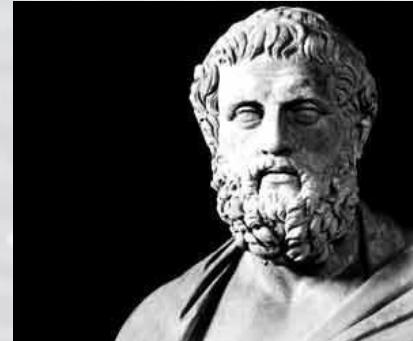


Foundations / A (Brief) History of AI

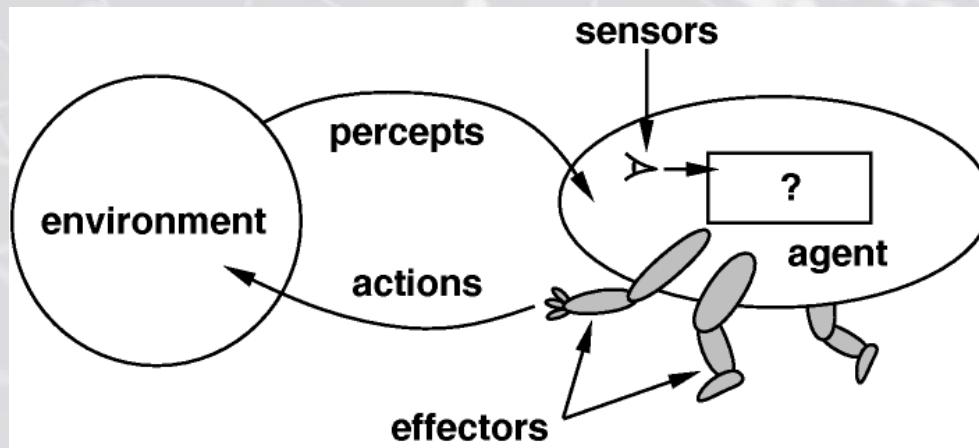
What is AI?

“Numberless are the world’s wonders – but non more wonderful than mankind” – Sophocles, Antigone



(*) “AI” encompasses two general domains:

- (1) **Behavior** (acting “humanly”)
- (2) **Rationality** (acting and thinking “rationally”)



What is AI?

Rational Agent Model for AI:

(*) A **rational agent** “behaves as well as possible.”

“Strong AI” vs. “Weak AI”

Strong AI: Machines are actually “thinking”

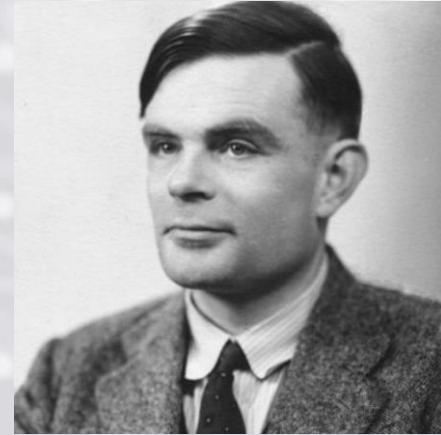
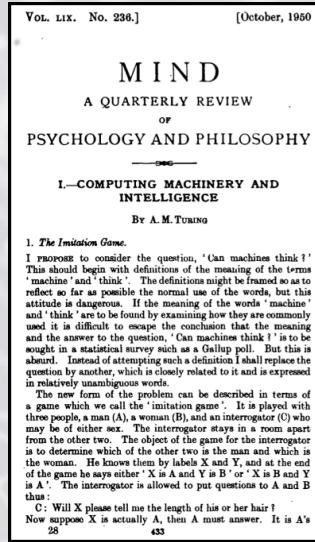
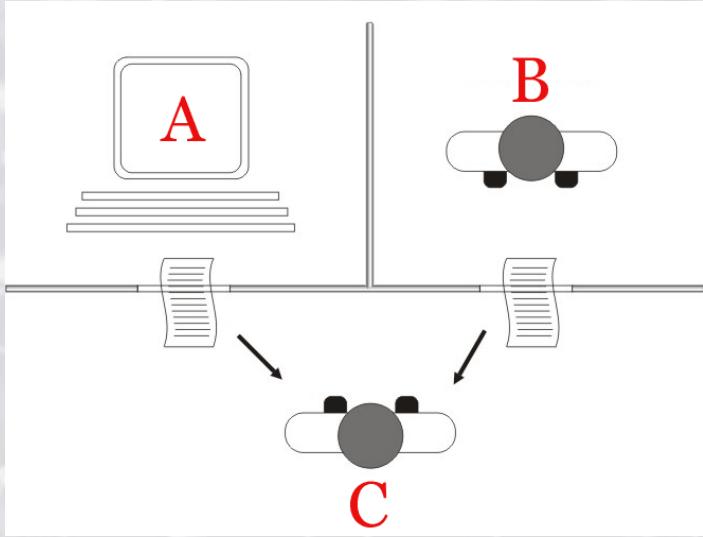
Weak AI: Machines *act* as if they were intelligent

	Weak Artificial Intelligence	Strong Artificial Intelligence
Definition	Machine cannot complete tasks on its own but is made to look intelligent.	Machine can actually think & perform tasks on its own just like a human.
Key Differences	Acts upon and is bound by the rules imposed on it	Can think and function very comparable to human beings.
Current Status	Advanced Stage	Initial Stage
Examples	Self-drive cars, Siri	No proper examples for Strong AI

- (*) Most (but not all) practitioners believe weak AI is (sometimes trivially) achievable – but disagree about the feasibility of strong AI.
- (*) Few mainstream researchers believe that anything significant hinges on the outcome of this debate.

What is AI?

(*) Turing Test / “Total” Turing Test (1950) <https://www.csee.umbc.edu/courses/471/papers/turing.pdf>



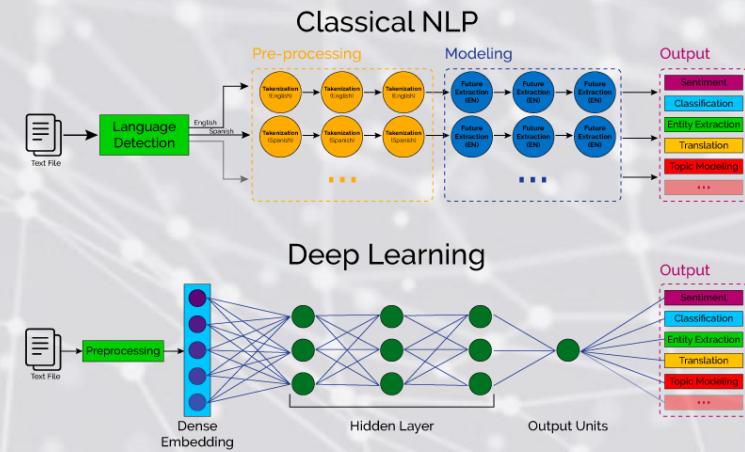
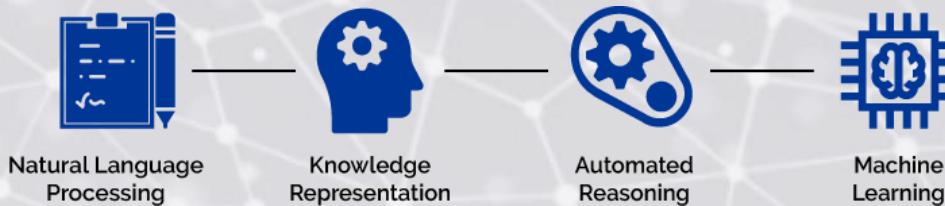
(*) Again, few researchers believe the Turing test is crucial to the future development of AI – however, many fascinating & canonical issues are raised by Turing in this paper on the nature of AI.

Is intelligence a well-defined concept?

(*) A 1955 study found that in 19/20 “expert disciplines”, an elementary mathematical model (e.g. regression, naïve Bayes) outperformed human practitioner.

Aspects of AI & AI Systems

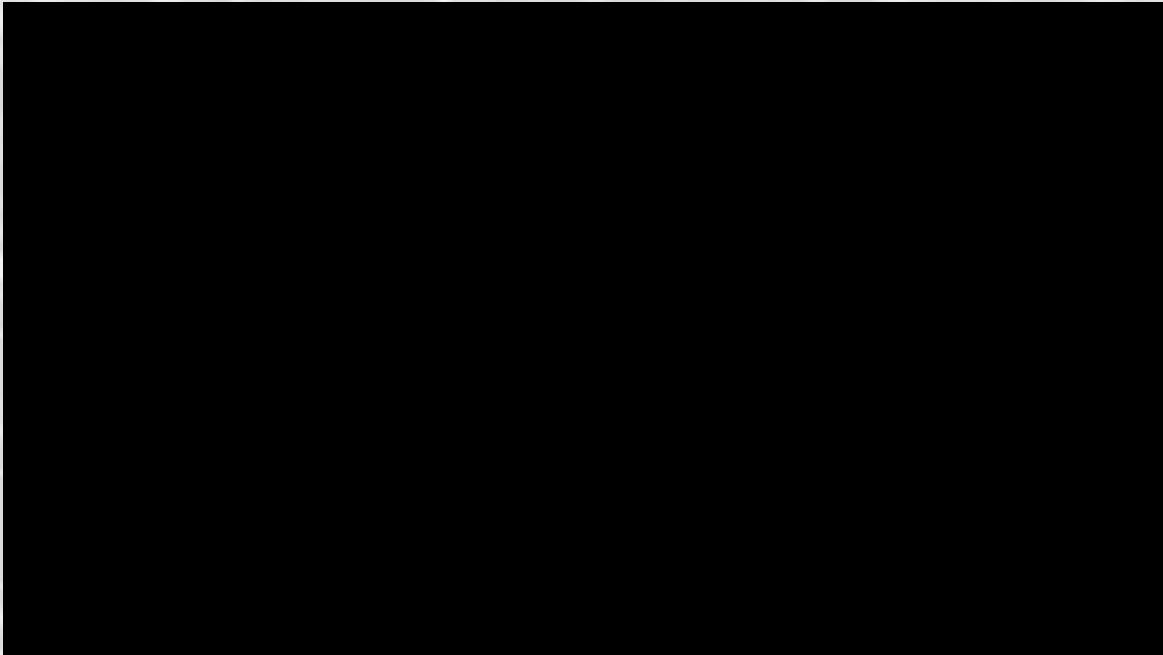
- (*) NLP: Natural Language Processing
- (*) Knowledge Representation: Store/retrieve what is known
- (*) Automated Reasoning: Use stored information for inference/deduction
- (*) Machine Learning: Detect/extrapolate patterns
- (*) Agent's Representation of the World: Environment-knowledge representation



