

Artificial general intelligence

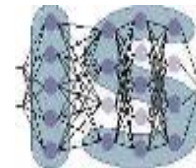
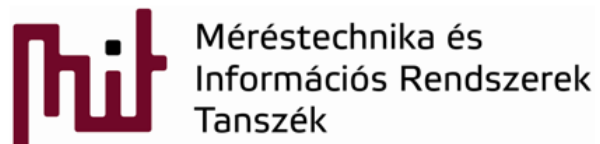
singularity vs. open AGI

Antal Péter, Bolgár Bence

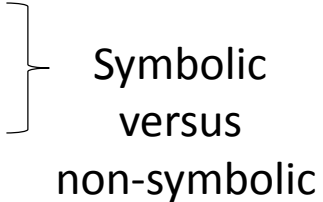
Számítógépes orvosbiológiai munkacsoport

Mesterséges Intelligencia kutatócsoport

BME, VIK, Méréstechnika és információs rendszerek tanszék



Agenda

- Resources
 - Examples for the phases/paradigms of AI
 - logic&search,
 - (understandable/white-box) expert systems,
 - black-box learning
 - Strong AI
 - Superintelligence/singularity: intelligence explosion
 - What AI already gave us:
 - rational models of (narrow) intelligence
 - broad range of theories and technologies
 - collaboration: open AGI
- 
- Symbolic
versus
non-symbolic

Resources

AGI resources: books

- Russell, Stuart J., and Peter Norvig. *Artificial intelligence: a modern approach*. 1st< edition
 - <http://aima.cs.berkeley.edu/>
 - http://project.mit.bme.hu/mi_almanach/
- ***Artificial general intelligence*. Ed. Goertzel, Ben., Cassio Pennachin. Vol. 2. New York: Springer, 2007.**
- Goertzel, Ben. *The AGI Revolution: An Inside View of the Rise of Artificial General Intelligence*. Humanity+ Press, 2016.

AGI resources: societies, conferences

- AGI Society (Artificial General Intelligence, AGI)
 - <http://www.agi-society.org/>
- AGI conferences 2008..2019
 - https://en.wikipedia.org/wiki/Conference_on_Artificial_General_Intelligence
 - 2008: AGI-08 Workshop on the Sociocultural, Ethical and Futurological Implications of Artificial General Intelligence
 - <http://agi-conf.org/2008/workshop/>
 - <http://agi-conf.org/2019/>

AGI resources: courses

- MIT 6.S099: Artificial General Intelligence
 - <https://agi.mit.edu/>
- CS 294-149: Safety and Control for Artificial General Intelligence (Fall 2018)
 - <https://inst.eecs.berkeley.edu/~cs294-149/fa18/>

AGI resources: podcasts

- Lex Friedman: MIT 6.S099: Artificial General Intelligence
 - <https://agi.mit.edu/>
- Sean Carroll's Mindscape Podcast
 - <http://www.preposterousuniverse.com/podcast/>
- Sam Harris: Making sense
 - <https://samharris.org/podcast/>
- PLAN: voting for podcasts!

Resources: your background

- Mentimeter
 - Earlier AI courses
 - None, without engineering BSc/MSc
 - Nothing specific, but with engineering background
 - Partial BSc-level AI
 - BSc-level AI
 - MSc-level AI: Decision support, Statistics, Machine learning
 - PhD-level AI
 - Expectations

Phases/paradigms of AI: symbolic AI

AI as “symbol manipulation”

- The Logic Theorist, 1955
 - ➔ see lectures on logic
- The Dartmouth conference ("birth of AI", 1956)
- List processing (Information Processing Language, IPL)
- Means-ends analysis ("reasoning as search")
 - ➔ see lectures on planning
- The General Problem Solver
- Heuristics to limit the search space
 - ➔ see lecture on informed search
- The physical symbol systems hypothesis
 - intelligent behavior can be reduced to/emulated by symbol manipulation
- The unified theory of cognition (1990, cognitive architectures: Soar, ACT-R)
- Newel&Simon: Computer science as empirical inquiry: symbols and search, 1975

Problem formulation

- A problem is defined by:
 - An **initial state**, e.g. *Arad*
 - **Successor function** $S(X)$ = set of action-state pairs
 - e.g. $S(\text{Arad}) = \{ \langle \text{Arad} \rightarrow \text{Zerind}, \text{Zerind} \rangle, \dots \}$initial state + successor function = state space
- **Goal test**, can be
 - Explicit, e.g. $x = \text{'at bucharest'}$
 - Implicit, e.g. $\text{checkmate}(x)$
- **Path cost** (additive)
 - e.g. sum of distances, number of actions executed, ...
 - $c(x, a, y)$ is the step cost, assumed to be ≥ 0

A **solution** is a sequence of actions from initial to goal state.

Optimal solution has the lowest path cost.

Iterative deepening search

- What?
 - A general strategy to find best depth limit l .
 - Goals is found at depth d , the depth of the shallowest goal-node.
 - Often used in combination with DF-search
- Combines benefits of DF- en BF-search

Iterative deepening search

function ITERATIVE_DEEPENING_SEARCH(*problem*) **return** a solution or failure

inputs: *problem*

for *depth* $\leftarrow 0$ to ∞ **do**

result \leftarrow DEPTH-LIMITED_SEARCH(*problem*, *depth*)

if *result* \neq *cutoff* **then return** *result*

ID-search, example

- Limit=0



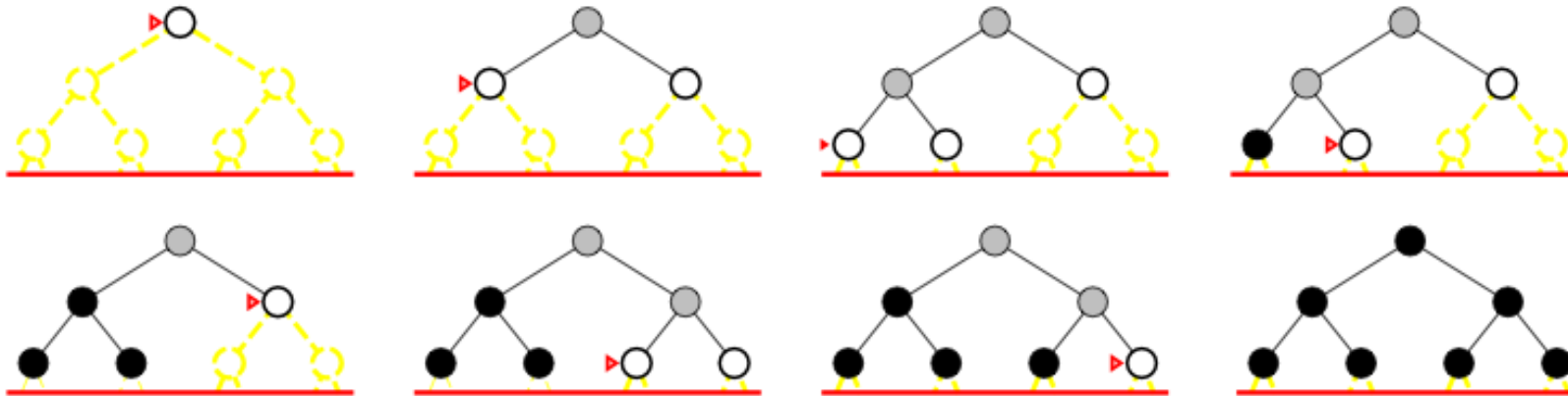
ID-search, example

- Limit=1



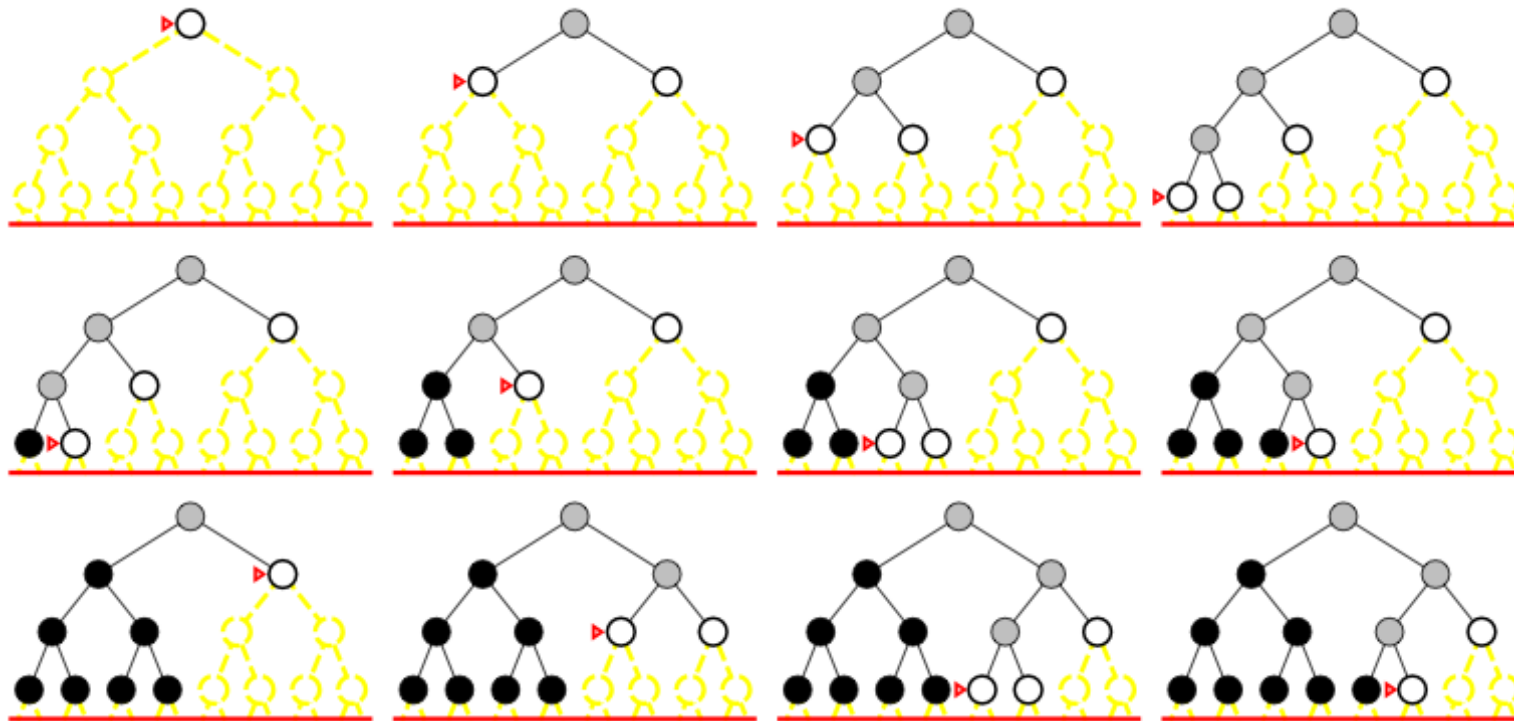
ID-search, example

- Limit=2



ID-search, example

- Limit=3



ID search, evaluation

- Completeness:
 - YES (no infinite paths)

ID search, evaluation

- Completeness:
 - YES (no infinite paths)
- Time complexity:
 - Algorithm seems costly due to repeated generation of certain states.
 - Node generation:
 - level d: once
 - level d-1: 2
 - level d-2: 3
 - ...
 - level 2: d-1
 - level 1: d

$$O(b^d)$$

$$N(IDS) = (d)b + (d-1)b^2 + \dots + (1)b^d$$

$$N(BFS) = b + b^2 + \dots + b^d + (b^{d+1} - b)$$

Num. Comparison for b=10 and d=5 solution at far right

$$N(IDS) = 50 + 400 + 3000 + 20000 + 100000 = 123450$$

$$N(BFS) = 10 + 100 + 1000 + 10000 + 100000 + 999990 = 1111100$$

ID search, evaluation

- Completeness:
 - YES (no infinite paths)
- Time complexity:
- Space complexity:
 - Cfr. depth-first search

$$O(b^d)$$

$$O(bd)$$

ID search, evaluation

- Completeness:

- YES (no infinite paths)

$$O(b^d)$$

- Time complexity:

$$O(bd)$$

- Space complexity:

- Optimality:

- YES if step cost is 1.
 - Can be extended to iterative lengthening search
 - Same idea as uniform-cost search
 - Increases overhead.

