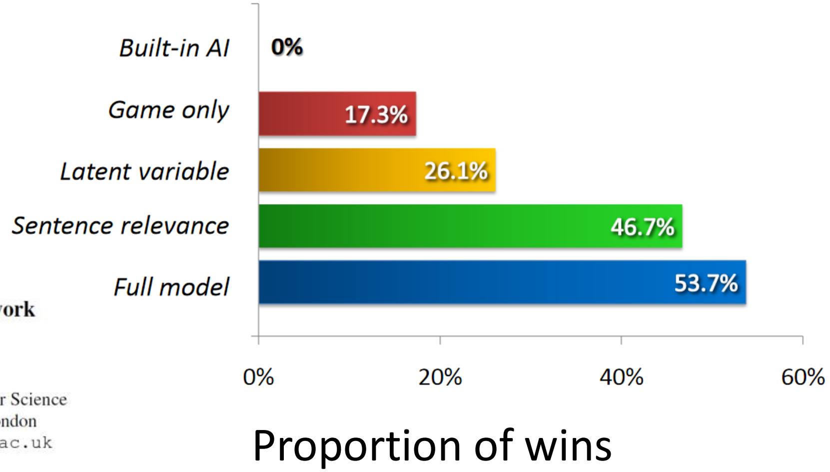


Learning to Win by Reading Manuals in a Monte-Carlo Framework

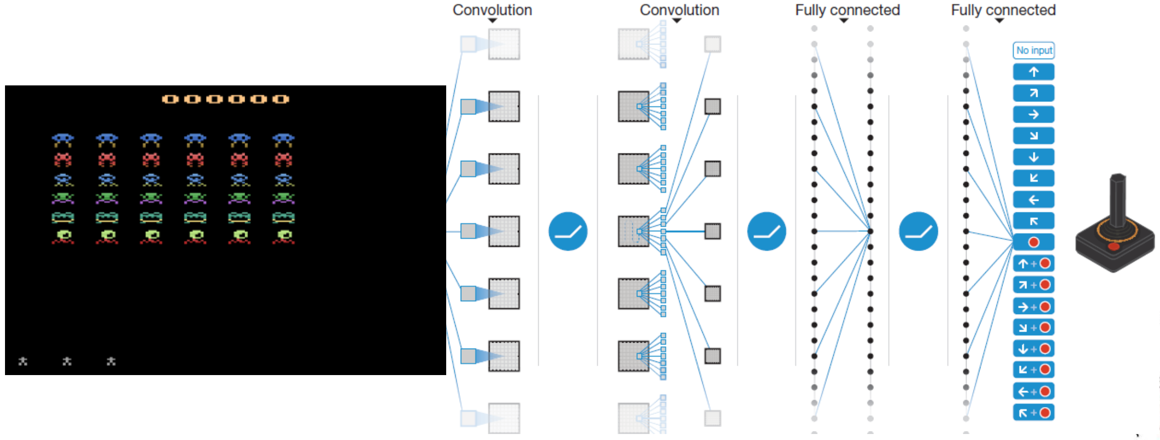
S.R.K. Branavan David Silver * Regina Barzilay

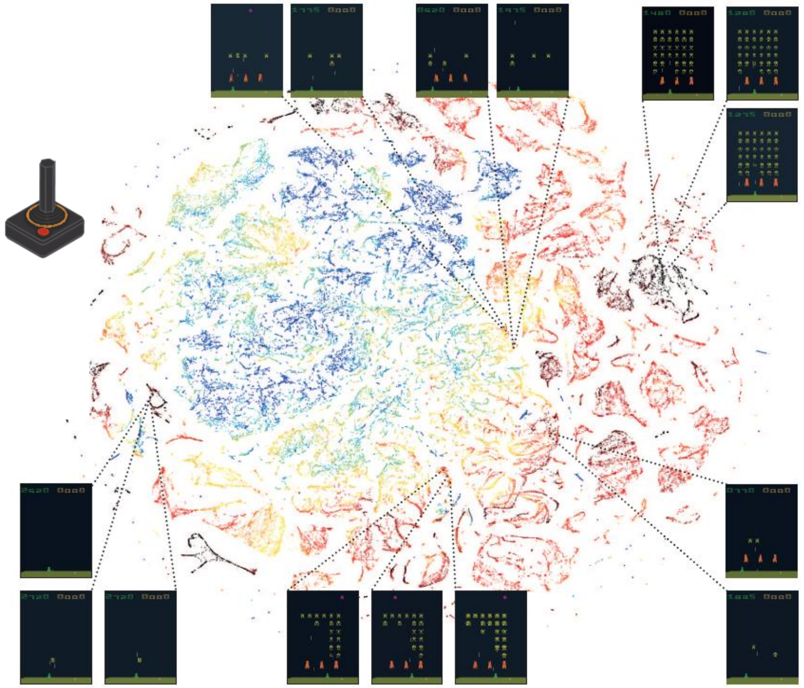
Computer Science and Artificial Intelligence Laboratory
Massachusetts Institute of Technology
{branavan, regina}@csail.mit.edu