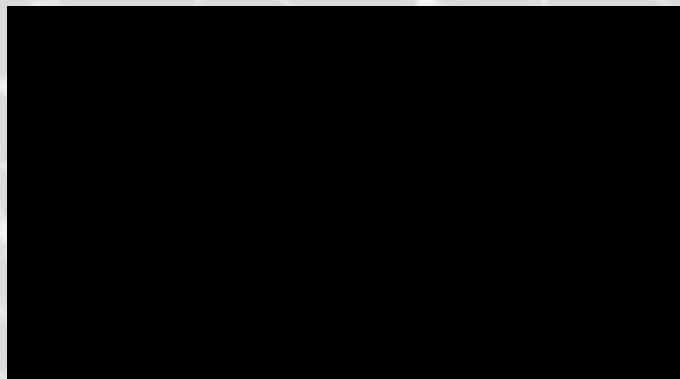
(Computer Vision + Robotics (sensors, actuators) + embodiment = “**Total Turing Test**”

Aspects of AI & AI Systems

(*) Sophia the robot (Hanson Robotics)



Ben Goertzel



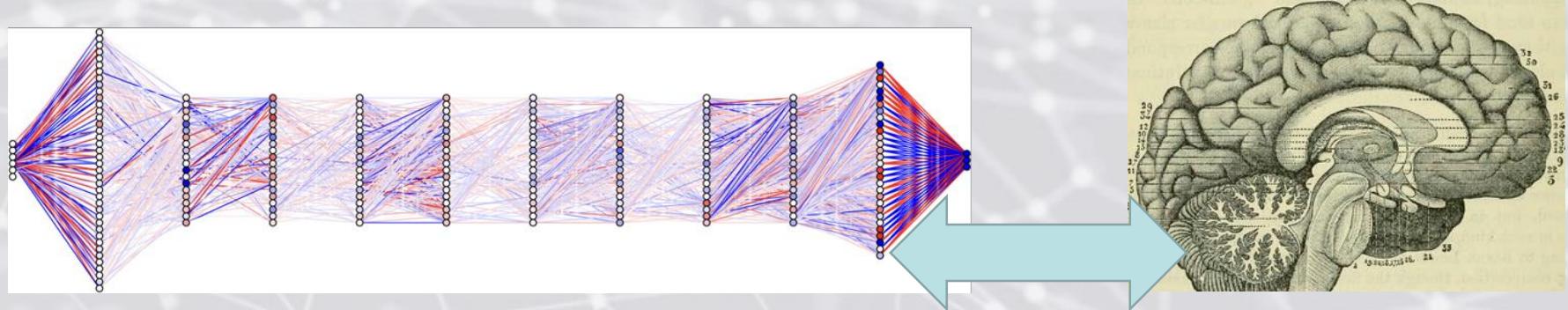
<https://singularitynet.io/> <http://goertzel.org/>

Related Disciplines

Neuroscience: Direct study of nervous system (brain functions, etc.)

Psychology: Emphasis on Behaviorism (Skinner) – difficult to directly test.

Cognitive Science: “AI + Psychology” (use testable theories)



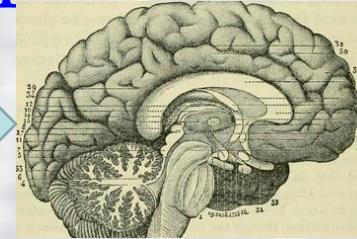
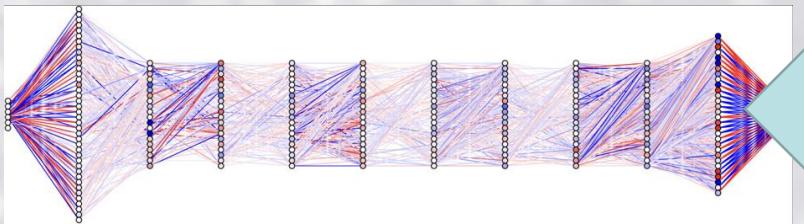
(*) Most researchers today accept the basic distinction between:

Phenomenal / observable
observable works in AI

vs.
Phenomenal
works in physical brains

(*) Nevertheless, many important biological models have fruitfully inspired models in AI (e.g. computer vision)

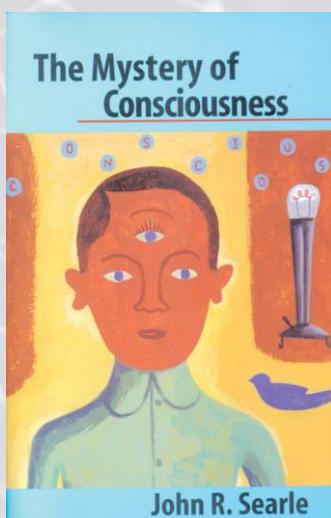
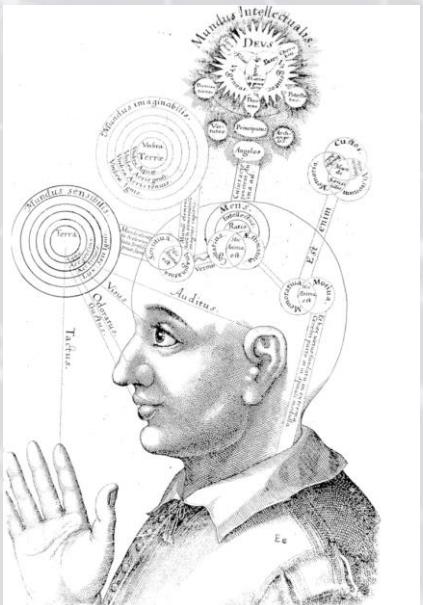
Related Disciplines



(*) Still though, rarely (to date) do researchers begin with the premise: “let’s build a biological brain”

Why? Because we still don’t know how a biological brain works!

Consciousness, for example, is still largely a mystery.



David Chalmers on consciousness:
<https://www.youtube.com/watch?v=uhRhtFhNzQ>

Computation Comparison

(*) Current computation benchmarks for computers are near that of the brain.

Q: But is computation enough? Probably not.

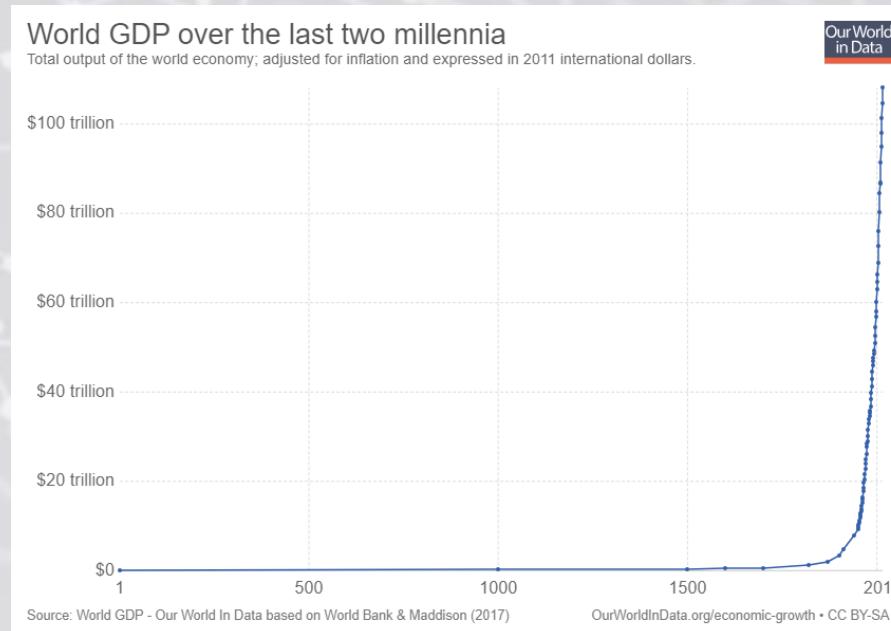
	Brain	Computer
Number of Processing Units	$\approx 10^{11}$	$\approx 10^9$
Type of Processing Units	Neurons	Transistors
Form of Calculation	Massively Parallel	Generally Serial
Data Storage	Associative	Address-based
Response Time	$\approx 10^{-3}$ s	$\approx 10^{-9}$ s
Processing Speed	Very Variable	Fixed
Potential Processing Speed	$\approx 10^{13}$ FLOPS ¹⁴	$\approx 10^{18}$ FLOPS
Real Processing Speed	$\approx 10^{12}$ FLOPS	$\approx 10^{10}$ FLOPS
Resilience	Very High	Almost None
Power Consumption per Day	20W	300W ¹⁵

Accelerated Growth

(*) A few hundred thousand years ago, in early human prehistory, growth was so slow that it took on the order of one million years for human productive capacity to increase sufficiently to sustain an additional one million individuals live at subsistence level.

(*) By 5000 BC, following the **Agricultural Revolution**, the rate of growth had increased to the point where the same amount of growth took just two centuries.

(*) Today, following the **Industrial Revolution**, the world economy grows on average by that amount of ninety minutes.



Accelerated Growth & Computation

1 The accelerating pace of change ...

Agricultural Revolution → 8,000 years → Industrial Revolution → 120 years → Light-bulb → 90 years → Moon landing → 22 years → World Wide Web → 9 years → Human genome sequenced

2 ... and exponential growth in computing power ...

Computer technology, shown here climbing dramatically by powers of 10, is now progressing more each hour than it did in its entire first 90 years

COMPUTER RANKINGS

By calculations per second per \$1,000



Analytical engine
Never fully built, Charles Babbage's invention was designed to solve computational and logical problems



Colossus

The electronic computer, with 1,500 vacuum tubes, helped the British crack German codes during WW II



UNIVAC I

The first commercially marketed computer, used to tabulate the U.S. Census, occupied 943 cu. ft.



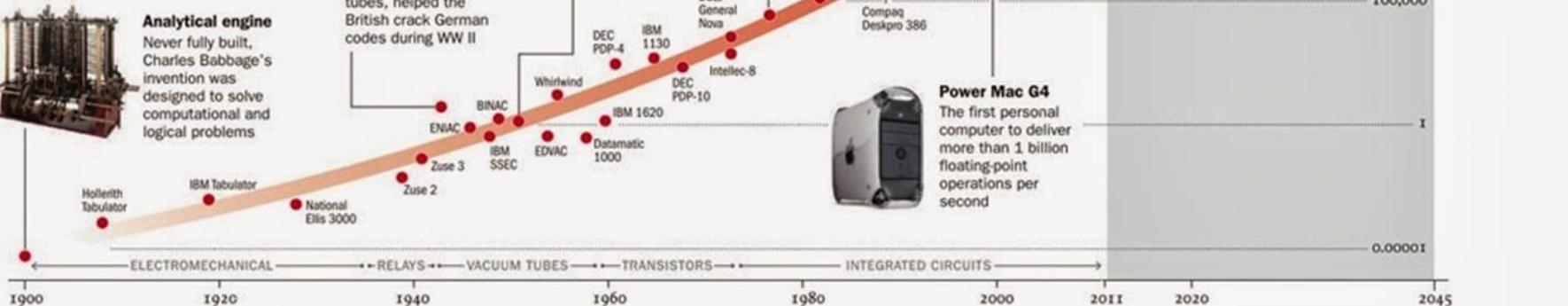
Apple II

At a price of \$1,298, the compact machine was one of the first massively popular personal computers



Power Mac G4

The first personal computer to deliver more than 1 billion floating-point operations per second



3 ... will lead to the Singularity

Accelerated Growth & Computation



TYPE I CIVILIZATION harnesses all the resources of a planet. Carl Sagan estimated that Earth rates about 0.7 on the scale.