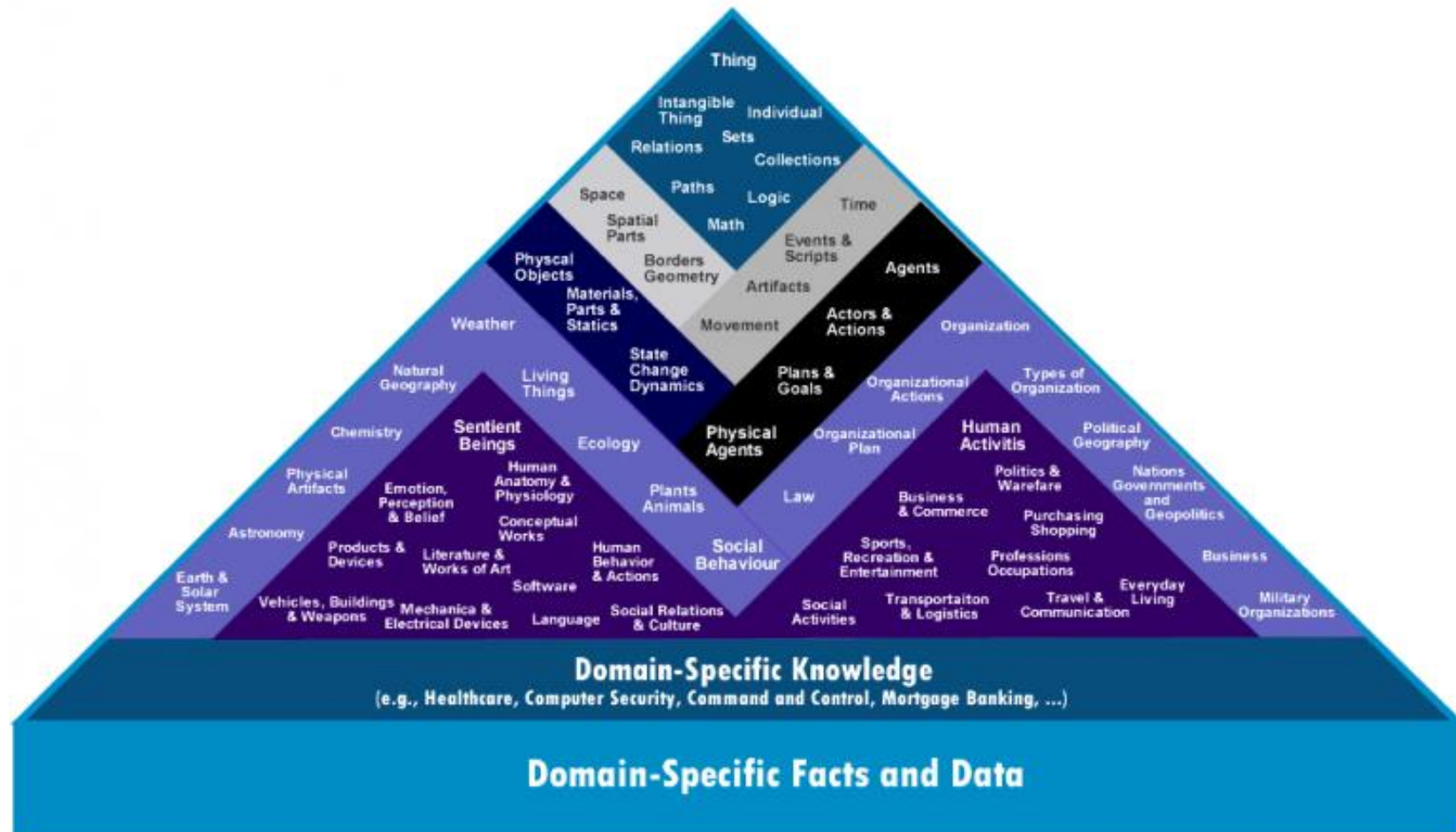
Summary of algorithms

Criterion	Breadth-First	Uniform-cost	Depth-First	Depth-limited	Iterative deepening	Bidirectional search
Complete?	YES*	YES*	NO	YES, if $l \geq d$	YES	YES*
Time	b^{d+1}	$b^{C*/e}$	b^m	b^l	b^d	$b^{d/2}$
Space	b^{d+1}	$b^{C*/e}$	bm	bl	bd	$b^{d/2}$
Optimal?	YES*	YES*	NO	NO	YES	YES

The Cyc project (1984-2016)



- Goal: common sense
- Estimations in 1984:
 - 250 000 rules
 - 350 man-year
- Language: CycL
- Access: OpenCyc
- Current state
 - 239,000 concept
 - 2,093,000 facts



Phases/paradigms of AI: expert AI

Optimal decision: decision theory probability theory+utility theory

- Decision situation:

- Actions
- Outcomes
- Probabilities of outcomes
- Utilities/losses of outcomes
- Maximum Expected Utility Principle (MEU)
- Best action is the one with maximum expected utility

$$a_i$$

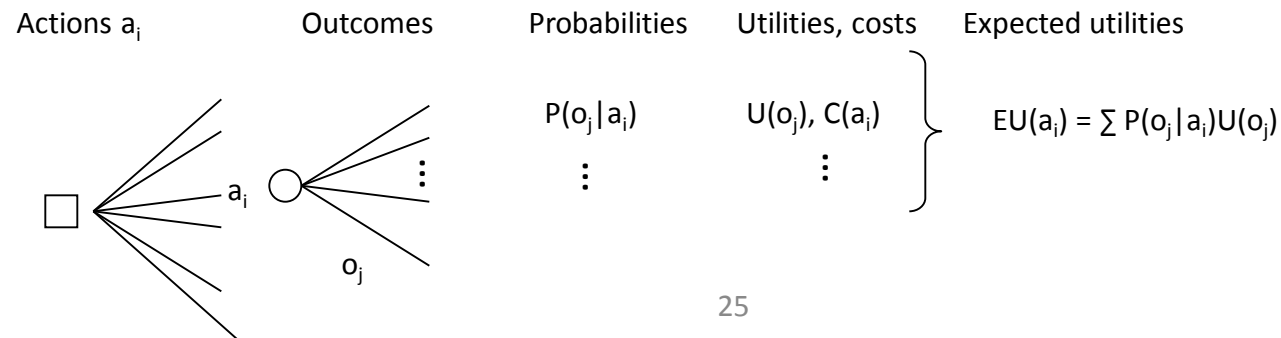
$$o_j$$

$$p(o_j | a_i)$$

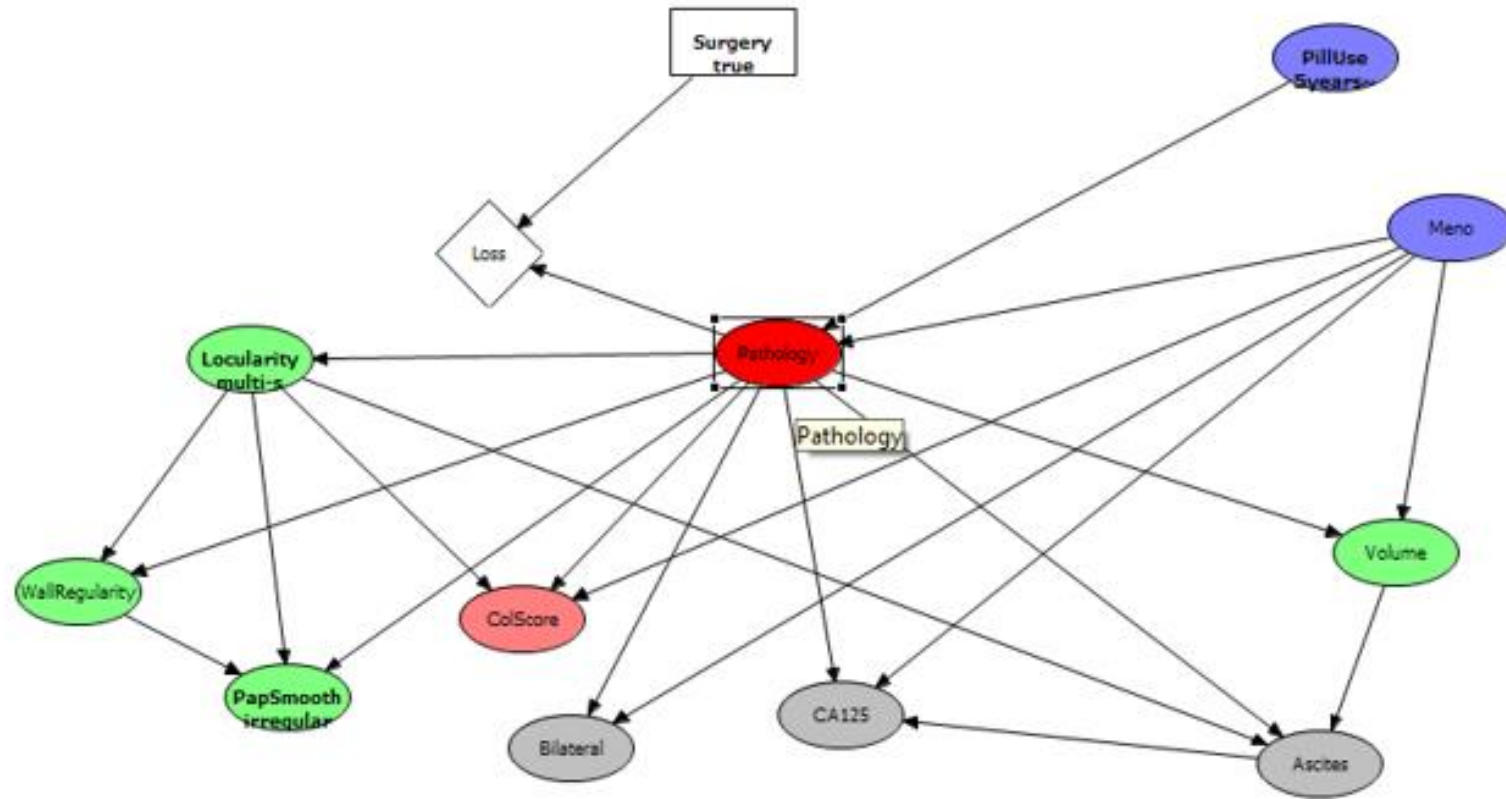
$$U(o_j | a_i)$$

$$EU(a_i) = \sum_j U(o_j | a_i) p(o_j | a_i)$$

$$a^* = \arg \max_i EU(a_i)$$



Decision networks



Antal, P., Fannes, G., Timmerman, D., Moreau, Y. and De Moor, B., 2004. Using literature and data to learn Bayesian networks as clinical models of ovarian tumors. *Artificial Intelligence in medicine*, 30(3), pp.257-281.