* Department of Computer Science
University College London
d.silver@cs.ucl.ac.uk



Playing computer games





LETTER

doi:10.1038/nature14236

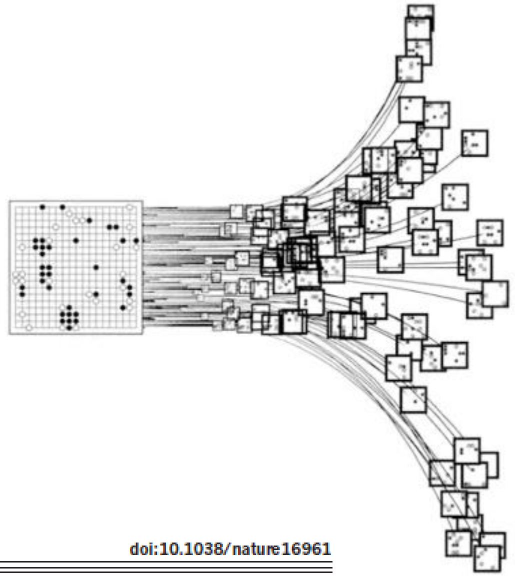
Human-level control through deep reinforcement learning

Volodymyr Mnih[1]*, Koray Kavukcuoglu[1]*, David Silver[1]*, Andrei A. Rusu[1], Joel Veness[1], Marc G. Bellemare[1], Alex Graves[1], Martin Riedmiller[1], Andreas K. Fidjeland[1], Georg Ostrovski[1], Stig Petersen[1], Charles Beattie[1], Amir Sadik[1], Ioannis Antonoglou[1], Helen King[1], Dharshan Kumaran[1], Daan Wierstra[1], Shane Legg[1] & Demis Hassabis[1]

Go:

AlphaGo

- Google DeepMind
- Monte Carlo tree search
- 2016: 9 dan
- 2017: wins against human champion



ARTICLE

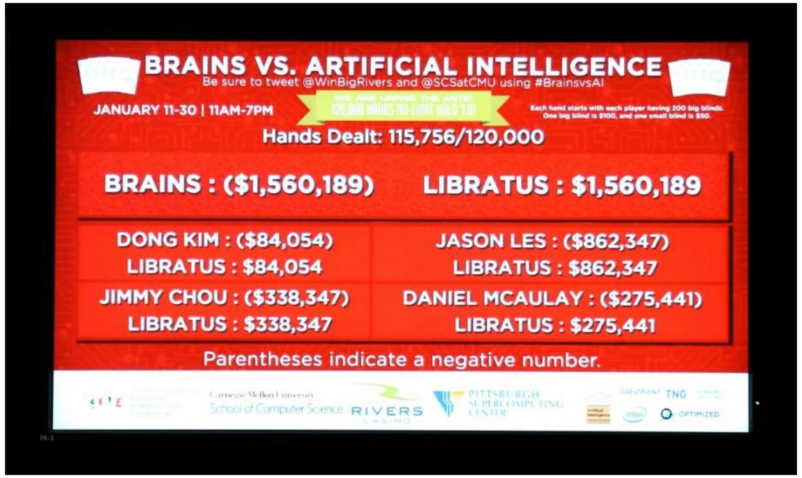
doi:10.1038/nature16961

Mastering the game of Go with deep neural networks and tree search

David Silver[1]*, Aja Huang[1]*, Chris J. Maddison[1], Arthur Guez[1], Laurent Sifre[1], George van den Driessche[1], Julian Schrittwieser[1], Ioannis Antonoglou[1], Veda Panneershelvam[1], Marc Lanctot[1], Sander Dieleman[1], Dominik Grewe[1], John Nham[2], Nal Kalchbrenner[1], Ilya Sutskever[2], Timothy Lillicrap[1], Madeleine Leach[1], Koray Kavukcuoglu[1], Thore Graepel[1] & Demis Hassabis[1]