TYPE II CIVILIZATION harnesses all the radiation of a star. Humans might reach Type II in a few thousand years.

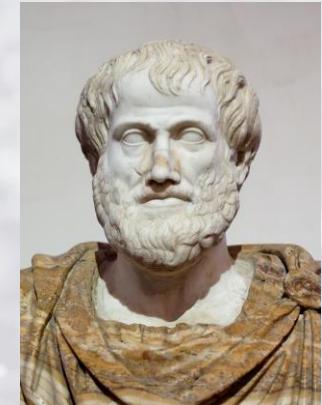
TYPE III CIVILIZATION harnesses all the resources of a galaxy. Humans might reach Type III in a few hundred thousand to a million years.

— The Kardashev Scale —

Rational Thinking

(*) Early codification of “right thinking – Aristotle (syllogisms)

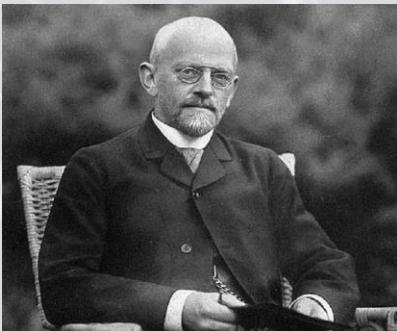
E	no M is P	$\forall x Mx \Rightarrow \neg Px$	figure 4	$\forall x Px \Rightarrow Mx$	figure 4 swap S,P
A	all S is M	$\forall x Sx \Rightarrow Mx$		$\forall x Mx \Rightarrow \neg Sx$	
E	no P is S	$\forall x Px \Rightarrow \neg Sx$		$\forall x Sx \Rightarrow \neg Px$	
A	all M is P	$\forall x Mx \Rightarrow Px$	$\exists x Px \wedge Mx$	$\exists x Sx \wedge Mx$	
I	some S is M	$\exists x Sx \wedge Mx$	$\forall x Mx \Rightarrow Sx$	$\forall x Mx \Rightarrow Px$	
I	some P is S	$\exists x Px \wedge Sx$	$\exists x Sx \wedge Px$	$\exists x Px \wedge Sx$	
A	all M is P	$\forall x Mx \Rightarrow Px$	$\forall x Px \Rightarrow \neg Mx$	$\forall x Sx \Rightarrow \neg Mx$	
E	no S is M	$\forall x Sx \Rightarrow \neg Mx$	$\forall x Mx \Rightarrow Sx$	$\forall x Mx \Rightarrow Px$	
O	some P is not S	$\exists x Px \wedge \neg Sx$	$\exists x Sx \wedge \neg Px$	$\exists x Px \wedge \neg Sx$	
I	some M is P	$\exists x Mx \wedge Px$	$\forall x Px \Rightarrow \neg Mx$	$\forall x Sx \Rightarrow \neg Mx$	
E	no S is M	$\forall x Sx \Rightarrow \neg Mx$	$\exists x Mx \wedge Sx$	$\exists x Mx \wedge Px$	
O	some P is not S	$\exists x Px \wedge \neg Sx$	$\exists x Sx \wedge \neg Px$	$\exists x Px \wedge \neg Sx$	



Logic: Is there a set of (finite, discoverable) laws that govern the operation of the mind?

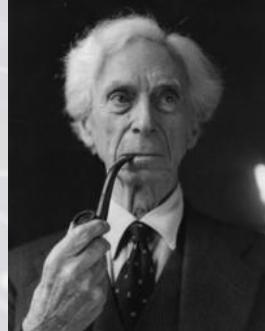
Issues with Logicist-AI methodology: Difficult to translate all real-world problems into symbolic form; tractability; some argue incompleteness (Gödel) renders logicist approach to mathematics/AI futile.

Hilbert \rightarrow Russell/Whitehead \rightarrow Gödel

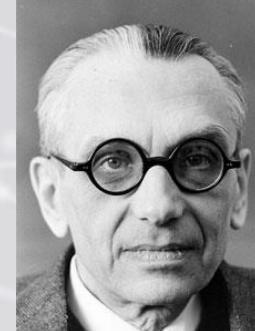


Hilbert's

“Millennium Problems”



“Principia”

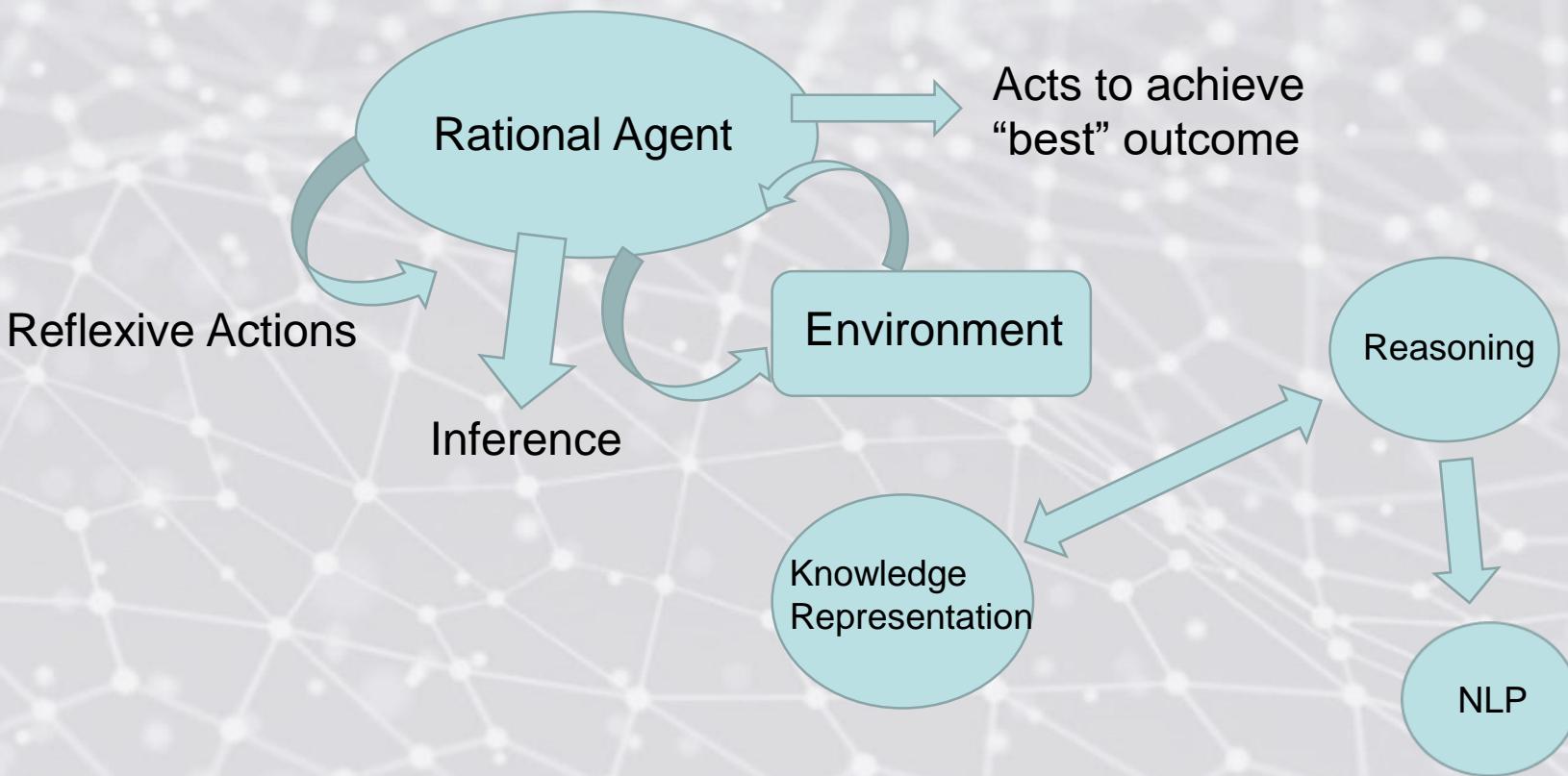


Incompleteness

Fundamental Fact: The attempt to establish an indubitable foundation for all of mathematics/science through logic directly inspired the gedanken experiments that led to the invention of the computer and the inception of AI as a formal discipline.

Rational Thinking

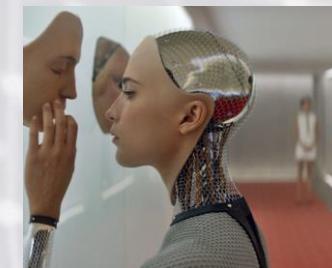
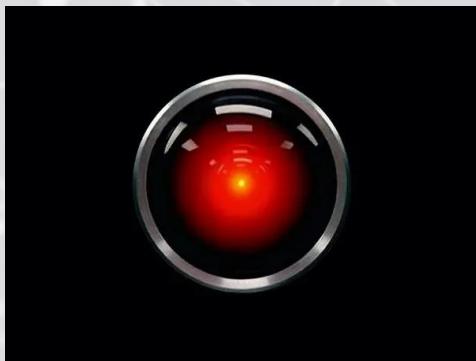
A more complete rational agent?



“Perfect rationality” – a good starting point for AI.

A Brief History of AI

Themes of AI & “Dangerous knowledge” in culture.



A Brief History of AI

What is the nature of knowledge, where does it come from?

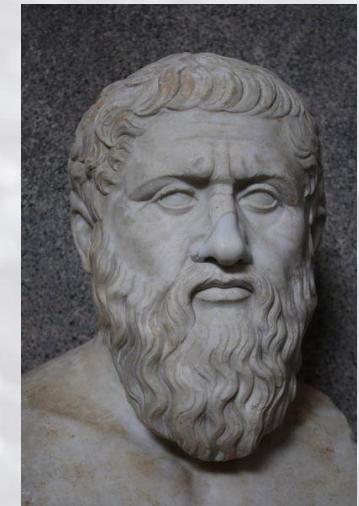
(Epistemology) Plato – Meno dialogue & *a priori* knowledge.

Induction vs. Deduction (Reasoning)

Q's: (*) Can we simply learn a huge list of inference rules?

(*) Does a look-up table constitute intelligence?

(*) How to proceed from knowledge to action?



Hobbes: “Reasoning is like numerical computation” (1651, Leviathan)

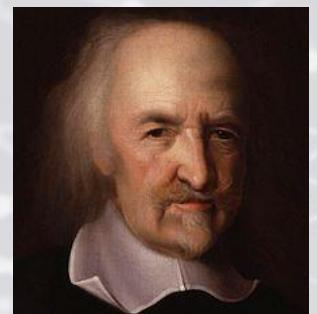
Leonardo: Designs for mechanical calculator (15C);

IBM built a replica in 1968



Leibniz: “Step Reckoner” – could perform all 4 arithmetic operations (1694)

Also conceived of a computational machine operating on concepts!