Phases/paradigms of AI: machine learning

Inductive learning

- Simplest form: learn a function from examples
-

f is the **target function**

An **example** is a pair $(x, f(x))$

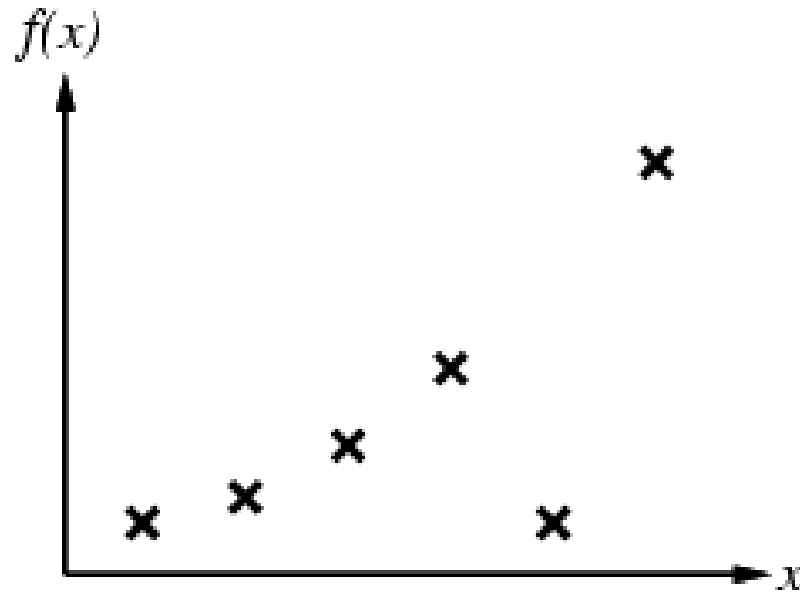
Problem: find a **hypothesis** h
such that $h \approx f$
given a **training set** of examples

(This is a highly simplified model of real learning:

- Ignores prior knowledge
- Assumes examples are given)
-

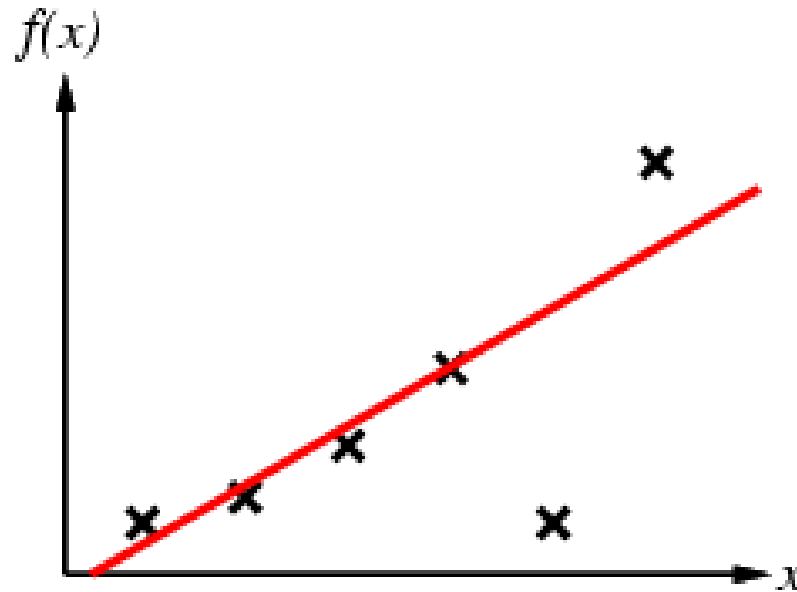
Inductive learning method

- Construct/adjust h to agree with f on training set
- (h is **consistent** if it agrees with f on all examples)
-
- E.g., curve fitting:
-



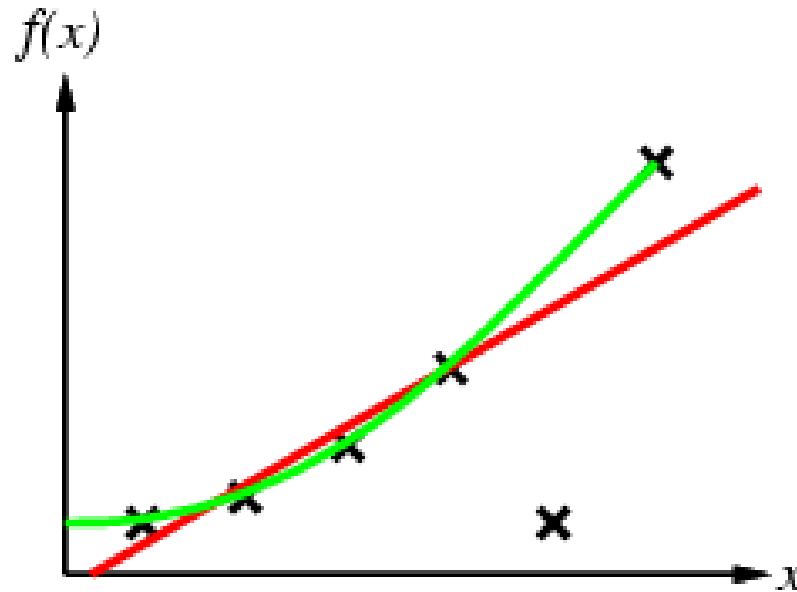
Inductive learning method

- Construct/adjust h to agree with f on training set
- (h is **consistent** if it agrees with f on all examples)
-
- E.g., curve fitting:
-



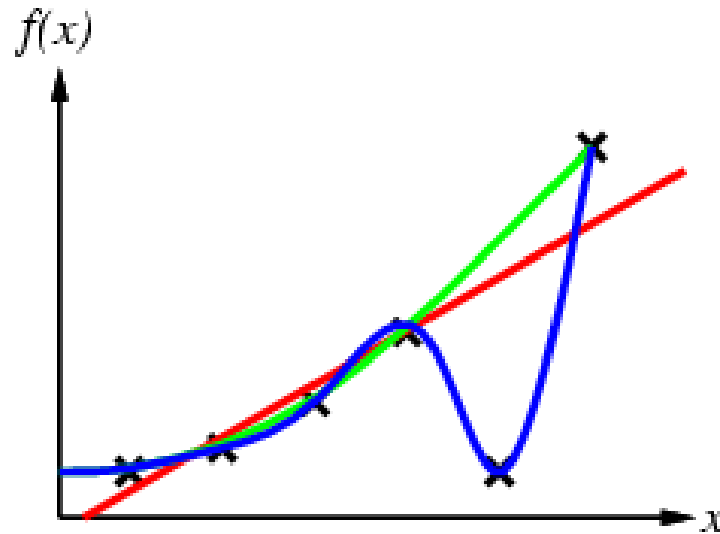
Inductive learning method

- Construct/adjust h to agree with f on training set
- (h is **consistent** if it agrees with f on all examples)
-
- E.g., curve fitting:
-



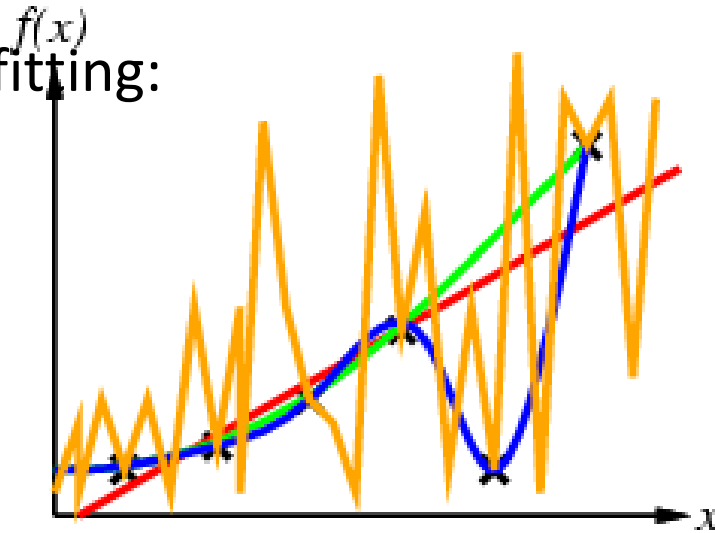
Inductive learning method

- Construct/adjust h to agree with f on training set
- (h is **consistent** if it agrees with f on all examples)
-
- E.g., curve fitting:
-



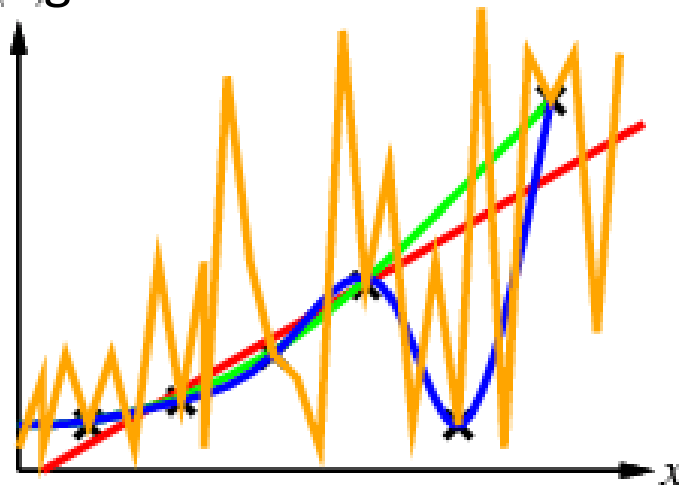
Inductive learning method

- Construct/adjust h to agree with f on training set
- (h is **consistent** if it agrees with f on all examples)
-
- E.g., curve fitting:



Inductive learning method

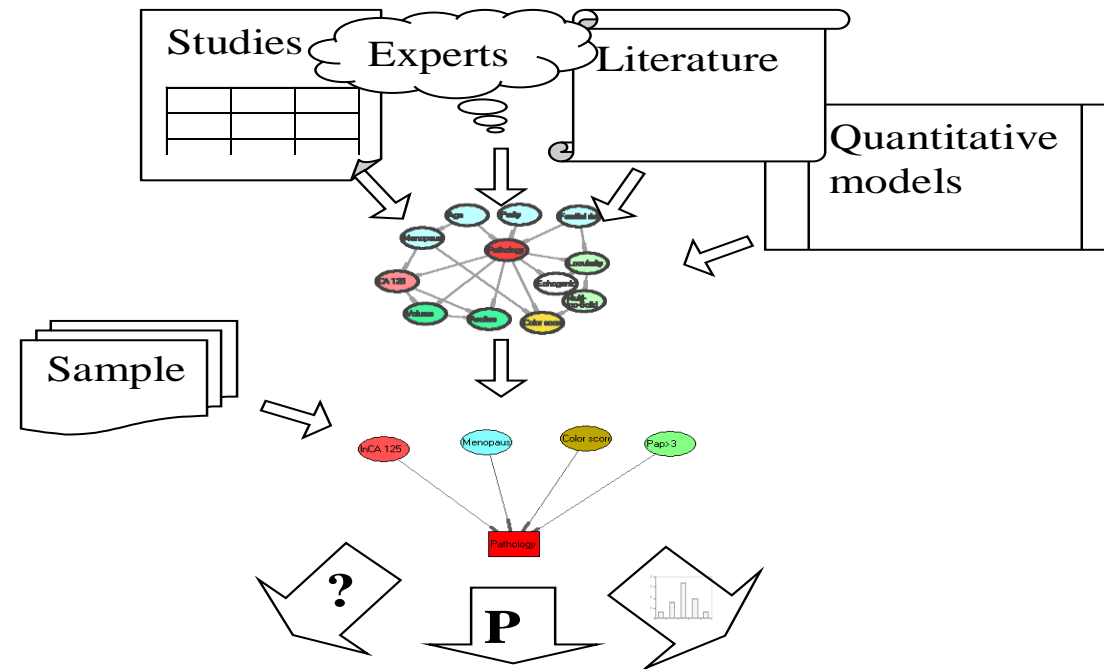
- Construct/adjust h to agree with f on training set
- (h is **consistent** if it agrees with f on all examples)
-
- E.g., curve fitting:



- Ockham's razor: prefer the simplest hypothesis consistent with data

Phases/paradigms of AI:
learning with background knowledge

Informed neural networks



Classification

Probability prediction

Credible region

P. Antal, G. Fannes, D. Timmerman, Y. Moreau, B. De Moor: Bayesian Applications of Belief Networks and Multilayer Perceptrons for Ovarian Tumor Classification with Rejection, *Artificial Intelligence in Medicine*, vol. 29, pp 39-60, 2003

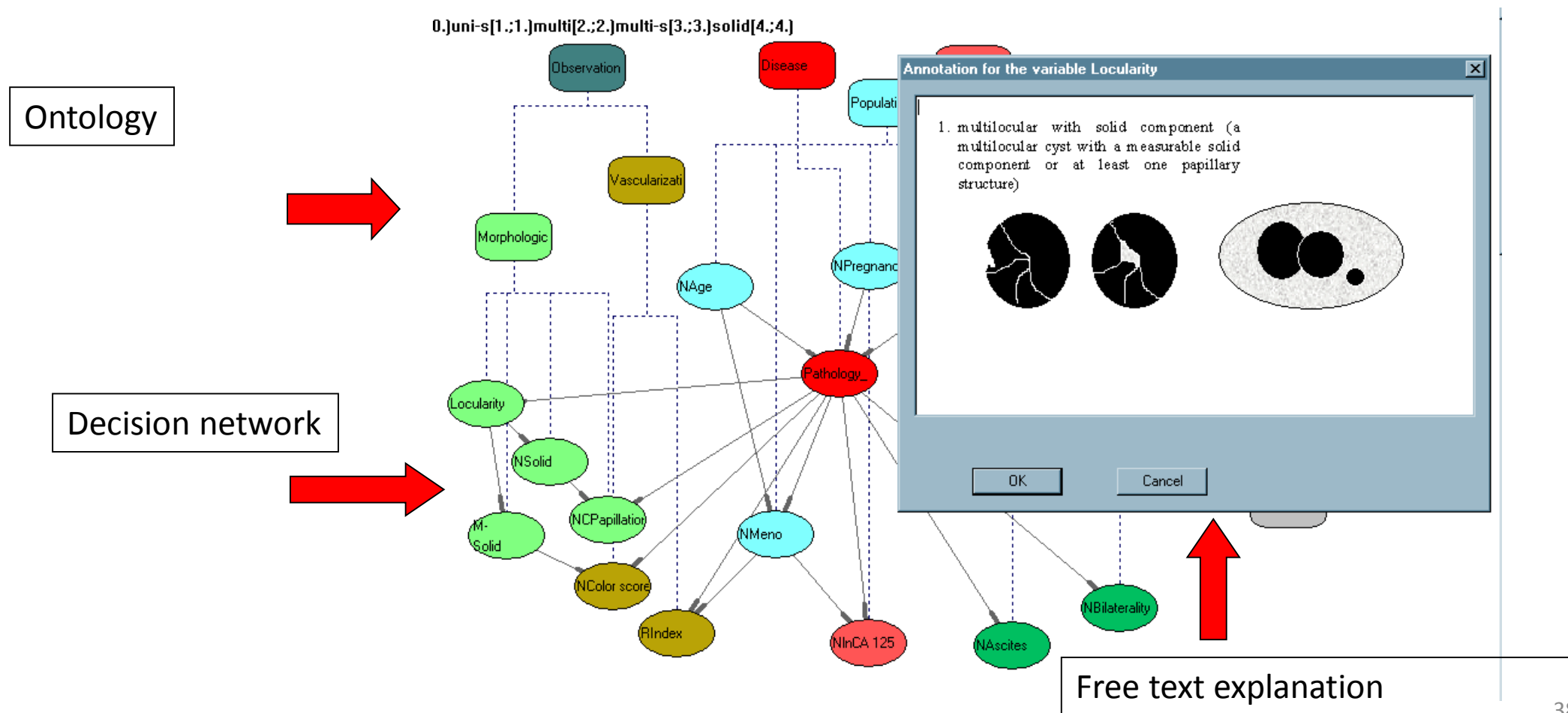
P. Antal, G. Fannes, H. Verrelst, B. De Moor, J Vandewalle: **Incorporation of prior knowledge in black-box models: Comparison of Transformation Methods from Bayesian Network to Multilayer Perceptrons**, in Working notes of the Fusion of Domain Knowledge with Data for Decision Support workshop, The Sixteenth Conference on Uncertainty in Artificial Intelligence (UAI-2000), June 30, 2000, Stanford University, pp. 11-16

Antal, P. et al. (2000). **How might we combine the information we know** about a mass better? In *Proc. of the 1st Montecarlo Conference on updates in Gynaecology (MCG)* (pp. 1-3).

Timmerman, D. (2004). The use of mathematical models to evaluate pelvic masses; **can they beat an expert operator?**. *Best Practice & Research Clinical Obstetrics & Gynaecology*, 18(1), 91-104.

Phases/paradigms of AI:
explanation generation

Evidence-based, explainable AI



Antal, P., Mészáros, T., De Moor, B., & Dobrowiecki, T. (2001). Annotated Bayesian Networks: a tool to integrate textual and probabilistic medical knowledge. In *Proceedings 14th IEEE Symposium on Computer-Based Medical Systems. CBMS 2001* (pp. 177-182).

Explainable AI

Query

Evidence-based decision network

The screenshot displays the BNet software interface, which is used for evidence-based decision networks. The interface is divided into several panels:

- Query Panel:** Located on the left, it contains a text area for entering a query. A red arrow points to this panel with the label "Query".
- Evidence-based decision network Panel:** Located on the right, it shows a graphical representation of a decision network. A red arrow points to this panel with the label "Evidence-based decision network".
- Result list Panel:** Located in the center, it displays a list of search results. A red arrow points to this panel with the label "Explanation".
- Relevant publications Panel:** Located at the bottom right, it shows a detailed view of a specific publication. A red arrow points to this panel with the label "Relevant publications".

The "Result list" panel shows a list of search results, including titles, authors, and scores. The "Relevant publications" panel shows a detailed view of a specific publication, including the title, authors, keywords, and abstract.

Antal, P., De Moor, B., Timmerman, D., Mészáros, T., & Dobrowiecki, T. (2002). Domain knowledge based information retrieval language: an application of annotated Bayesian networks in ovarian cancer domain. In *Proceedings of 15th IEEE Symposium on Computer-Based Medical Systems (CBMS 2002)* (pp. 213-218).

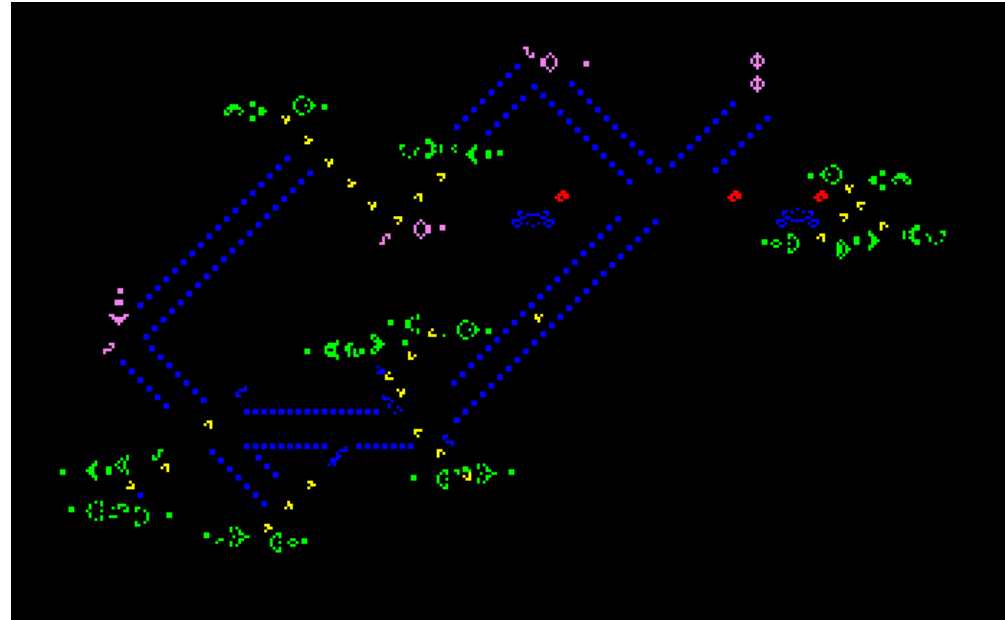
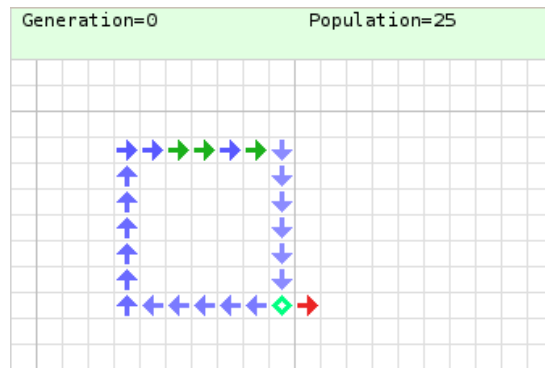
Phases/paradigms of AI:
strong AI

Strong AI, zombie arguments

- Simulation and tests for reality (~testing outside from inside)
 - ~BC300: Zhuangzi's (Chuang-Tzu's), Butterfly Dream
 - ~1700: G.Berkeley, subjective idealism
 - Movies: Matrix, Inception,...
 - .. N. Bostrom: https://en.wikipedia.org/wiki/Simulation_hypothesis
 - A. Becker: What Is Real?: The Unfinished Quest for the Meaning of Quantum Physics
- Simulation and tests for human mind (~testing inside)
 - Experience: any formally defined (discrete) computation is a program on a universal Turing machine.
 - Experience: any (narrow) intelligence can have a functionally equivalent computational model.
 - Zombie arguments:
 - Assumption: there are (discrete) computational models for conscious minds.
 - Paradox: any execution using arbitrary substrate and realization will give rise to qualia/consciousness.
 - Example: Chinese room (using epiphenomenal patterns in a cellular automaton, see next slide)
 - https://en.wikipedia.org/wiki/Philosophical_zombie
- Reductionism, emergence, downward causation,...