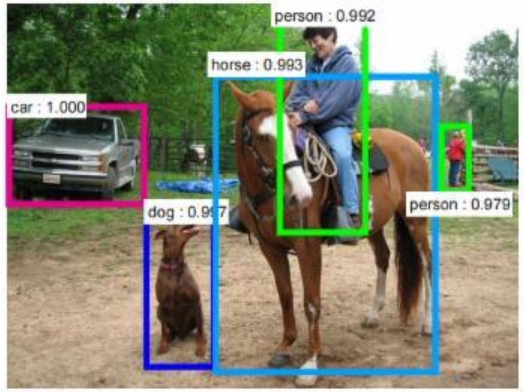
Poker: Libratus

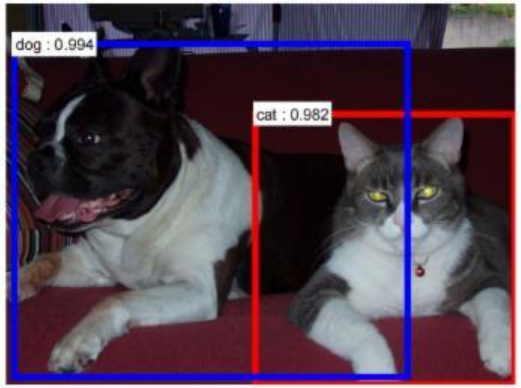
- 2017: Carnegie Mellon University MI: Libratius
- Pittsburgh Supercomputing Center:
  - 1.35 petaflops computation
  - 274 Terabytes memory



Vision: YOLO

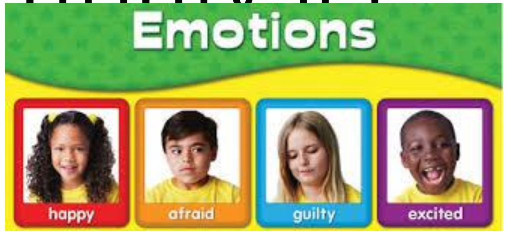
- YOLO (you only look once)