A Brief History of AI

Hobbes: “Reasoning is like numerical computation” (1651, Leviathan)

Leonardo: Designs for mechanical calculator (15C);

IBM built a replica in 1968



Leibniz: “Step Reckoner” – could perform all 4 arithmetic operations (1694)

Also conceived of a computational machine operating on concepts!

Pascal: “Pascaline” (1652): Arithmetic / Mechanical Calculator

Descartes: *Cogito* (Meditations, 1641)



Rationalists – Dualism: Part of Mind/Spirit outside body & external world

Materialism: Brain constitutes mind

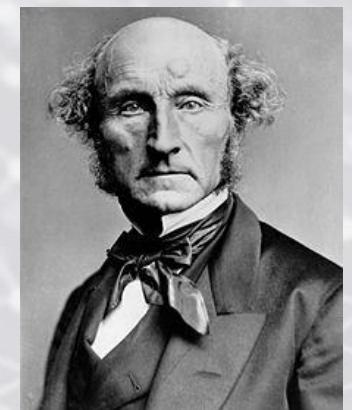
Sources and Nature of Knowledge

Empiricism: Bacon (16C), Locke, Hume



Inductive Reasoning: General rules acquired by repeated exposure/association. (can we prove induction?)

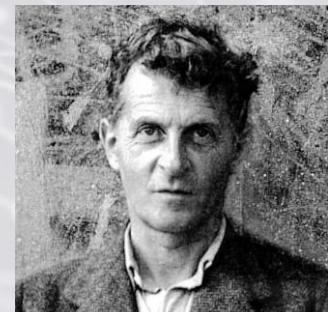
Paradoxes of Induction: Black Swan



Utilitarianism: Mill (19C) – Ethics: maximize/quantify utility

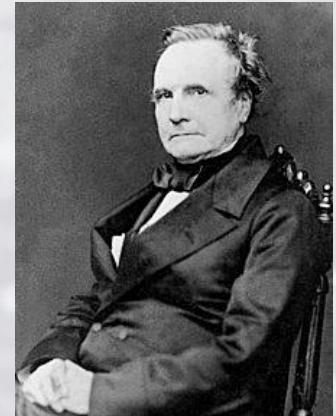
Phenomenology: Husserl -> Heidegger: attempts to square subjective experiences with rationalism.

Logical Positivism (20C): Russell -> Wittgenstein -> Carnap – (Verificationism) only meaningful problems are those solvable by logical analysis (against metaphysics).

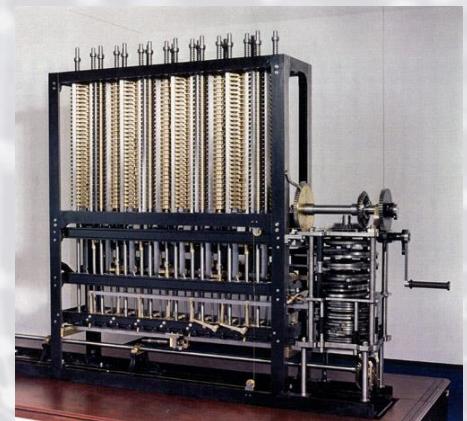


A Brief History of AI

C. Babbage: Difference Engine (1820s) – computes polynomial coefficients from Newton's Difference (classical interpolation).



Analytical Engine (AE, 1837): Proposed general-purpose computer; integrating loops, memory, logic unit (Turing-complete machine).



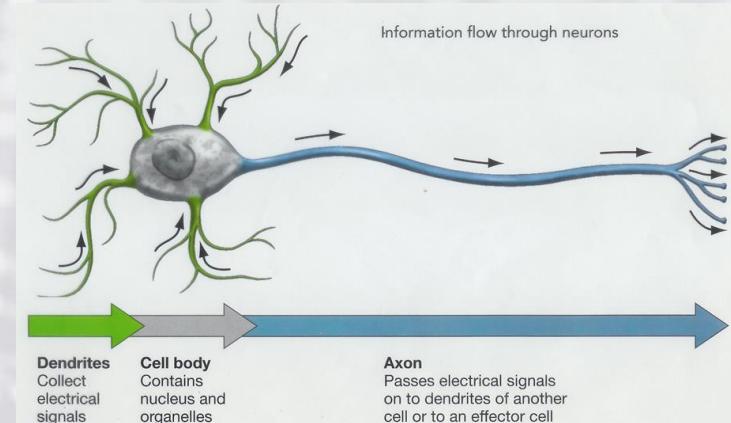
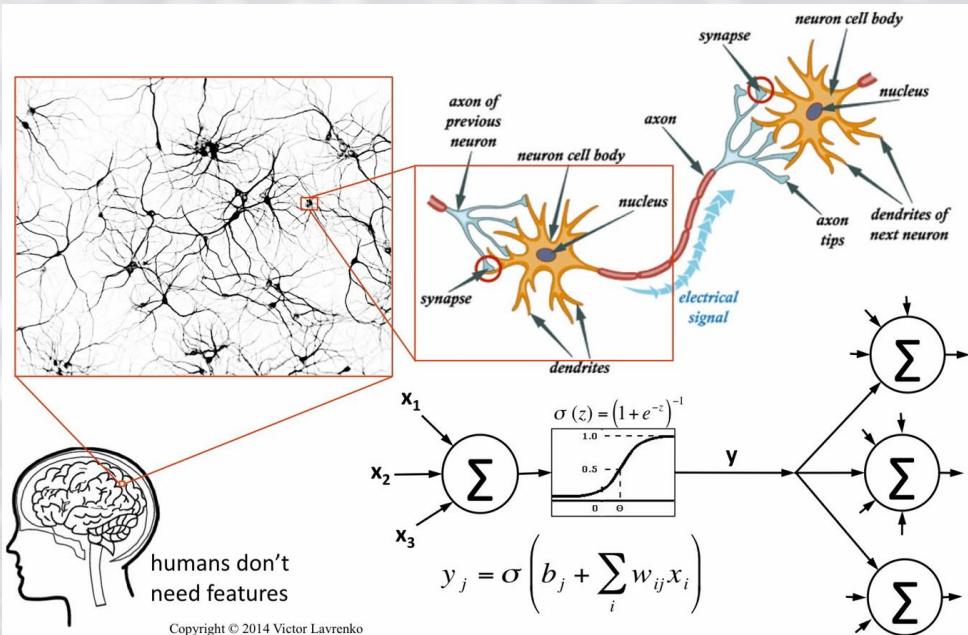
Ada Lovelace: (first “programmer”?) Wrote programs for AE; speculated about creative ability of AI (chess/music).

Leonardo: Designs for mechanical calculator (15C);

“[The Analytical Engine] might act upon other things besides number, were objects found whose mutual fundamental relations could be expressed by those of the abstract science of operations, and which should be also susceptible of adaptations to the action of the operating notation and mechanism of the engine... Supposing, for instance, that the fundamental relations of pitched sounds in the science of harmony and of musical composition were susceptible of such expression and adaptations, the engine might compose elaborate and scientific pieces of music of any degree of complexity or extent.” – Ada Lovelace



Gestation 1943-1955

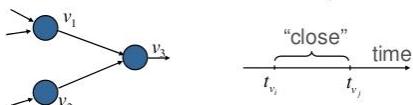


Hebb's Postulate

“When an axon of cell A is near enough to excite a cell B and repeatedly or persistently takes part in firing it, some growth process or metabolic change takes place in one or both cells such that A’s efficiency, as one of the cells firing B, is increased.”

- In other words: if two neurons fire “close in time” then strength of synaptic connection between them increases.

$$\Delta w_{ij}(t) = \eta v_i v_j g(t_{v_i}, t_{v_j})$$

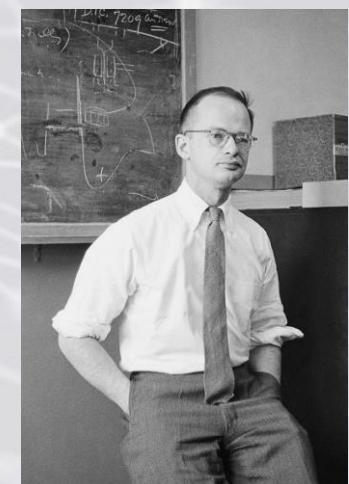
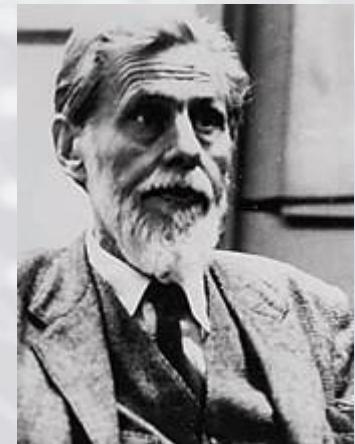
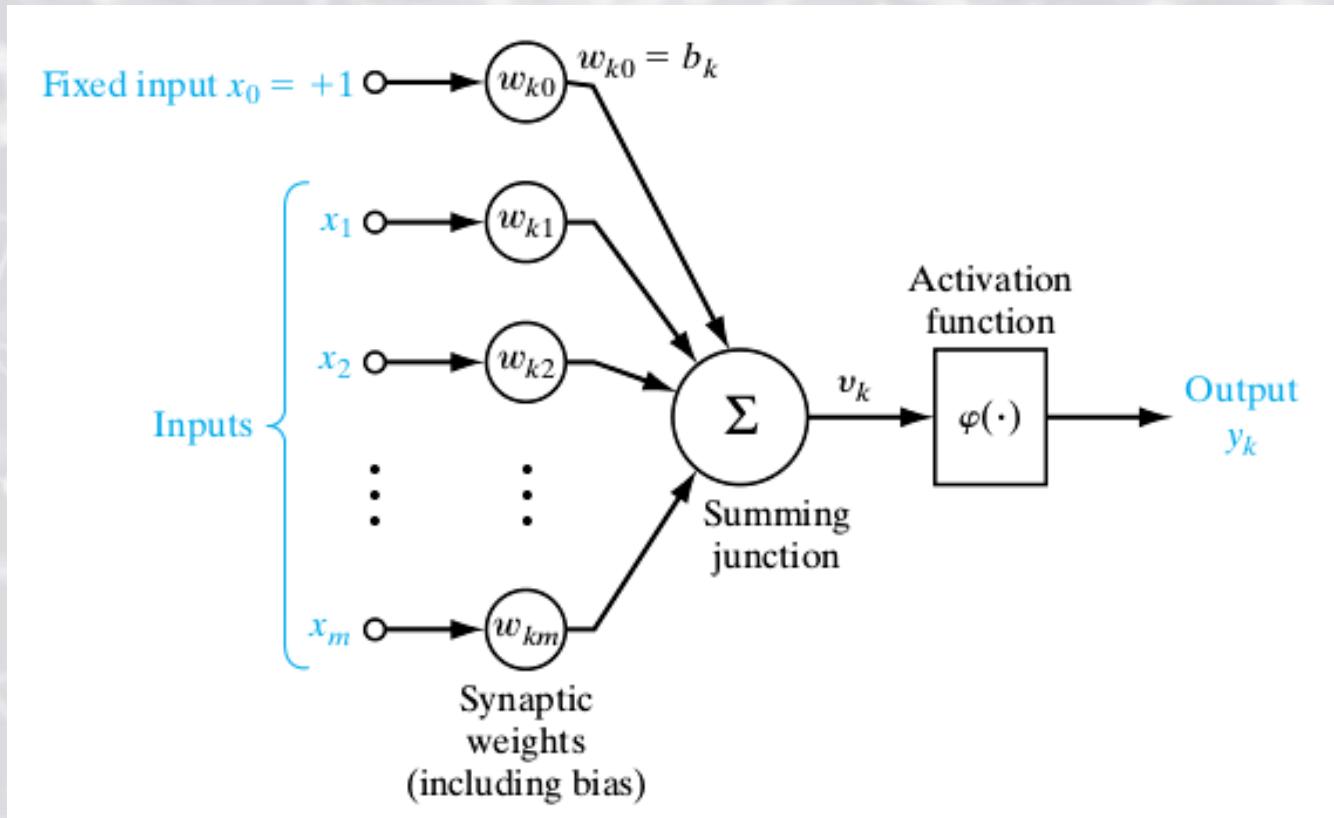


- Weights reflect correlation between firing events.



