

History of AI, Current Trends, Prospective Trajectories

Winter Academy on Artificial Intelligence and International Law

Asser Institute – 11 February 2019

Giovanni Sileno
g.sileno@uva.nl



What is Artificial Intelligence?

What is Artificial Intelligence?

What is Artificial Intelligence?

- What is made by humans?



What is Artificial ~~Intelligence~~?

- What is made by humans?
- What is induced by humans?



What is Artificial Intelligence?

- What is made by humans?
- What is induced by humans?
- What is simulated, not true?



"It's a sunset, Billy. It's not selling anything."

What is ~~Artificial~~ Intelligence?

What is ~~Artificial~~ Intelligence?

- Problem-solving ability?



What is ~~Artificial~~ Intelligence?

- Problem-solving ability?
- Capacity of abstraction?



What is ~~Artificial~~ Intelligence?

- Problem-solving ability?
- Capacity of abstraction?
- Capacity of organization?



What is ~~Artificial~~ Intelligence?

- Problem-solving ability?
- Capacity of abstraction?
- Capacity of organization?
- Creativity?



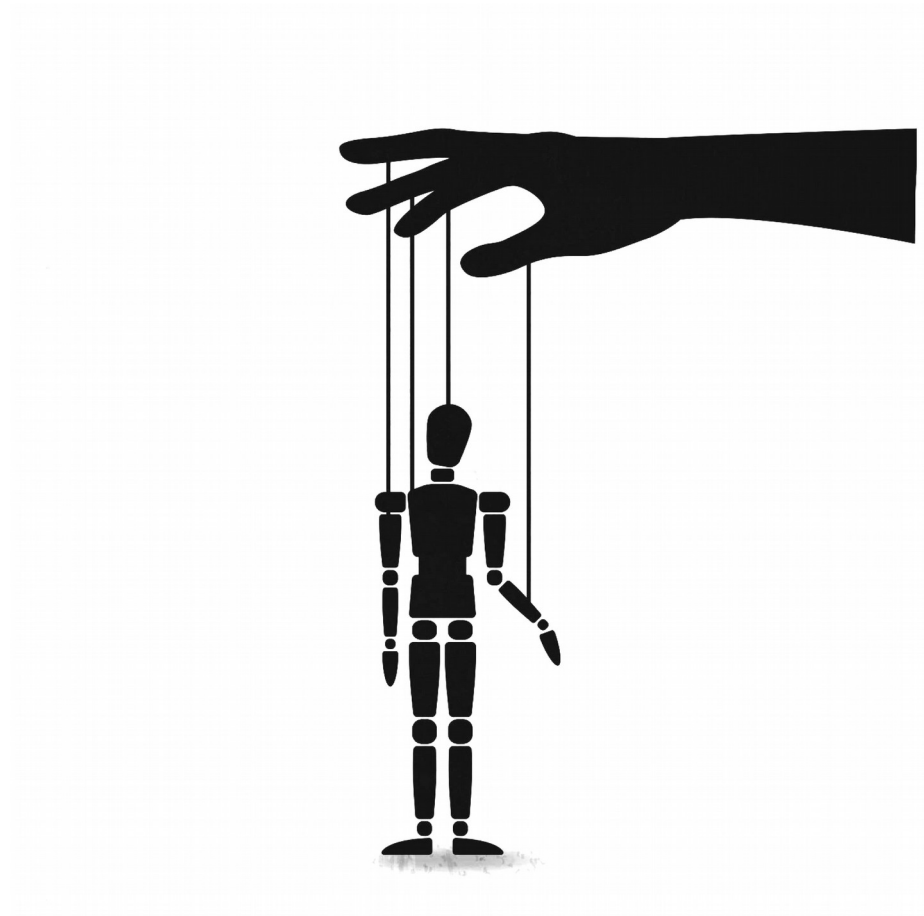
What is ~~Artificial~~ Intelligence?

- Problem-solving ability?
- Capacity of abstraction?
- Capacity of organization?
- Creativity?
- Self-awareness?



What is ~~Artificial~~ Intelligence?

- Problem-solving ability?
- Capacity of abstraction?
- Capacity of organization?
- Creativity?
- Self-awareness?
- Manipulation ability?



AI as a discipline

- Most disciplines emerge around specific domains of knowledge, settling upon methods deemed adequate to that domain.

Biology



Life and living organisms

Physics



Laws of the universe

Law



Legal systems and justice

Computer science



Computational systems

but Artificial Intelligence?