Universal computation in cellular automata using epiphenomenal patterns

- Von Neumann, J. and A. W. Burks (1966): Theory of self-reproducing automata

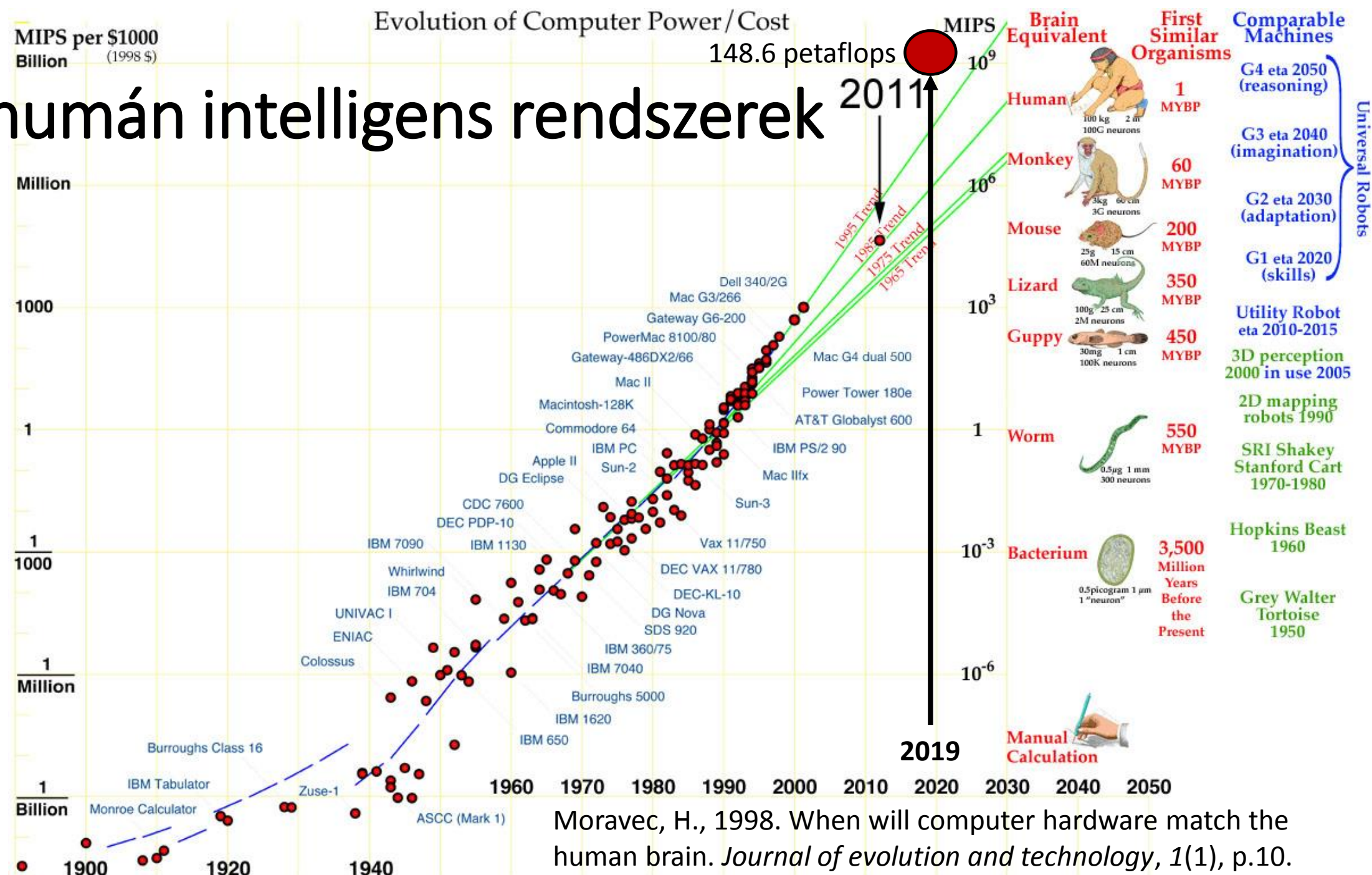


Phases/paradigms of AI: singularity

Intelligence explosion, „singularity”

„Let an ultraintelligent machine be defined as a machine that can far surpass all the intellectual activities of any man however clever. Since the design of machines is one of these intellectual activities, an ultraintelligent machine could design even better machines; there would then unquestionably be an ‘**intelligence explosion**,’ and the intelligence of man would be left far behind. Thus the first ultraintelligent machine is the last invention that man need ever make, provided that the machine is docile enough to tell us how to keep it under control.” [Good \(1965\)](#),

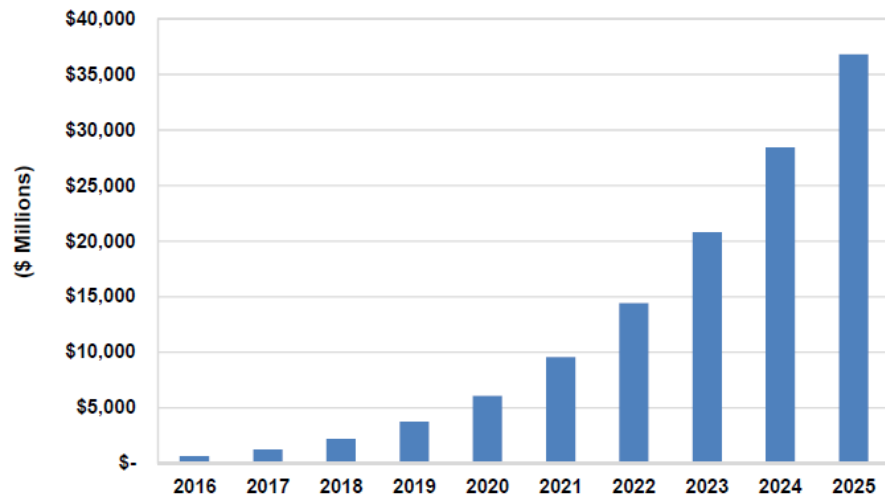
Post-humán intelligens rendszerek



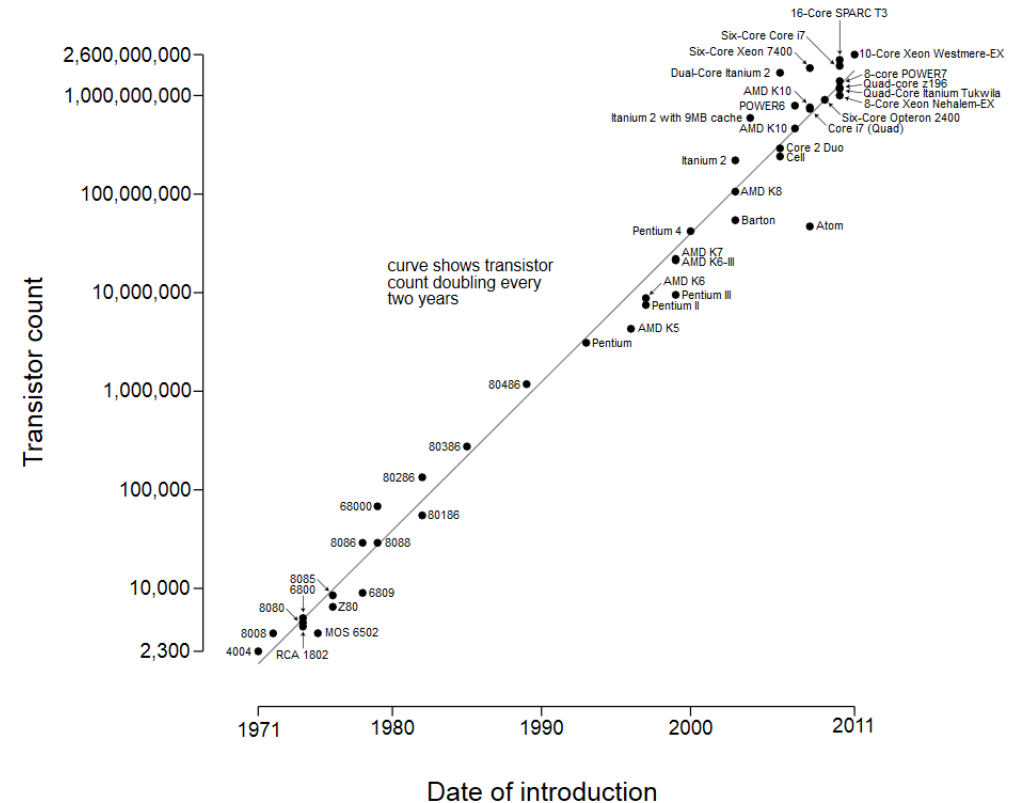
Moravec, H., 1998. When will computer hardware match the human brain. *Journal of evolution and technology*, 1(1), p.10.