Gestation 1943-1955

McCulloch & Pitts Neuron Model (1943)



Gestation

(1956) Dartmouth Workshop on AI

John McCarthy: Founded Stanford AI Lab, invented Lisp, Advice Taker program

Alan Newell: RAND/CMU, received Turing award

Claude Shannon: Founder of information theory

Herbert Simon: CMU, received Nobel prize in Economics, Turing award

Marvin Minsky: initiated MIT AI Lab, built SNARC (early NN machine)



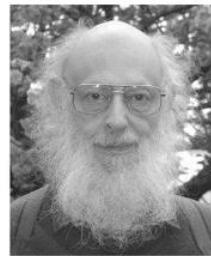
John McCarthy



Marvin Minsky



Claude Shannon



Ray Solomonoff



Alan Newell



Herbert Simon



Arthur Samuel



Oliver Selfridge



Nathaniel Rochester



Trenchard More

Gestation

(*) The Dartmouth Workshop charter:

“We propose a 2 month, 10 man study of artificial intelligence be carried out ... The study is to proceed on the basis of the conjecture that every aspect of learning or any other feature of intelligence can in principle be so precisely described that a machine can be made to simulate it. An attempt will be made to find out how to make machines that use language, form abstractions and concepts, solve kinds of problems now reserved for humans, and improve themselves. We think that a significant advance can be made in one or more of these problems if a carefully selected group of scientists work on it together for a summer.”

(*) A more tempered perspective: I.J. Good in 1965

(British mathematician who worked with Turing at Bletchley Park):

“Let an ultra-intelligent 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 ultra- intelligent machine could design even better machines; there would unquestionably be an “intelligence explosion,” and the intelligence of man would be left far behind. Thus the first ultra-intelligent 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.”



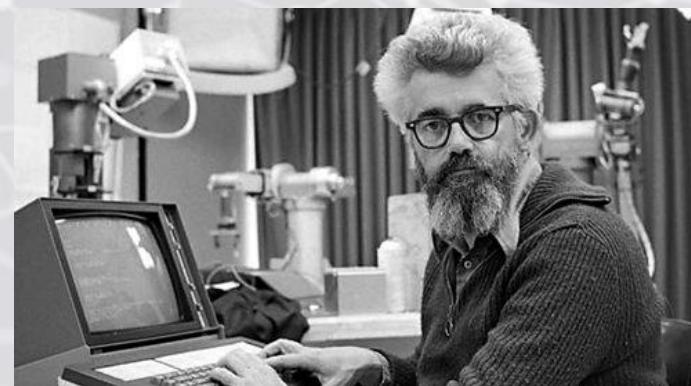
1952-1969

Early Enthusiasm/Great Expectations

(*) For the two decades following the Dartmouth workshop, AI research was largely dominated by the workshop participants and their immediate colleagues.

(*) John McCarthy referred to this era as the “Look, Ma, no hands!” era of AI research.; during these days researchers built systems designed to refute claims of the form “No machine could ever do X!”

(*) Such skeptical claims were common at the time. To counter them, the AI researchers created small systems that achieved X in a “**microworld**” (a well-defined, limited domain that enabled a pared-down version of the performance to be demonstrated), thus providing a proof of concept and showing that X could, in principle, be done by a machine.



1952-1969

Early Enthusiasm/Great Expectations

(*) LT: **Logic Theorist** (1956) designed by Allen Newell and Herbert Simon, dubbed the “first AI program”, it was deliberately engineered to mimic the problem solving skills of humans. In total it successfully proved 38 of the first 52 theorems from Russell/Whitehead’s *Principia*.

Logic Theorist

In about 12 minutes LT produced, for theorem 2.45:

```
-( $p \vee q \rightarrow \neg p$ )
1.  $A \rightarrow (A \vee B)$ 
2.  $B \rightarrow (p \vee q)$ 
3.  $(A \rightarrow (A \vee B)) \rightarrow (\neg B \rightarrow \neg A)$ 
4.  $(p \rightarrow (p \vee q)) \rightarrow ((\neg (p \vee q)) \rightarrow \neg p)$ 
5.  $\neg(p \vee q) \rightarrow \neg p$ 
Q. E. D.
```

(Theorem 2.45, to be proved.)
(Theorem 2.2.)
(Subst. p for A , q for B in 1.)
(Theorem 2.2.)
(Subst. p for A , $(p \vee q)$ for B in 3.)
(Detach right side of 4, using 2.)

(*) GPS: **General Problem Solver** (1959), also developed by Newell

And Simon; used a separate knowledge representation module – intended as a universal solver machine (any problems expressed as WFFs could be solved, in principle) – limited due to combinatorial explosion.

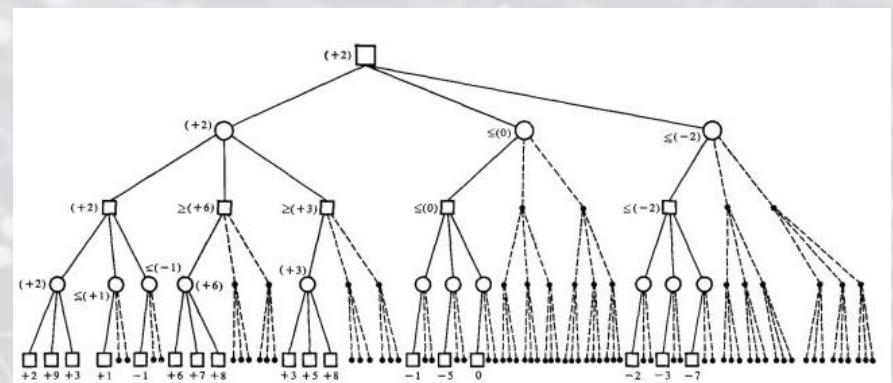
(*) **Geometry Theorem Prover** (IBM, 1959)



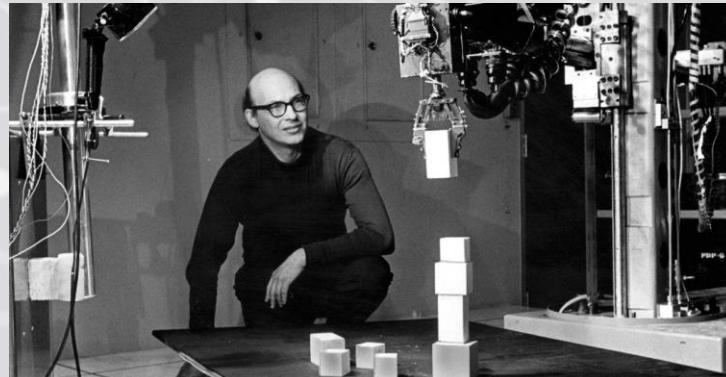
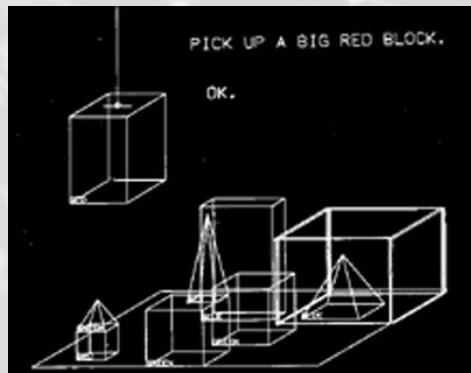
1952-1969

Early Enthusiasm/Great Expectations

(*) **Arthur Samuel** (IBM, Stanford) was an early pioneer in AI (first to coin term “machine learning”); began seminal work on **AI checkers** program in 1959, invents **alpha-beta pruning** and **minimax** algorithms (among others).



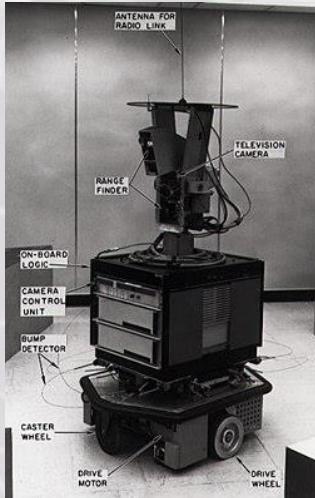
(*) Minsky supervised students in microworld problems (e.g. *blockworld*)



1952-1969

Early Enthusiasm/Great Expectations

(*) **Shakey the Robot**, developed at Stanford, (so named because of its tendency to tremble during operation) demonstrated how logical reasoning could be integrated with perception and used to plan and control physical activity. It was the first general-purpose mobile robot to be able to reason about its actions; project combined research in robotics, computer vision and NLP. As notable for one of the first applications of the **A* algorithm**.



<https://www.youtube.com/watch?v=7bsEN8mwUB8>

(*) The **ELIZA** (MIT, 1964-1966) program showed how a computer could impersonate a Rogerian psychotherapist. ELIZA simulated conversation by using a pattern matching and substitution methodology that gives the illusion of understanding (note that ELIZA is incapable of learning new patterns of speech/words through interaction alone).

Demo: <http://www.manifestation.com/neurotoys/eliza.php3>

1952-1969

Early Enthusiasm/Great Expectations

- (*) Following a number of early successes of applied AI in these microworld domains, enthusiasm was high for AI to solve the “big problems” (e.g. computer vision, NLP, etc.); H. Simon declared (1957): “There are now in the world machines that think, that learn and that create.”
- (*) However, the methods that produce these early successes often proved difficult to extend to a wider variety of problems or to harder problem instances.
- (*) One reason for this is the “**combinatorial explosion**” of possibilities that must be explored by methods that rely on something like exhaustive search. (note that the inception of **computational complexity** as a formal discipline only began in the mid 1960s)
- (*) For instance: to prove a theorem using 5-lines and a deductive system containing 5 axioms, one could simply enumerate the 3,125 possible combinations; proving a 50-line proof by contrast requires **$\sim 8.9 \times 10^{34}$ possible sequences (!)** – which is computationally infeasible for even the fastest supercomputers.

1966-1973: AI Winter

(*) During the U.S.-Soviet space race, the U.S. government funded research in **machine translation**. At the time, researchers believed (naively) that NLP (in addition to computer vision) would be solved in a matter of a few years.