AI as a discipline

- As a discipline, AI is not primarily connected to a knowledge domain, but to a **purpose**:

conceiving artificial systems that are intelligent

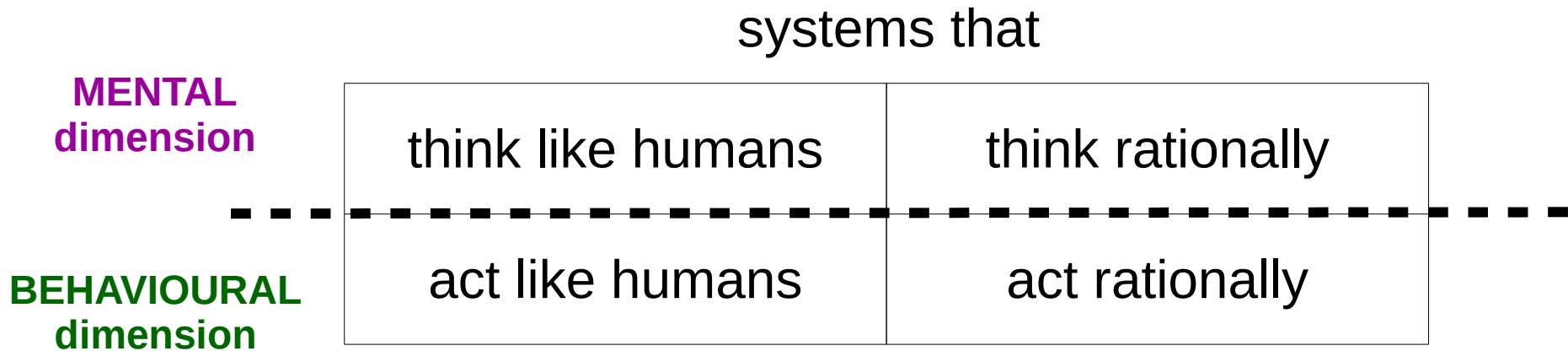
- All other disciplines (and their methods, or refinements of their methods) become for AI ***instrumental*** to that purpose (or sub-goals derived from that purpose).
- **But what is meant by this purpose?**