AI: computational power, data, methods, money

Chart 1.1 Artificial Intelligence Revenue, World Markets: 2016-2025



(Source: Tractica)



- [10 \$\mu\text{m}\$](#) – 1971
- [6 \$\mu\text{m}\$](#) – 1974
- [3 \$\mu\text{m}\$](#) – 1977
- [1.5 \$\mu\text{m}\$](#) – 1982
- [1 \$\mu\text{m}\$](#) – 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

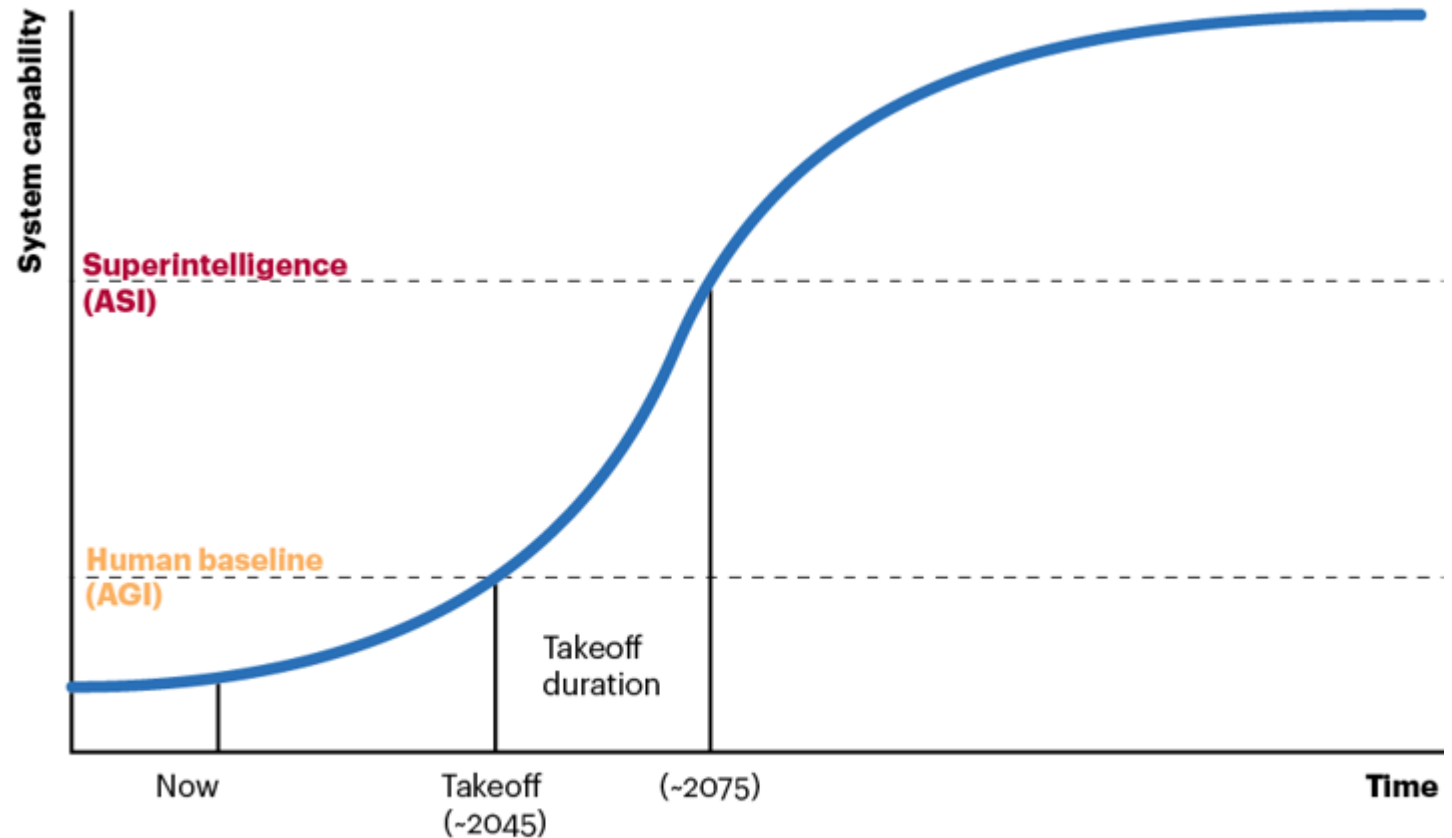
2012: single atom transistor (~0.1n, 1A)

Feedback, self-improving AI

- Chip design and production (optimal wiring, ...fault diagnosis)
- AI environments (TensorFlow, fastAI,...)
- Compiler technologies (GPUs...)
- Active learning

Artificial general/super intelligence,

- Narrow AI:
 - In any well-defined task AI will beat human performance
- Human-level AI
- Beyond human-level AI



Note: AI is artificial intelligence, ASI is artificial superintelligence, and AGI is artificial general intelligence.

Sources: WaitButWhy.com, Nick Bostrom, *Superintelligence: Paths, Dangers, Strategies*; A.T. Kearney analysis

Mesterséges Általános Intelligencia (MÁI)

- MI dimenziók

- Teljesítményszintek: jelölt, haladó, mester, nagymester
- **Szűk (narrow) versus általános (general) intelligencia**
 - Józan ész (common sense), naív fizika,...

- **MÁI – Emberközpontú megközelítés**

- The Turing Test (Turing)
- The Coffee Test (Wozniak)
- The Robot College Student Test (Goertzel)
- The Employment Test (Nilsson)
- The flat pack furniture test (Severyns)
- ...

- **MÁI – Racionális megközelítés**

- Egységes elmélet
- Közös alap MSZI-k számára
- Hordozható intelligencia, analógikus gondolkodás, kreativitás,...

MSZI-1

MSZI-2

....

MSZI-n

MÁI: Józan ész, naív fizika, vizualitás, okozatiság, analógia,...

Pharma productivity gap: explosion in resources only!

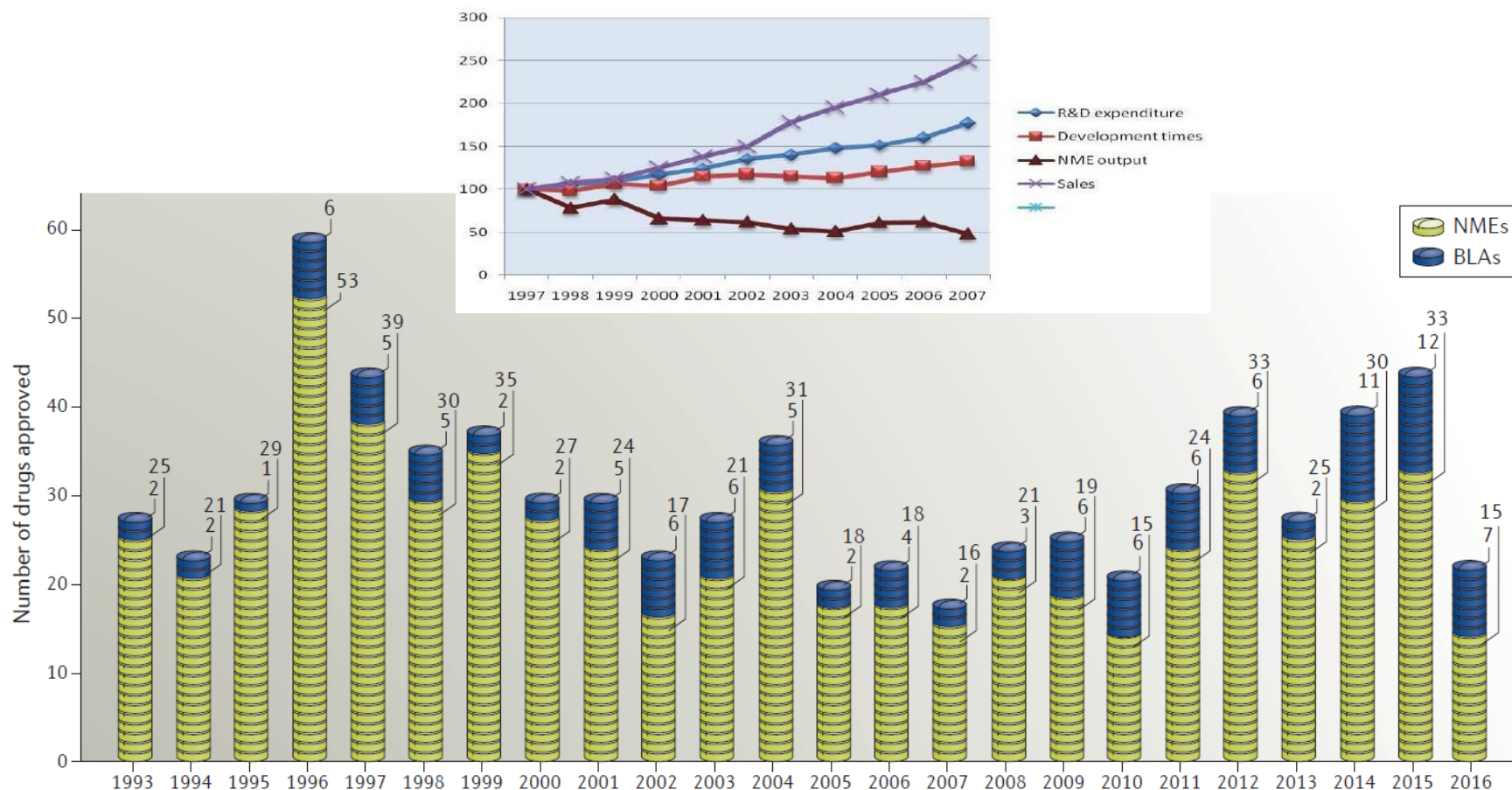
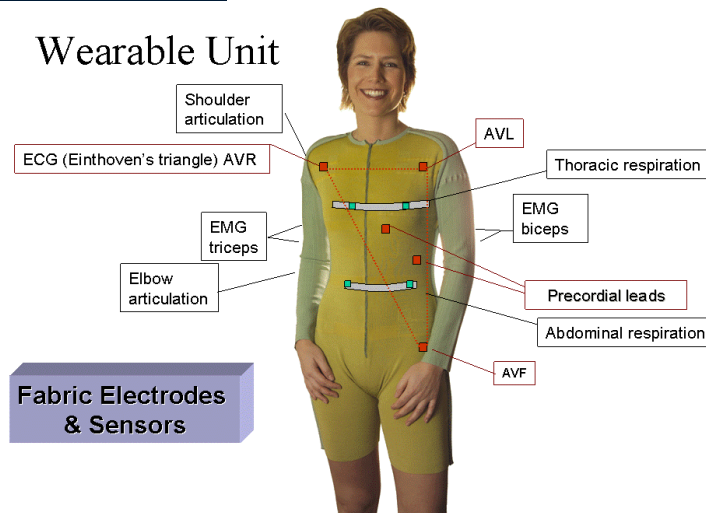
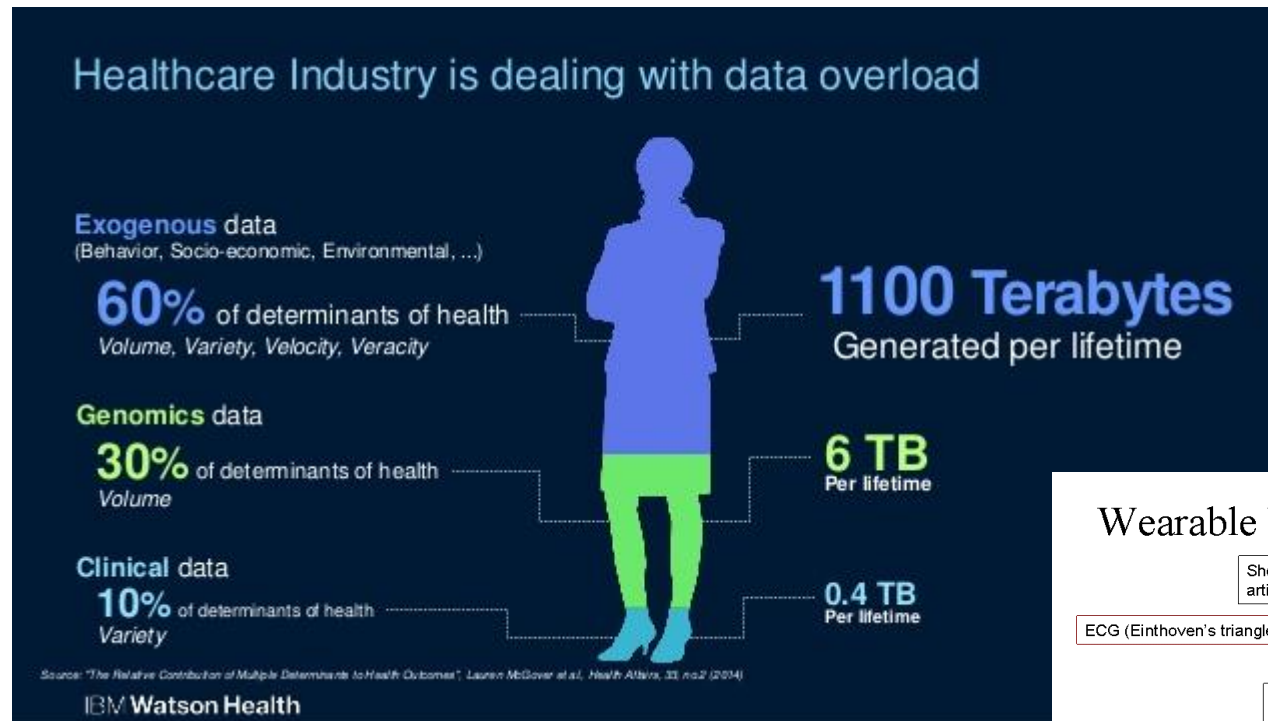