(*) After nearly a decade of funding research in machine translation, researchers discovered that they were still a long way from “solving” the problem. A famous mistranslation encapsulated this difficulty:

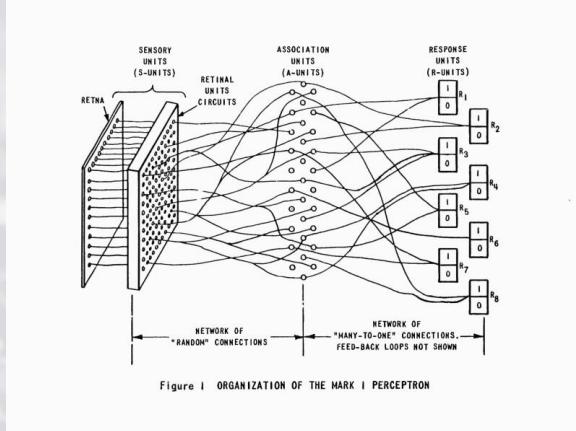
“the spirit is willing, but the flesh is weak” (originally in Russian) was translated: *“the vodka is good but the meat is rotten”!* (in 1966 all funding was halted for this project)

(*) Note that to overcome **the combinatorial explosion** in AI, one needs algorithms that exploit structure in the target domain and take advantage of prior knowledge by using heuristic search, planning, and flexible abstract representations – capabilities that were poorly developed by early AI systems.

(*) The performance of these early systems also suffered because of the poor methods for handling uncertainty, reliance on brittle and ungrounded symbolic representations, **data scarcity**, and severe **hardware limitations** of memory capacity and processor speed.

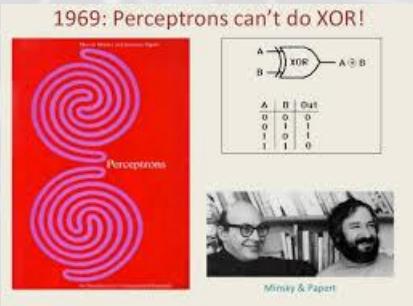
1966-1973: AI Winter

(*) **Rosenblatt** (1962) proved that the **perceptron learning rule** converges to correct weights in a finite number of steps, provided the training examples are linearly separable.



(*) **Minsky and Papert** (1969) proved that perceptrons cannot represent non-linearly separable target functions.

(*) Later it was shown that by using continuous activation functions (rather than thresholds), a fully connected network with a single hidden layer can in principle represent any function (UAT (1989): **universal approximation theorem**). The well-known **backpropagation algorithm** (essential to deep learning) algorithm was later “rediscovered” by **Hinton** et al.



$$\begin{aligned}
 y_i &= f(x_i) \\
 &= f(\sum_j w_{ij} y_j) \\
 E &= \sum_\alpha (y_i^{(\alpha)} - t_i^{(\alpha)})^2 \\
 \delta_i^{(\alpha)} &= \frac{\partial E}{\partial x_i^{(\alpha)}} \\
 &= \frac{\partial E}{\partial y_i^{(\alpha)}} \frac{\partial y_i^{(\alpha)}}{\partial x_i^{(\alpha)}} \\
 &= 2(y_i^{(\alpha)} - t_i^{(\alpha)}) f'(x_i^{(\alpha)}) \\
 \frac{\partial E}{\partial w_{ij}} &= \sum_\alpha \delta_i^{(\alpha)} \frac{\partial x_i^{(\alpha)}}{\partial w_{ij}} \\
 &= \sum_\alpha \delta_i^{(\alpha)} y_j^{(\alpha)}
 \end{aligned}$$

AI in the 1980s: Expert Systems

- (*) The ensuing years saw a great proliferation of **expert systems** (rule-based programs that made simple inferences from a knowledge base of facts, elicited from human domain experts and painstakingly hand-coded in a formal language).
- (*) Hundreds of these expert systems were built; however, the smaller systems provided little benefit, and the larger ones proved expensive to develop, validate, and keep updated, and were generally cumbersome to use.
- (*) At this point, a critic could justifiably bemoan: “*the history of AI research to date, consisting always of very limited success in particular areas, followed immediately by failure to reach the broader goals at which these initial successes seem at first to hint*”; AI became something of an unwanted epithet at this time.

1990s-2000s: Resurgence of AI

(*) By the early 1990s, a second AI winter began to thaw. Optimism was rekindled by the introduction of new techniques, which seemed to offer alternatives to the traditional logicist paradigm (often referred to as *Good Old-Fashioned AI* (GOFAI)) which had reached its apogee in the 1980s.

(*) In particular, two newly popular techniques: **Neural Networks (NNs)** and **Genetic Algorithms** (GAs), promised to overcome some of the shortcomings of the GOFAI approach.

(*) The resurgence of AI in the 1990s was also prompted especially by the rediscovery of the backpropagation algorithm, the UAT for NNS, a proliferation of data (due in part to the widespread use of the internet), and new develops processing techniques.

1990s-2000s: Resurgence of AI

(*) These new techniques (NNs and GAs) boasted a more *organic* performance on the whole.

(*) NNs for instance exhibited a useful, “graceful degradation” property and they could learn from experience – thereby finding natural ways to generalize from examples by discovering hidden statistical patterns in their input.

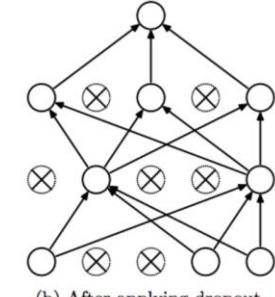
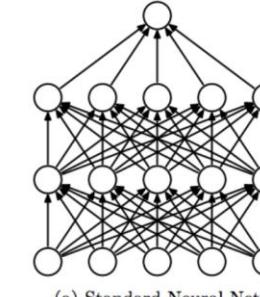
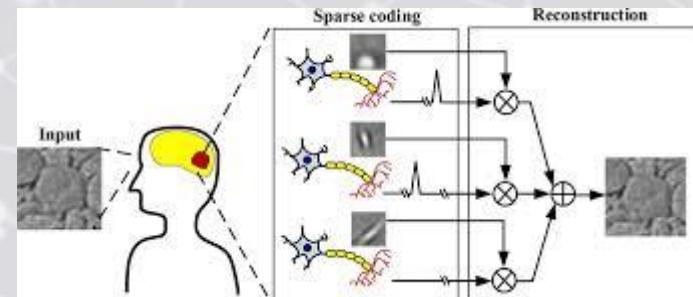
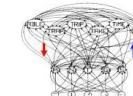
(*) NNs were also seen as more biologically-plausible models (cf. GOFAI); the brain-like qualities of NNs contrasted favorably with the rigid and brittle performance of traditional, rule-based GOFAI systems.

(*) The philosophy of **connectionism** gained traction in cognitive science / neuroscience, which further supported modeling with NNs. Connectionism emphasizes the importance of massively parallel sub-symbolic processing .

1. Distributed Representation (vs. localist representation)

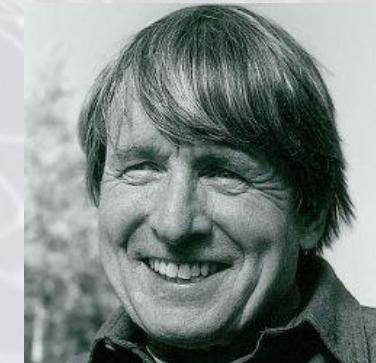
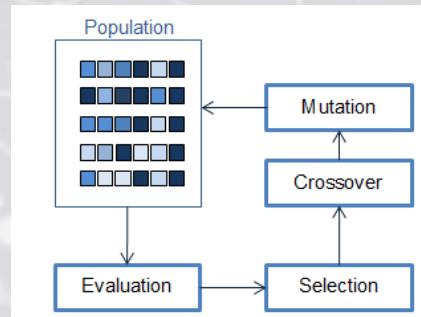
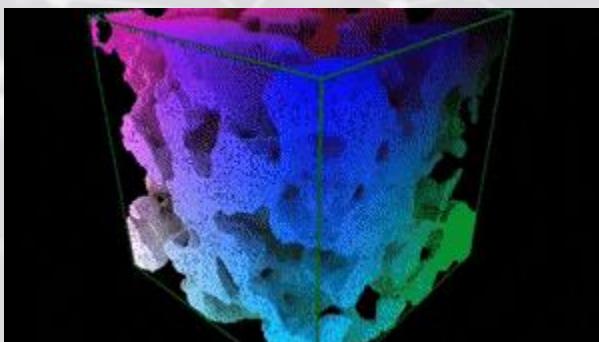
An object is represented by multiple units; the same unit participates in multiple representations:

$$\begin{array}{c} A \\ \bullet \quad \circ \quad \bullet \quad \bullet \end{array}$$
$$\begin{array}{c} Z \\ \bullet \quad \circ \quad \bullet \quad \circ \end{array}$$



1990s-2000s: Resurgence of AI

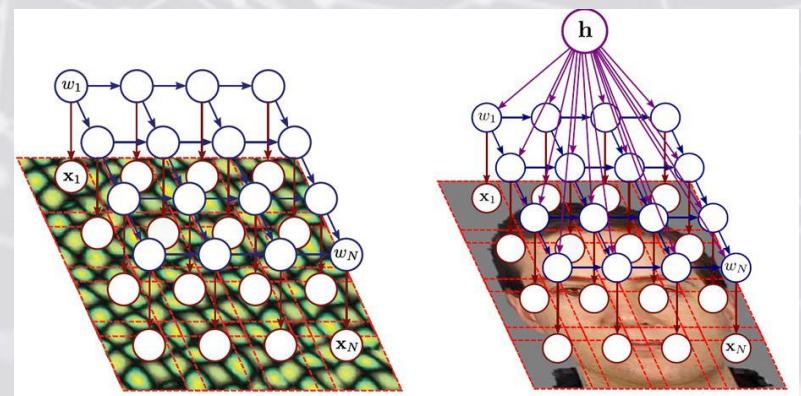
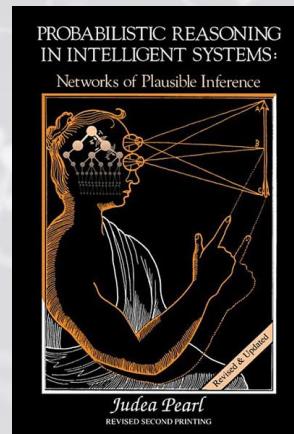
- (*) In evolution-based models (e.g. GAs), a population of candidate solutions (which can be data structures or programs) is maintained, and new candidate solutions are generated randomly by mutating or recombining variants in the existing population.
- (*) Periodically, the population is pruned by applying a selection criterion (a *fitness function*) that allows only the better candidates to survive into the next generation and thereby combine their “genetic material” with other “fit” subjects.
- (*) When it works, this kind of algorithm can produce efficient solutions to a very wide range of problems – solutions that may be strikingly novel and unintuitive, often looking more like natural structures than anything that a human engineer would design. Furthermore, this can happen without much need for direct human input (beyond the specification of the problem/learning parameters).