Categories of AIs

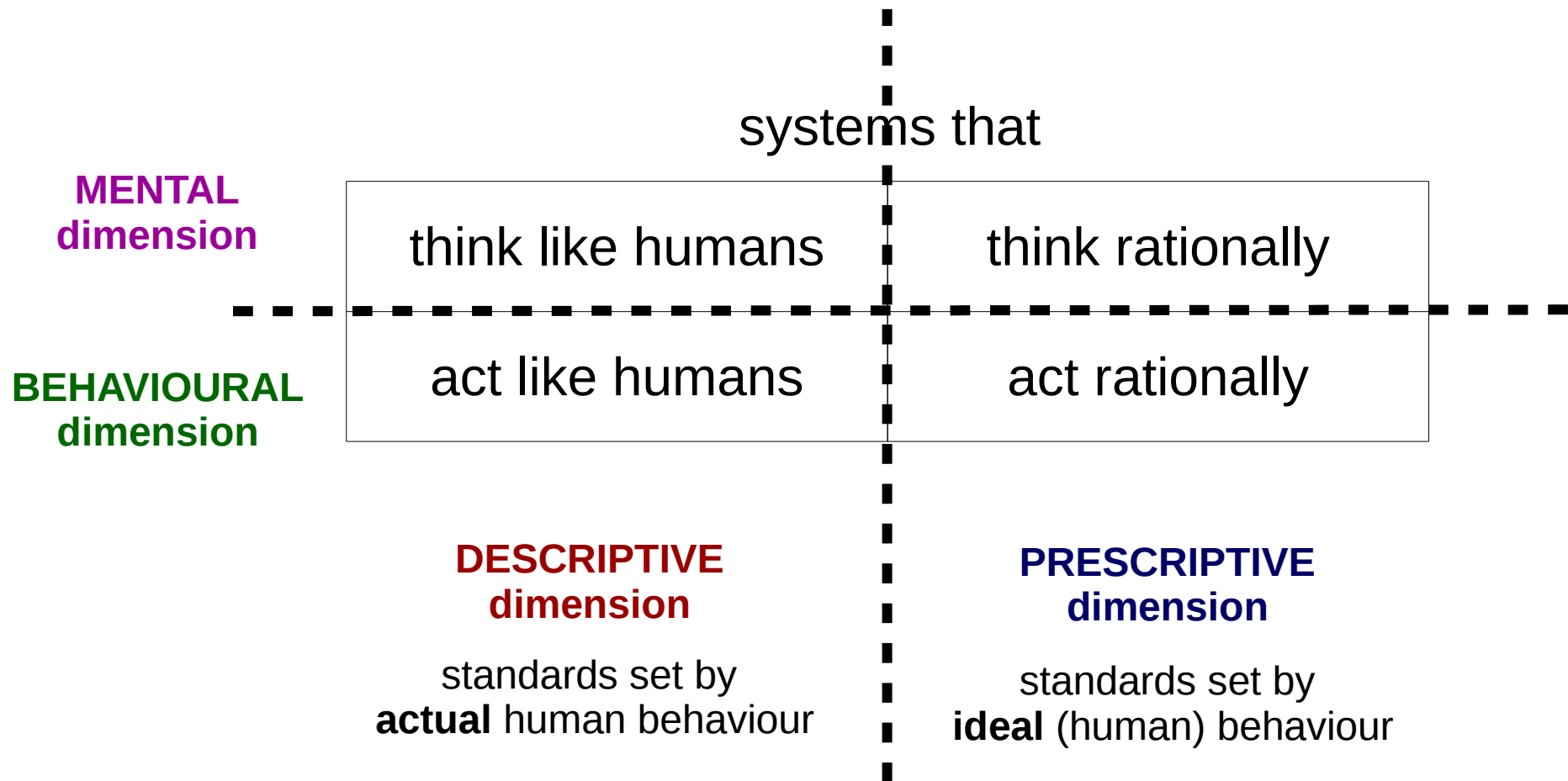
systems that

think like humans	think rationally
act like humans	act rationally

Categories of AIs



Categories of AIs

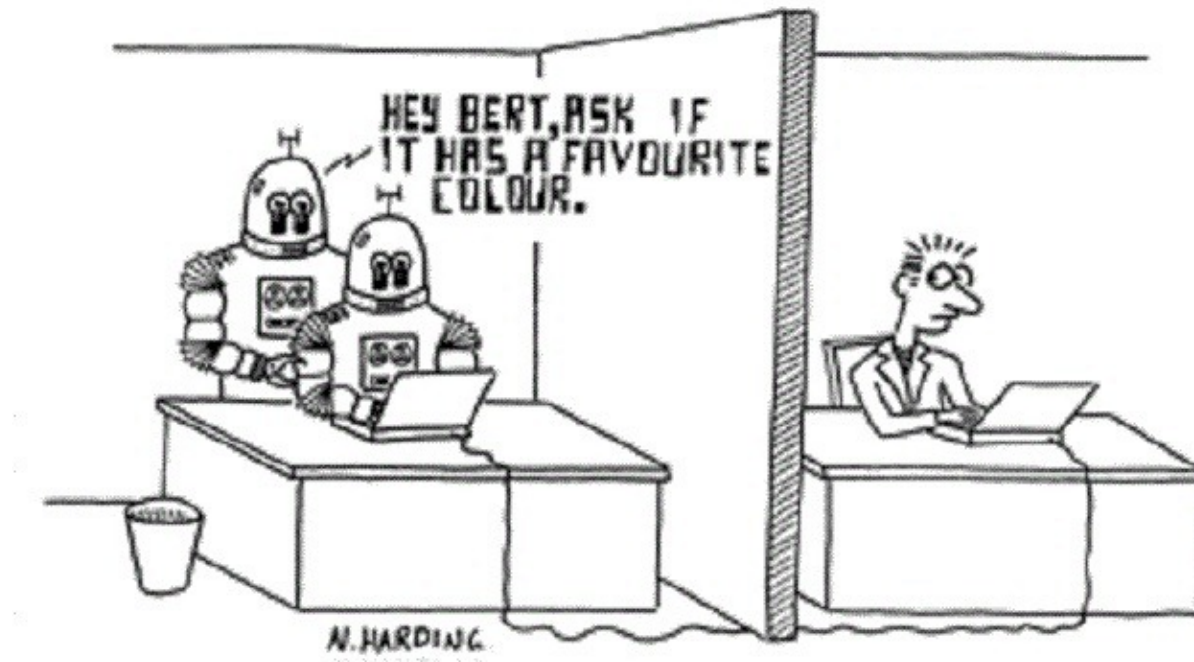


systems that

think like humans	think rationally
act like humans	act rationally

Turing test approach

artificial and natural not distinguishable behind a neutral interface



systems that

think like humans	think rationally
act like humans	act rationally

Cognitive modeling approach

AI reproducing cognitive functions observed by humans

NATURA ARTIS MAGISTRA argument