Figure 1 | **Novel FDA approvals since 1993.** New molecular entities (NMEs) and biologics licence applications (BLAs) approved by the Center for Drug Evaluation and Research (CDER) since 1993 (see also

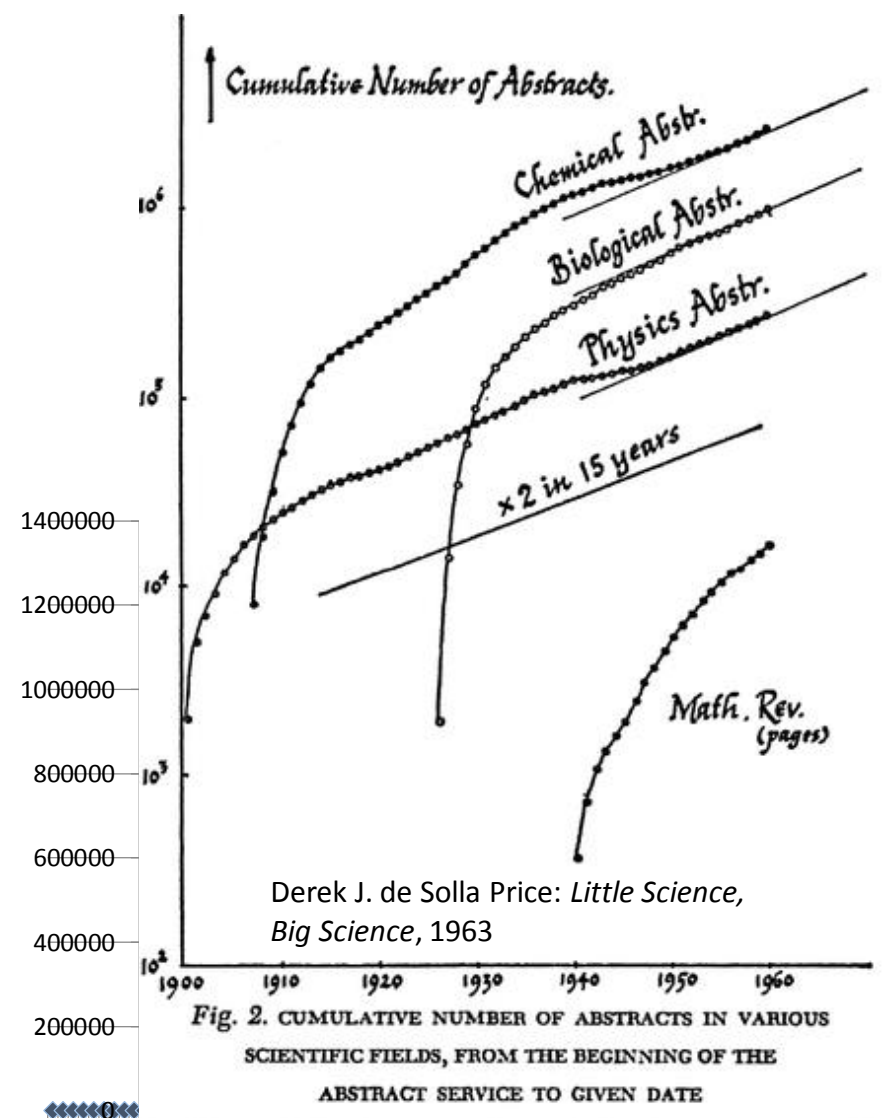
TABLE 1). Approvals by the Center for Biologics Evaluation and Research (CBER) are not included in this drug count (see TABLE 3). Data are from Drugs@FDA.

Mullard, A., 2017. 2016 FDA drug approvals. *Nature Reviews Drug Discovery*, 16(2), pp.73-76.

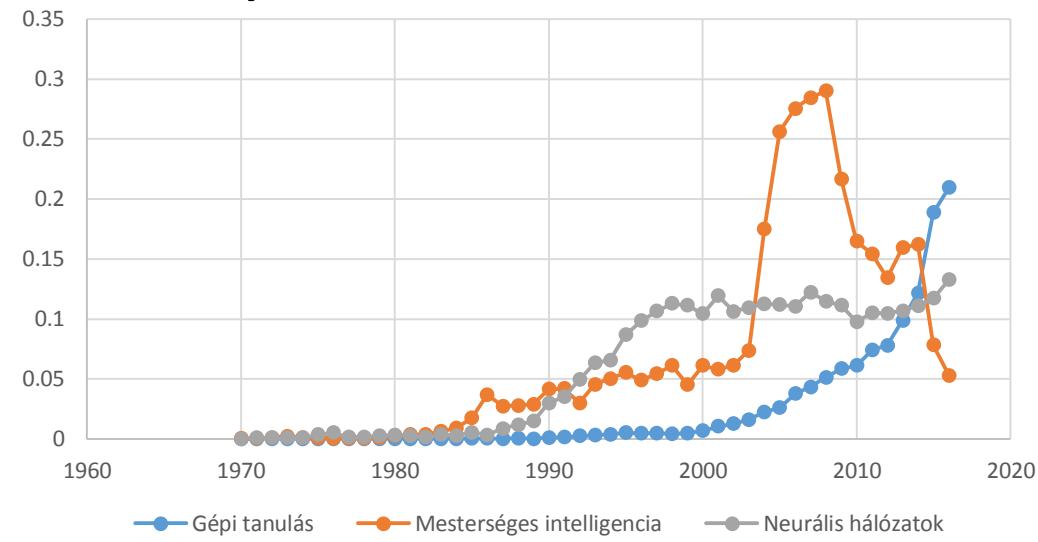
Data: Big data in life sciences



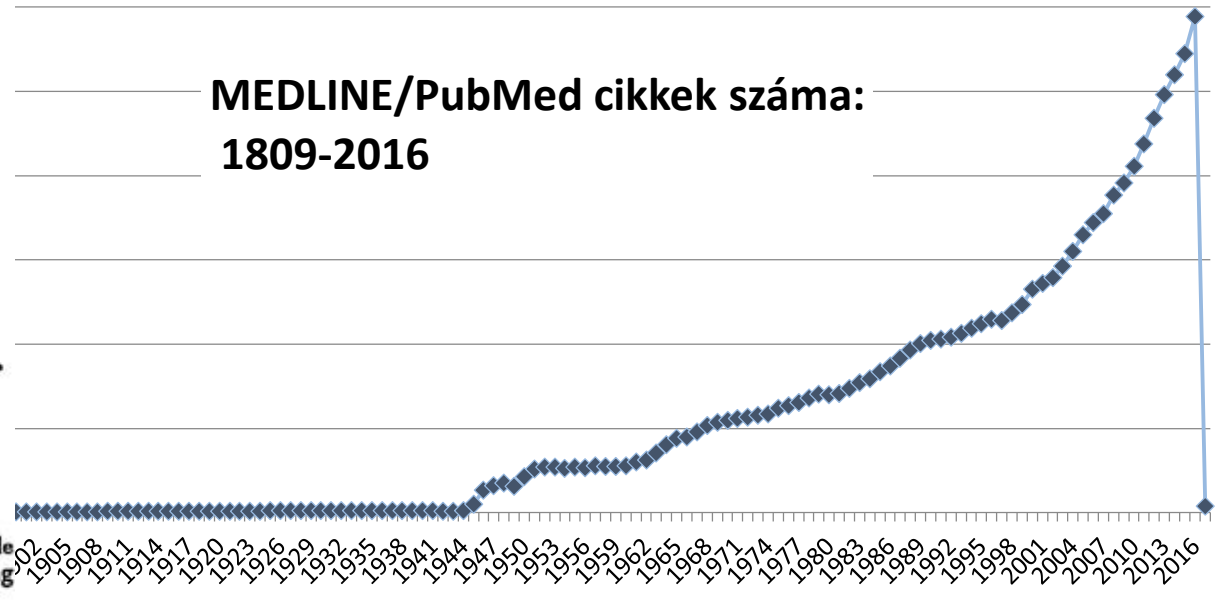
Free text knowledge: publications, patents