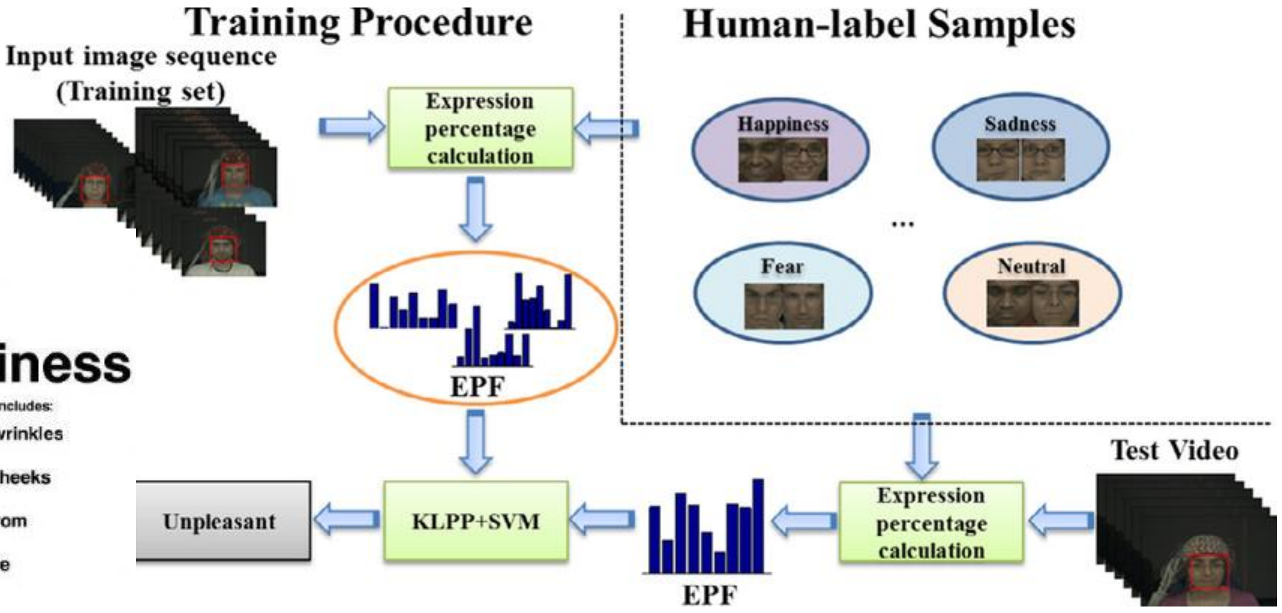


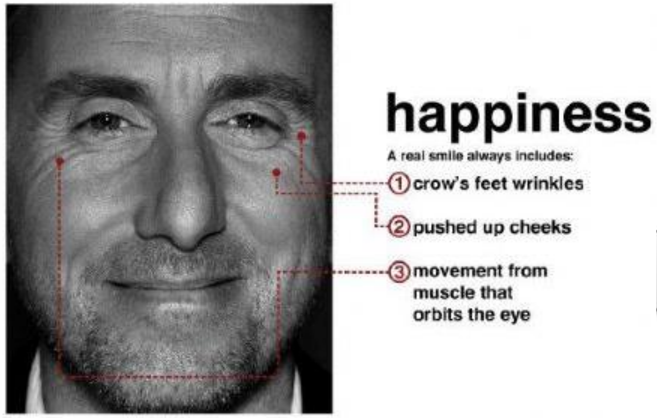


https://www.ted.com/talks/joseph_redmon_how_a_computer_learns_to_recognize_objects_instantly#t-409586

Emotion detection, sentiment analysis

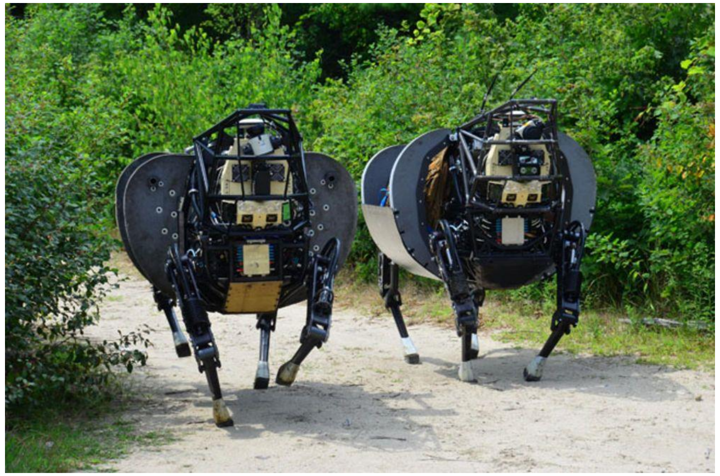






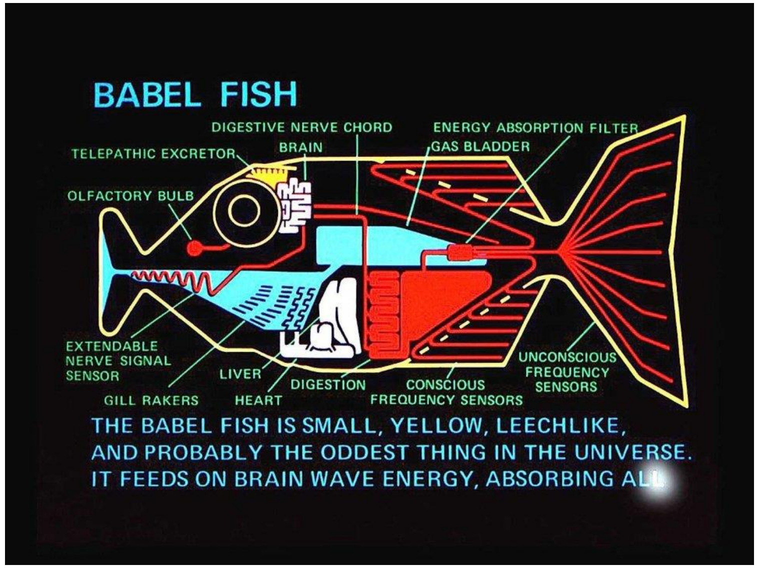
https://www.ted.com/talks/rana_el_kaliouby_this_app_knows_how_you_feel_from_the_look_on_your_face