Holland

1990s-2000s: Resurgence of AI

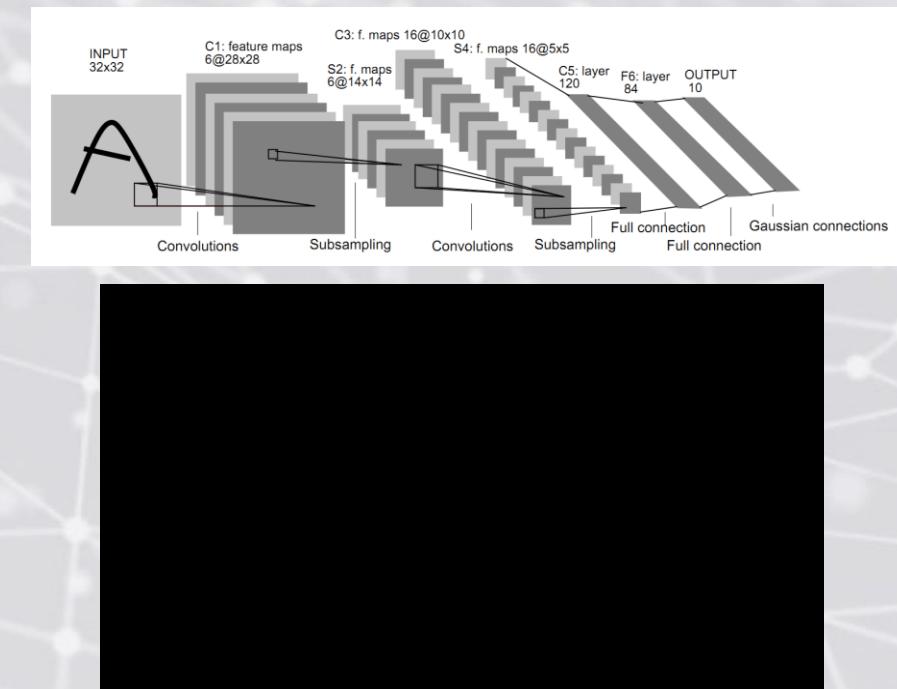
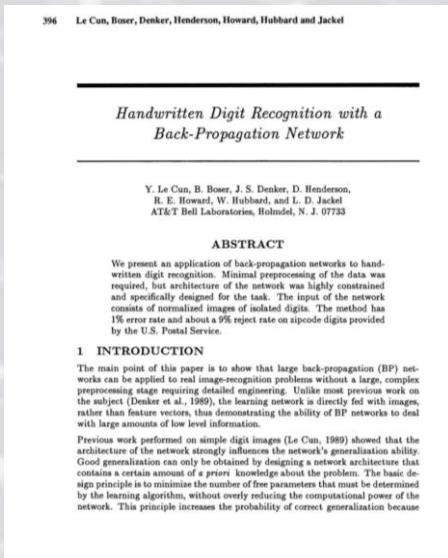
- (*) Behind many of the new state-of-the-art techniques in AI lies a set of mathematically well-specified tradeoffs. The ideal one is that of a **Bayesian agent**, one that makes probabilistically optimal use of available information.
- (*) This ideal is, however, unattainable because it is too computationally demanding to be implemented in any physical computer. Accordingly, *one can view AI as a quest to find shortcuts: ways of tractably approximating the Bayesian ideal by sacrificing some optimality while preserving enough to get high performance.*
- (*) In 1988 Judea Pearl published a highly influential book on **probabilistic graphical models** (e.g. *Bayesian nets*). Bayesian networks provide a concise way of representing probabilistic and conditional independence relations that hold in some particular domain. Exploiting these independence relations is essential to overcoming the combinatorial explosion (graphical models have led to improvements in Monte Carlo approximation techniques, deep learning, and causal models, among other domains).



1990s-2000s: Resurgence of AI

(*) An important benchmark in the history of AI that helped usher in the recent, “deep learning” phenomenon, was **Yann LeCun** (Facebook) et al.’s seminal 1990 paper, for which researchers applied a convolutional neural network (CNN) to perform handwritten digit recognition.

(*) Remarkably, this technique, which has been extraordinarily influential for subsequent deep learning and computer vision approaches, achieved a **1% error rate** (on par with human error). This approach was later adopted by the US post office to automate mail sorting (using handwritten zip codes).



1990s-2000s: Resurgence of AI

(*) In 1997 **Deep Blue** (IBM), a chess-playing computer defeated the world champion, Garry Kasparov by a final score of 4-2. (Note that Kasparov accused IBM of cheating and demanded a rematch, IBM refused).

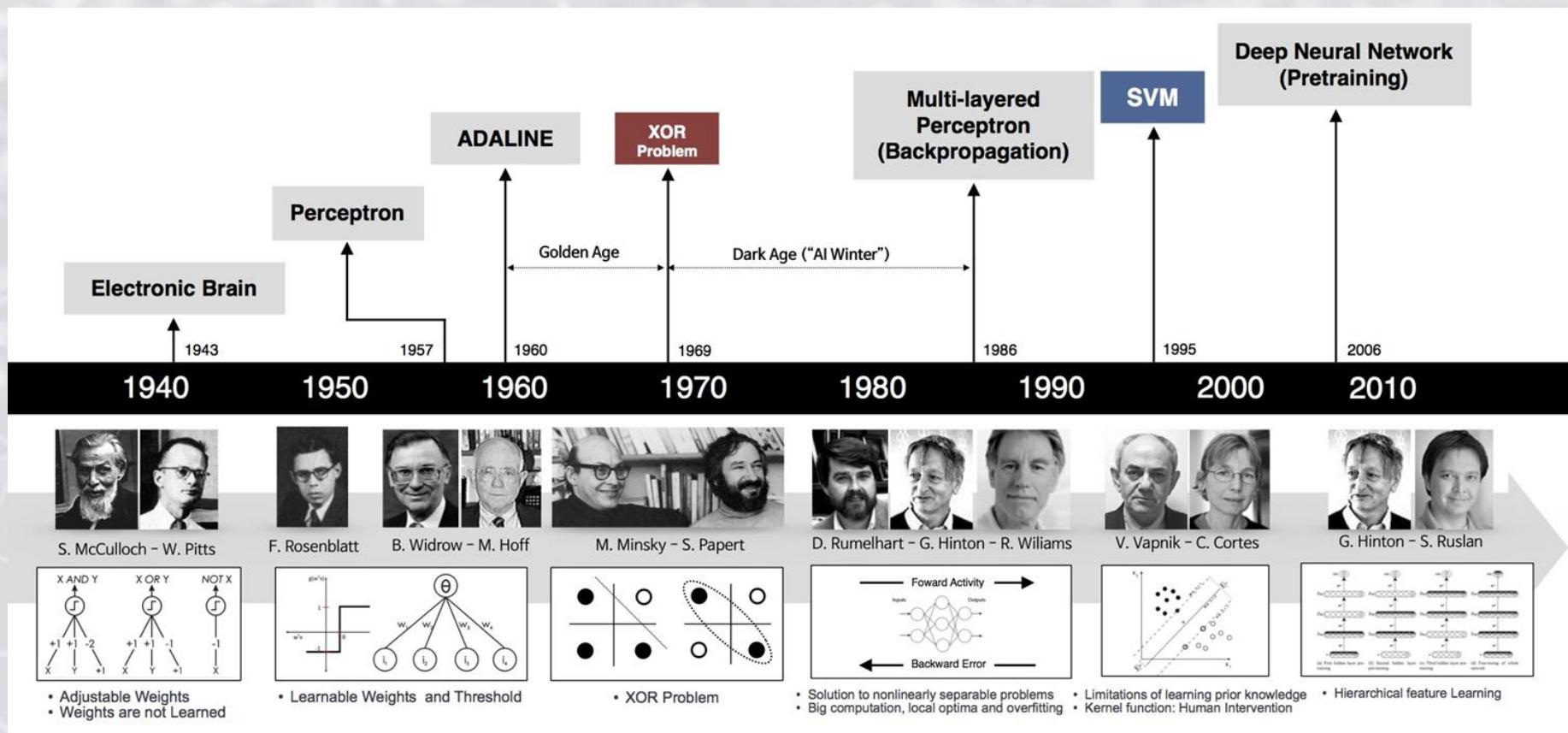
(*) Deep Blue employed custom VLSI chips to execute alpha-beta pruning search in parallel; notably Deep Blue is an example of GOFAI rather than deep learning – which is to say it used a relatively naïve (by contemporary standards), brute force approach.

(*) These issues aside, Deep Blue's victory had a significant impact on capturing the imaginations of both the public (and industry) regarding AI and its future potential.

(*) Notably, it was once thought (in the early days of AI) that the invention of master-level chess AI systems would presuppose the invention of AGI (artificial general intelligence) – clearly this is not the case, as we still await the invention of AGI (John McCarthy once lamented: “As soon as it works, no one calls it AI anymore.”)

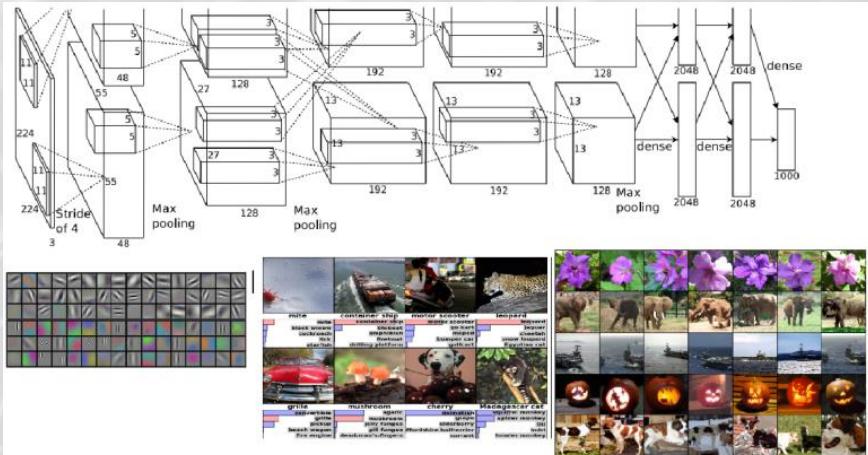


2010s and Beyond



2010s and Beyond

(*) **AlexNet** (2012), Alex Krizhevsky et al.: A new benchmark for deep learning, achieved top-5 error rate of 15% on ImageNet challenge (23k categories).



Krizhevsky



Goodfellow

(*) **Adversarial Learning and GANs** (2014),
Goodfellow et al.

The diagram shows a comparison between a "panda" image and a "gibbon" image. On the left is a photograph of a panda. Next to it is the equation $x + .007 \times \text{sign}(\nabla_x J(\theta, x, y))$, where x is labeled "panda" with 57.7% confidence. To the right is a photograph of a gibbon. Below the equations are the labels "nematode" and "gibbon" with their respective confidence levels: 8.2% and 99.3%.