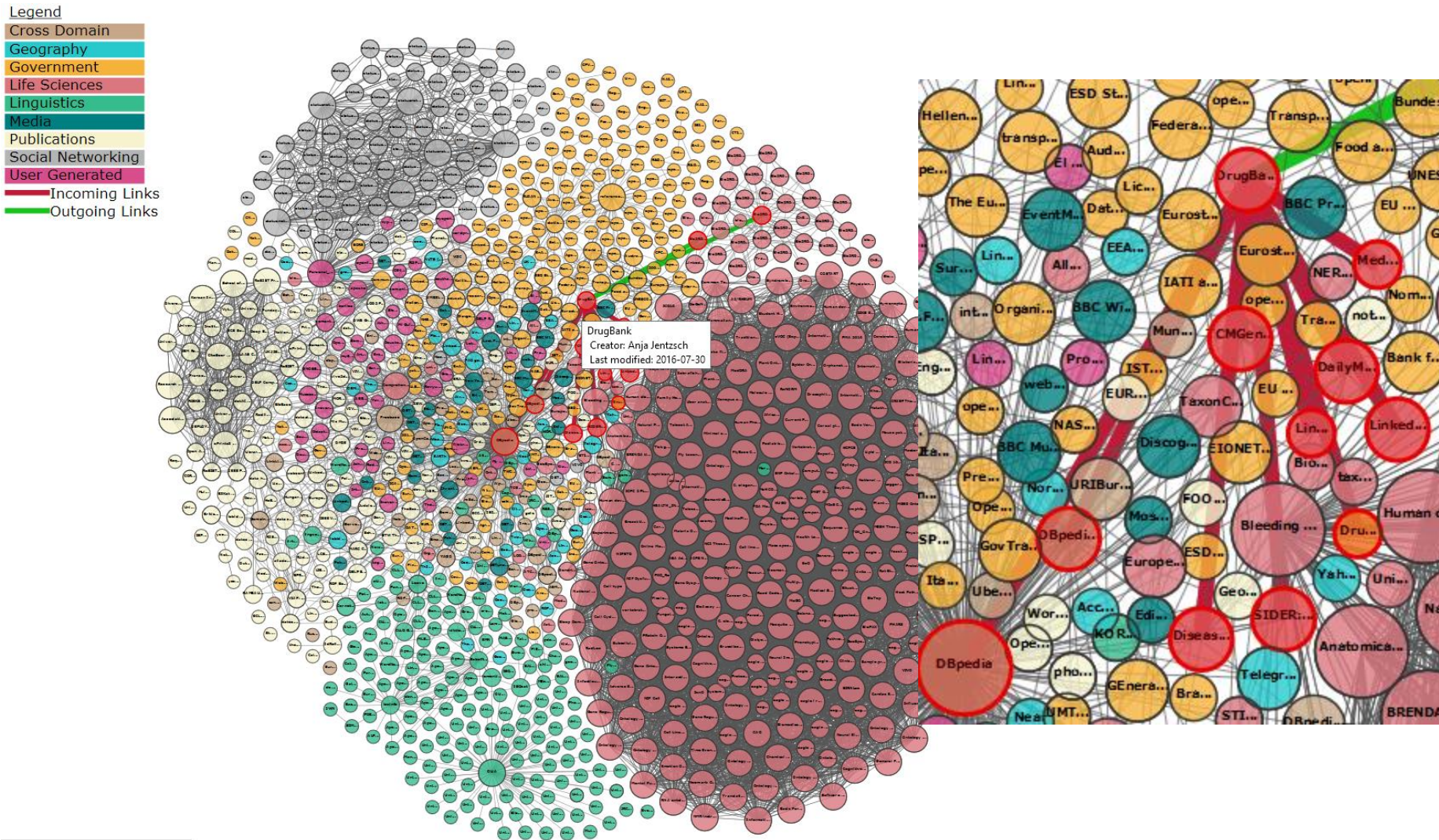
It will be noted that after an initial period of rapid expansion to a stable growth rate, the number of abstracts increases exponentially, doubling in approximately 15 years.



MEDLINE/PubMed cikkek száma:
1809-2016

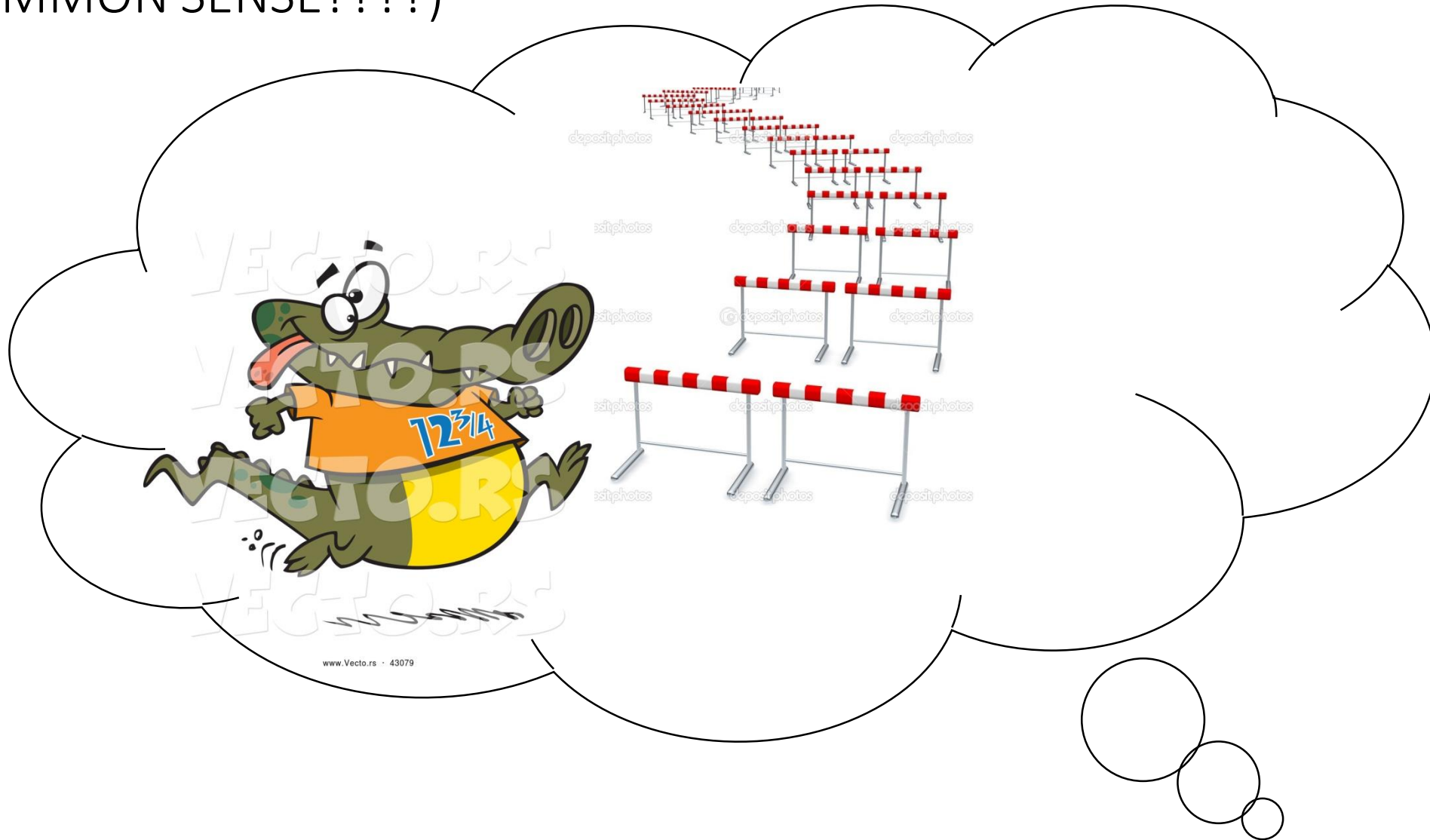


Knowledge: Linked open data



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

WHY CAN'T MY COMPUTER UNDERSTAND ME? (COMMON SENSE????)



Paradox of AI: solved = trivial (\sim not intelligent)

?

What have romans ever given us?!

- What have they [ROMANS] ever given us [in return]?!
- XERXES: The aqueduct?
- REG: What?
- XERXES: The aqueduct.
- REG: Oh. Yeah, yeah. They did give us that. Uh, that's true. Yeah.
- COMMANDO #3: And the sanitation.
- LORETTA: Oh, yeah, the sanitation, Reg. Remember what the city used to be like?
- REG: Yeah. All right. I'll grant you the aqueduct and the sanitation are two things that the Romans have done.
- MATTHIAS: And the roads.
- REG: Well, yeah. Obviously the roads. I mean, the roads go without saying, don't they? But apart from the sanitation, the aqueduct, and the roads--
- COMMANDO: Irrigation.
- XERXES: Medicine.
- COMMANDOS: Huh? Heh? Huh...
- COMMANDO #2: Education.
- COMMANDOS: Ohh...
- REG: Yeah, yeah. All right. Fair enough.
- COMMANDO #1: And the wine.
- COMMANDOS: Oh, yes. Yeah...
- FRANCIS: Yeah. Yeah, that's something we'd really miss, Reg, if the Romans left. Huh.
- COMMANDO: Public baths.
- LORETTA: And it's safe to walk in the streets at night now, Reg.
- FRANCIS: Yeah, they certainly know how to keep order. Let's face it. They're the only ones who could in a place like this.
- COMMANDOS: Hehh, heh. Heh heh heh heh heh heh.
- REG: All right, but apart from the sanitation, the medicine, education, wine, public order, irrigation, roads, a fresh water system, and public health, what have the Romans ever done for us?
- XERXES: Brought peace.
- REG: Oh. Peace? Shut up!

What have AI ever given us?

- Search methods: internet search, route finding
- Logic: software testing
- Linguistics: [real-time] translation, compiler technologies
- Decision theory: expert systems
- Game theory: economics, computer games
- Recommendation systems: web shops, movies,...
- Unbiased uncertain reasoning methods: human biases
- Machine learning: function approximation, online learning
- Causality research
- **Collaboration: open AGI**

Phases/paradigms of AI: open AI research

Open AGI: collaboration

- Linked Open Data (LOD)
- Open collaborative environments
 - openAI
 - <https://openai.com/>
 - fastAI
 - <https://www.fast.ai/>
 - ...
- Open society
 - Ethics guidelines for trustworthy AI
 - <https://ec.europa.eu/digital-single-market/en/news/ethics-guidelines-trustworthy-ai>

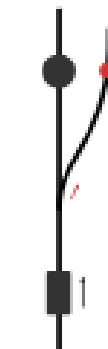
Phases/paradigms of AI:
ethics for AI, existential risk

Intelligent decisions: the trolley problem

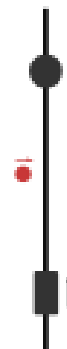
- [Recall: self-driving cars]
- There is a runaway [trolley](#) barreling down the [railway tracks](#). Ahead, on the tracks, there are five people tied up and unable to move. The trolley is headed straight for them. You are standing some distance off in the train yard, next to a lever. If you pull this lever, the trolley will switch to a different set of tracks. However, you notice that there is one person on the side track. You have two options:
 - (1) Do nothing, and the trolley kills the five people on the main track.
 - (2) Pull the lever, diverting the trolley onto the side track where it will kill one person.

Which is the most ethical choice?

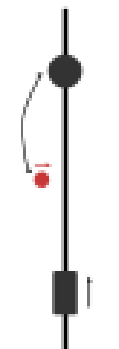
[../wiki/Trolley_problem]



the switch
Foot, 1967



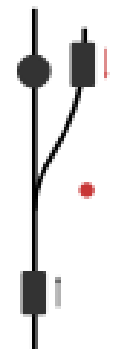
the fat man
Thomson, 1978



the fat villain



the loop
Carr, 1987



the man in the yard
Unger, 1982

AGI: ethical/moral dilemmas

- Mass unemployment.
- Safety AI/provably beneficial/trustworthy AI.
- The value alignment problem.
- Political/civilization-level risk.
 - Social credit systems.
 - https://en.wikipedia.org/wiki/Social_Credit_System
- Existential risk.

Summary

- Resources: vote for podcasts
- Examples for the phases/paradigms of AI
- Strong AI
- Superintelligence/singularity: intelligence explosion
- What AI already gave us:
 - rational models of (narrow) intelligence
 - broad range of theories and technologies
 - **collaboration: open AGI**