Walking, movements



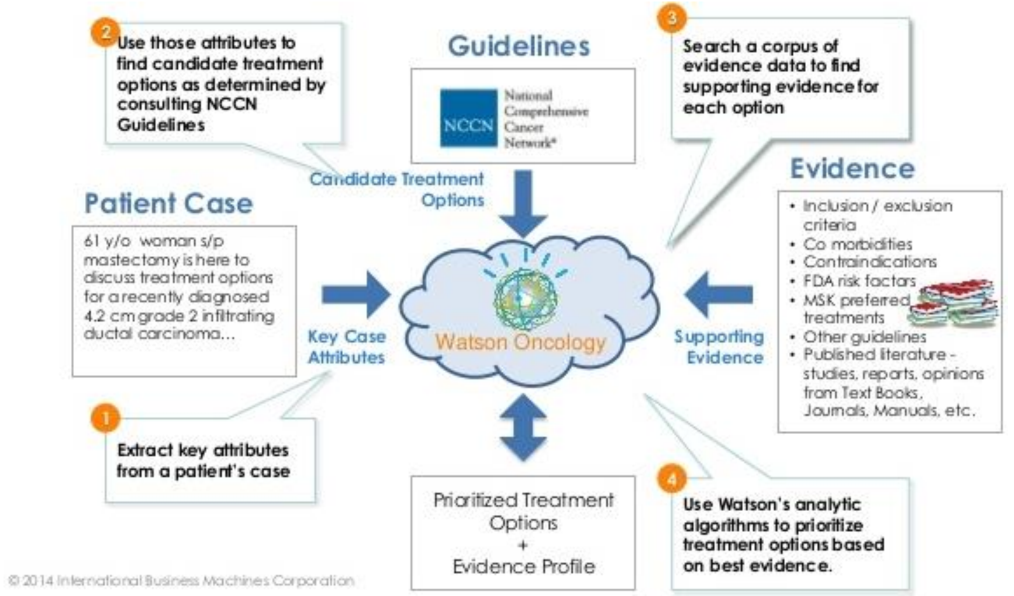
Real-time translation

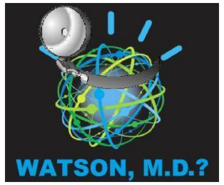
Pilot Translating Earpiece



D.Adams: [Galaxis útikalauz stopposoknak]
Hitchhiker's Guide to the Galaxy"