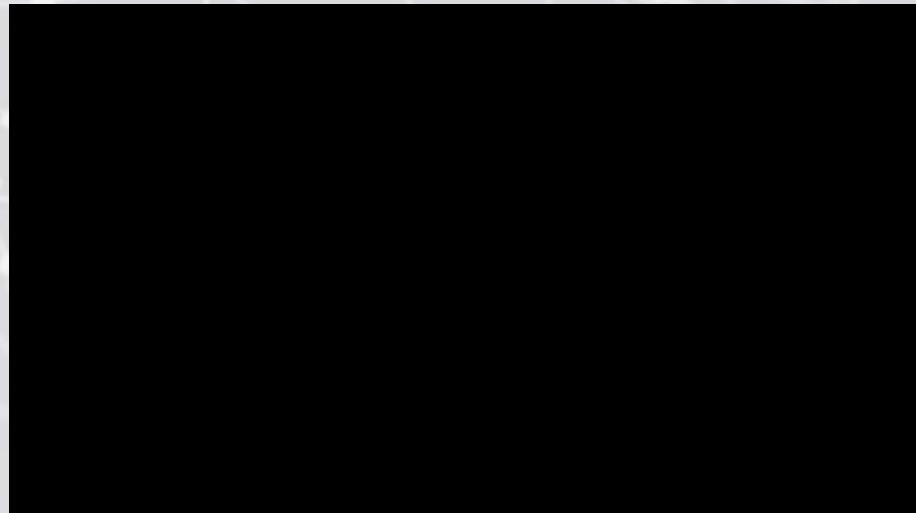


2010s and Beyond

(*) **IBM Watson** (2011)

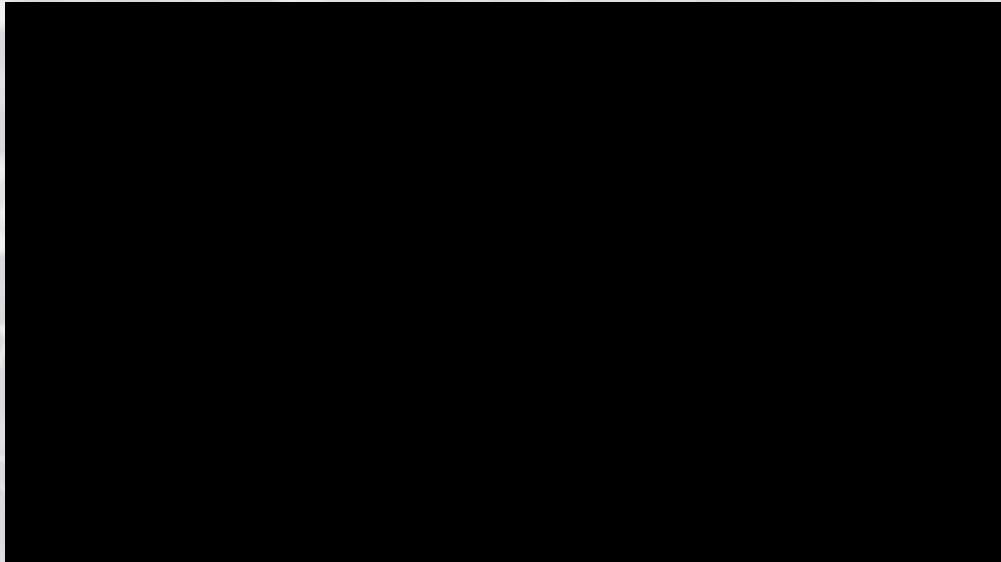


(*) **Google DeepDream** (2015)



2010s and Beyond

(*) DeepMind (2014): Playing Atari (at super-human levels) using Deep Reinforcement Learning



Hassabis

(*) DeepMind AlphaGo (2015)



AI: The Future

Final Considerations:

(*) Is human-level AGI (artificial general intelligence) possible?

(*) Is deep learning the answer? (or is it still a “microworld”?)

(*) Supervised vs. unsupervised learning: it turns out that more/”better” data might trump effectiveness of an algorithm! Need to consider the “knowledge bottleneck” – automate the learning process, bootstrap new patterns.

(*) “The first ultra-intelligent machine is the last invention that man need ever make”

10% chance: 2030
50% chance: 2050
90% chance: 2100

Polling results for researchers in AI when asked about arrival date of HMLI (human-level machine intelligence) and its potential impact

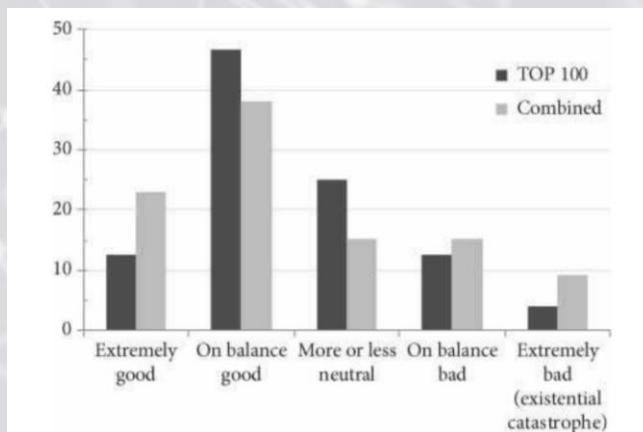
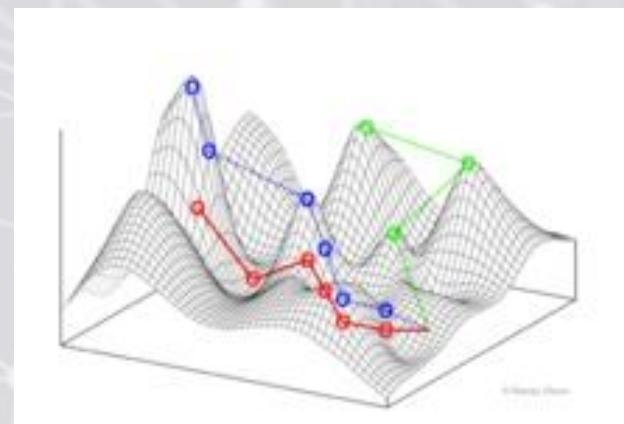
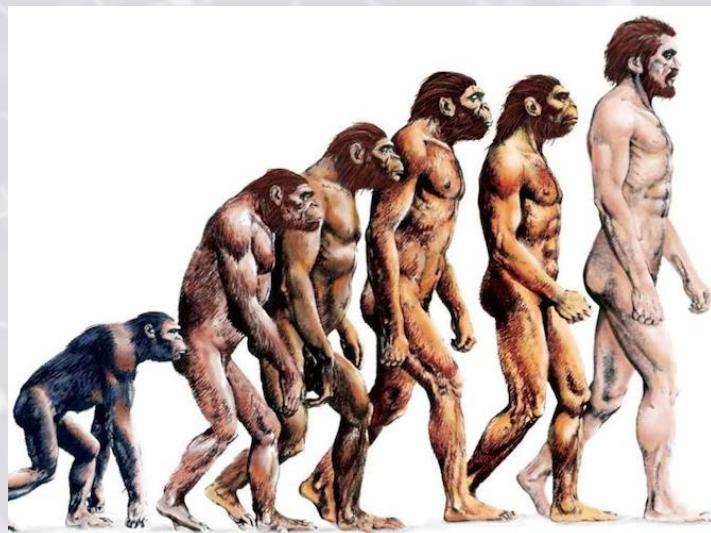


Figure 2 Overall long-term impact of HMLI.⁸³

Can we use evolution to re-discover intelligence?

(*) We know that blind evolutionary processes can produce human-level general intelligence, since they have already done so at least once!

(*) Q: It stands to reason that evolutionary processes **with foresight** – that is, genetic programs designed and guided by an intelligent human programmer – can achieve a similar outcome with far greater efficiency.



Can we use evolution to re-discover intelligence?

A back-of-the-envelope approximation for the complexity of “inventing” intelligence with foresight:

- (*) One can argue that the key insights for AI are embodied in the structure of the **nervous system**, which came into existence less than a billion years ago.
- (*) Evolutionary algorithms require not only variations to select among but also a fitness function to evaluate variants, and this is typically the most computationally expensive component. A fitness function for the evolution of AI plausibly requires simulation of neural development, learning and cognition to evaluate fitness.
- (*) We can make a crude estimate of the number of neurons in biological organisms that we might need to simulate to mimic evolution’s fitness function.
- (*) If, for instance, we consider that the honeybee brain consists of $\sim 10^6$ neurons, a fruit fly, $\sim 10^5$, and ants $\sim 250k$, then erring on the side of conservative, there are approximately 10^{19} insects on Earth, there would be roughly a total of 10^{24} insect neurons.

Can we use evolution to re-discover intelligence?

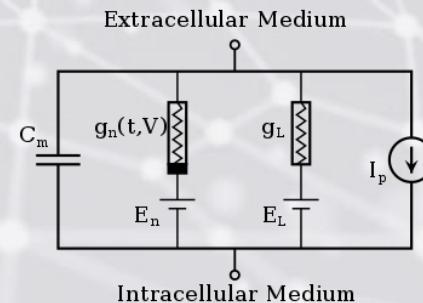
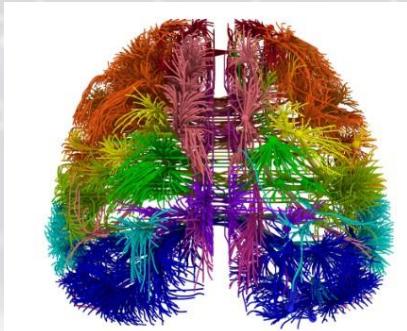
A back-of-the-envelope approximation for the complexity of “inventing” intelligence with foresight:

(*) If, for instance, we consider that the honeybee brain consists of $\sim 10^6$ neurons, a fruit fly, $\sim 10^5$, and ants $\sim 250k$, then erring on the side of conservative, there are approximately 10^{19} insects on Earth, there would be roughly a total of **10²⁴ insect neurons**.

(*) This figure can further be augmented an additional order of magnitude when we consider aquatic life, birds, reptiles, mammals, etc., to reach **10²⁵**.

(*) The computational cost of simulating one neuron depends on the level of detail that one includes in the simulation. Extremely simple neural models use $\sim 1k$ floating-point operations per second (FLOPS) to simulate one neuron in real-time.

(*) A more electrophysiologically realistic *Hodgkin-Huxley model* uses 1.2 million FLOPS.



Can we use evolution to re-discover intelligence?

A back-of-the-envelope approximation for the complexity of “inventing” intelligence with foresight:

(*) A more electrophysiologically realistic *Hodgkin-Huxley model* uses 1.2 million FLOPS.

(*) A more detailed, multi-compartmental model would add another three or four orders of magnitude (while higher-level models that abstract systems of neurons might subtract two or three orders of magnitude).

(*) If we were to simulate 10^{25} neurons over a billion years of evolution (longer than the existence of nervous systems as we know them), and we allow our computers to run for one year, these figures would give us a requirement in the range of **10^{31} - 10^{44} FLOPS**.

(*) Contemporary super computers provide only roughly 10^{18} FLOPS, which means this brute force search would on the surface require well over a trillion years to execute!