Clinical decision support systems

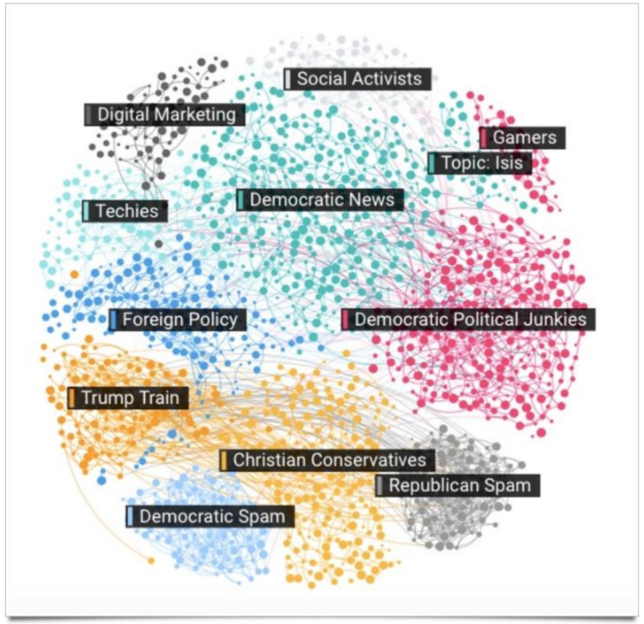




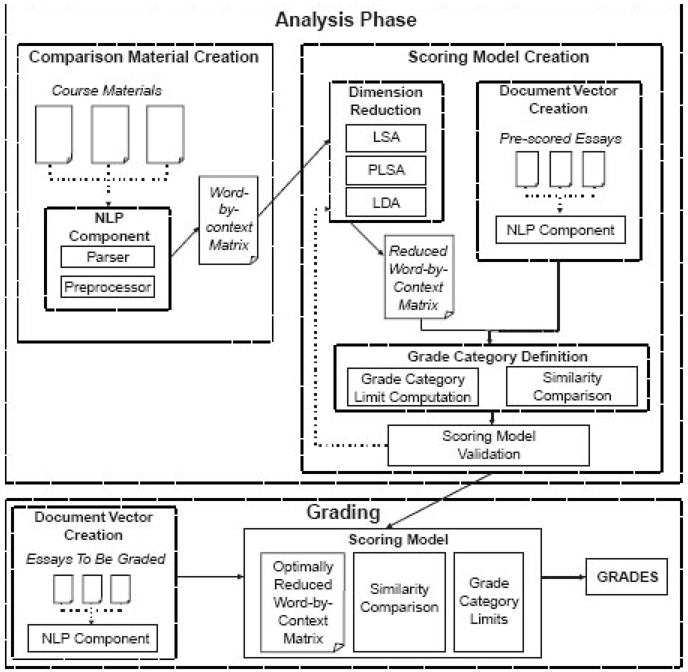
Watson for Oncology – assessment and advice cycle
www.avanteoconsulting.com/machine-learning-accelerates-cancer-research-discovery-innovation/

Political analytics: MogIA

- **~„big data failed, AI correctly predicted the upset victory” (correct prediction of election in the US 3 times in a row)**

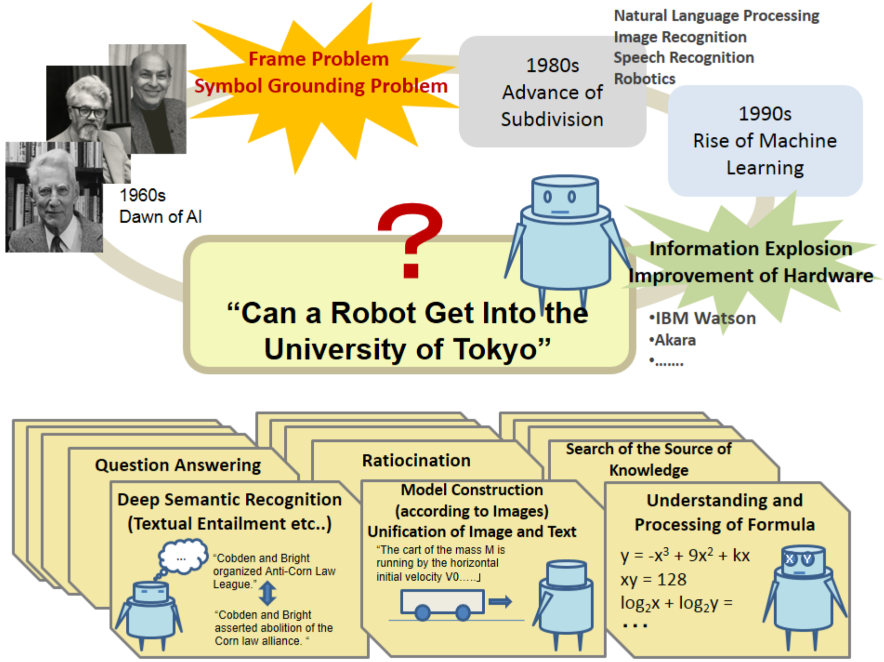


Automated essay scoring (AES)



University entry exam: Todai robot

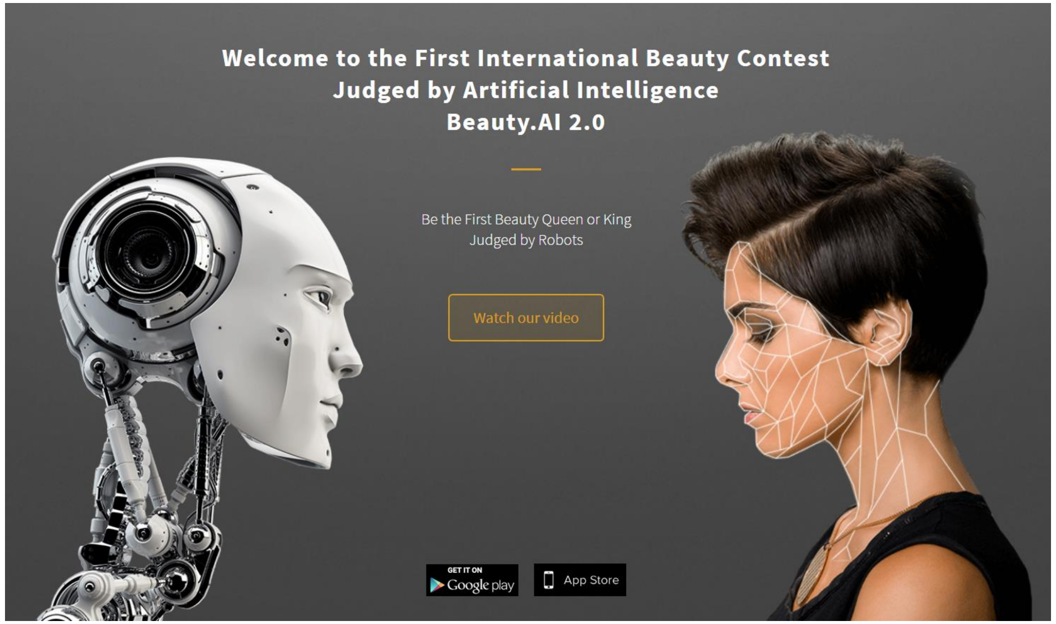
http://21robot.org/?lang=english



Legal applications of AI

- Juridical decisions:
  - Human experts: 66% identical decision.
  - Katz, D.M., Bommarito II, M.J. and Blackman, J., 2017. **A general approach for predicting the behavior of the Supreme Court of the United States**. *PloS one*, *12*(4), p.e0174698.
    - 1816-2015 esetek
    - 70%< accuracy
  - COMPAS CORE

Beauty.AI



http://beauty.ai/

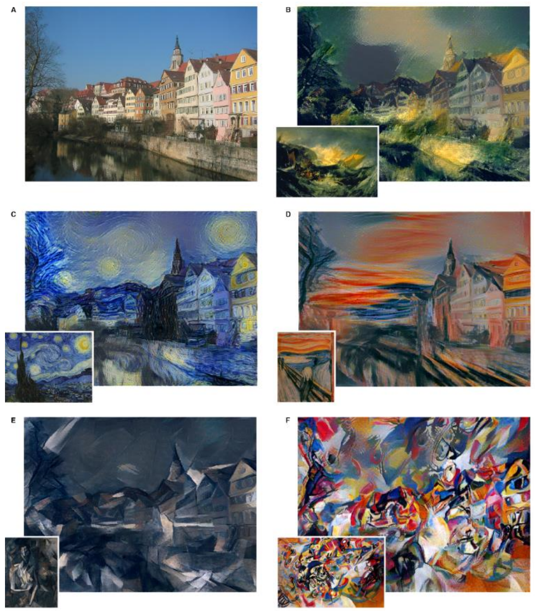
- **A beauty contest was judged by AI and the robots didn't like dark skin, Guardian**
- **Another AI Robot Turned Racist, This Time At Beauty Contest, Unilad**

Chatbot: Tay

- **Turing-test, Loebner-prize**
- **Tay** was an artificial intelligence chatterbot released by Microsoft Corporation on March 23, 2016. Tay caused controversy on Twitter by releasing inflammatory tweets and it was taken offline around 16 hours after its launch.[1] Tay was accidentally reactivated on March 30, 2016, and then quickly taken offline again.

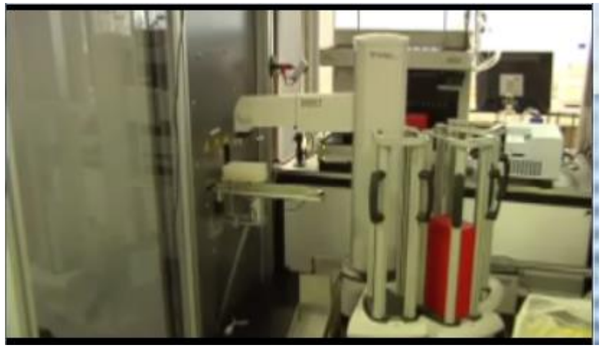
Reproduction of artistic style

- Gatys, L.A., Ecker, A.S. and Bethge, M., 2015. A neural algorithm of artistic style. *arXiv preprint arXiv:1508.06576*.



Automated scientific discovery

- Langley, P. (1978). Bacon: A general discovery system.
- …
- ...
- R.D.King et al.: **The Automation of Science**, Science, 2009
- Sparkes, Andrew, et al.: Towards **Robot Scientists** for autonomous scientific discovery, 2010

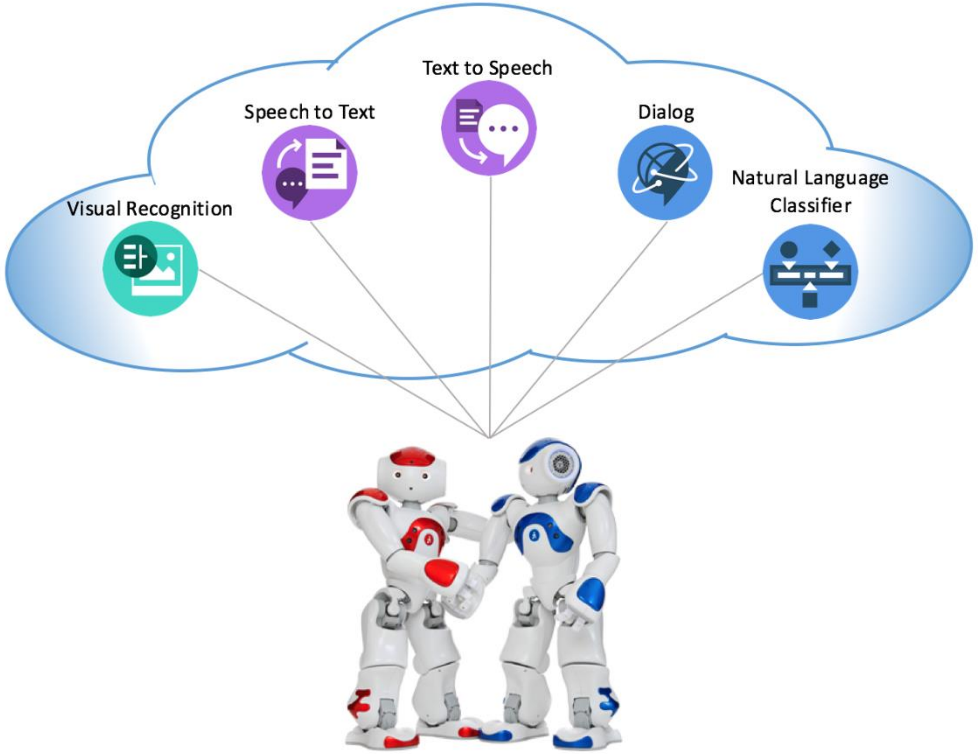


„Adam”

„Eve”

Humour?

"*Artificial intelligence* is when you get a college degree, but you're still stupid when you graduate."



Summary

- Driving forces of AI
    - Logic
    - Computational theories
        - Complexity theories
        - Computational linguistics
    - Computing power
    - Data flood
    - Scalable probabilistic reasoning
    - (Machine) learning theories
    - Causality research
- State of the art of (narrow) AI
- Next lecture: beyond narrow AI: AGI

- Suggested reading:

Russell, S., 2017. Artificial intelligence: The future is superintelligent. *Nature*, *548*(7669), pp.520-522.