If these cognitive functions are required for our intelligence



they might be required to achieve artificial intelligence

EXPLAINABILITY argument

If they explain our internal working



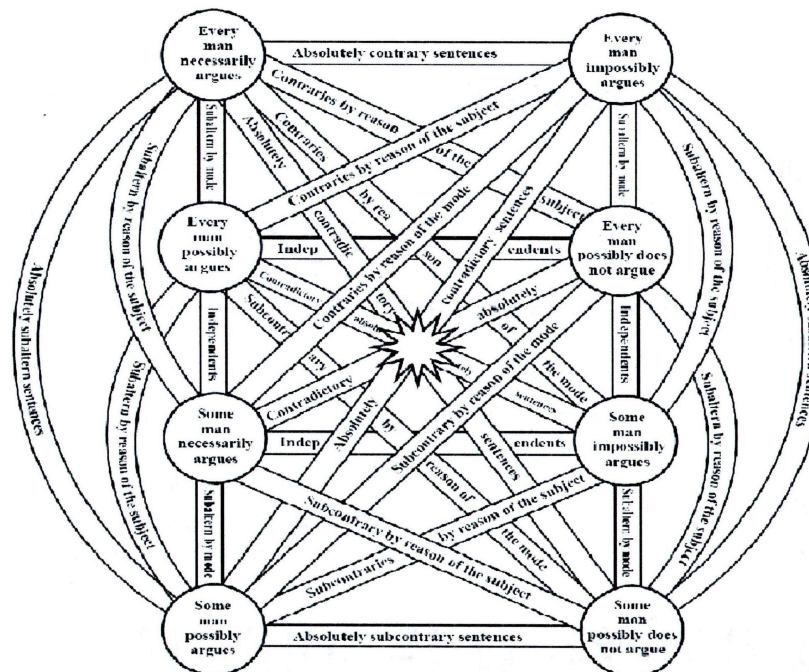
they can help to interpret AI functioning

systems that

think like humans	think rationally
act like humans	act rationally

The “Laws of Thought” approach

AI producing logically valid inferences

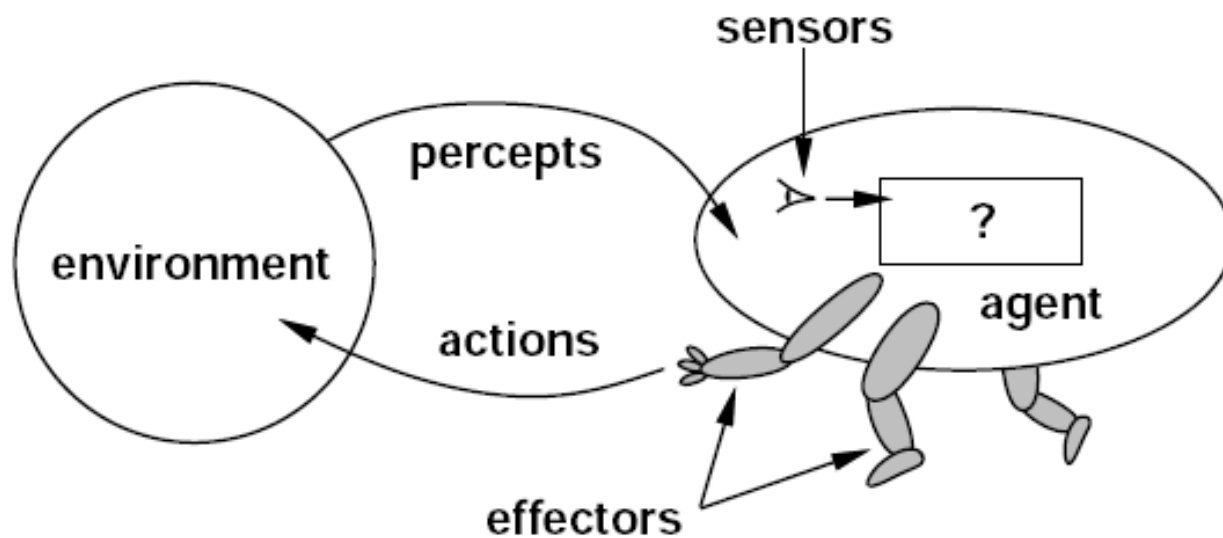


systems that

think like humans	think rationally
act like humans	act rationally

The “Rational Agent” approach

AI decision-making following standards of rationality



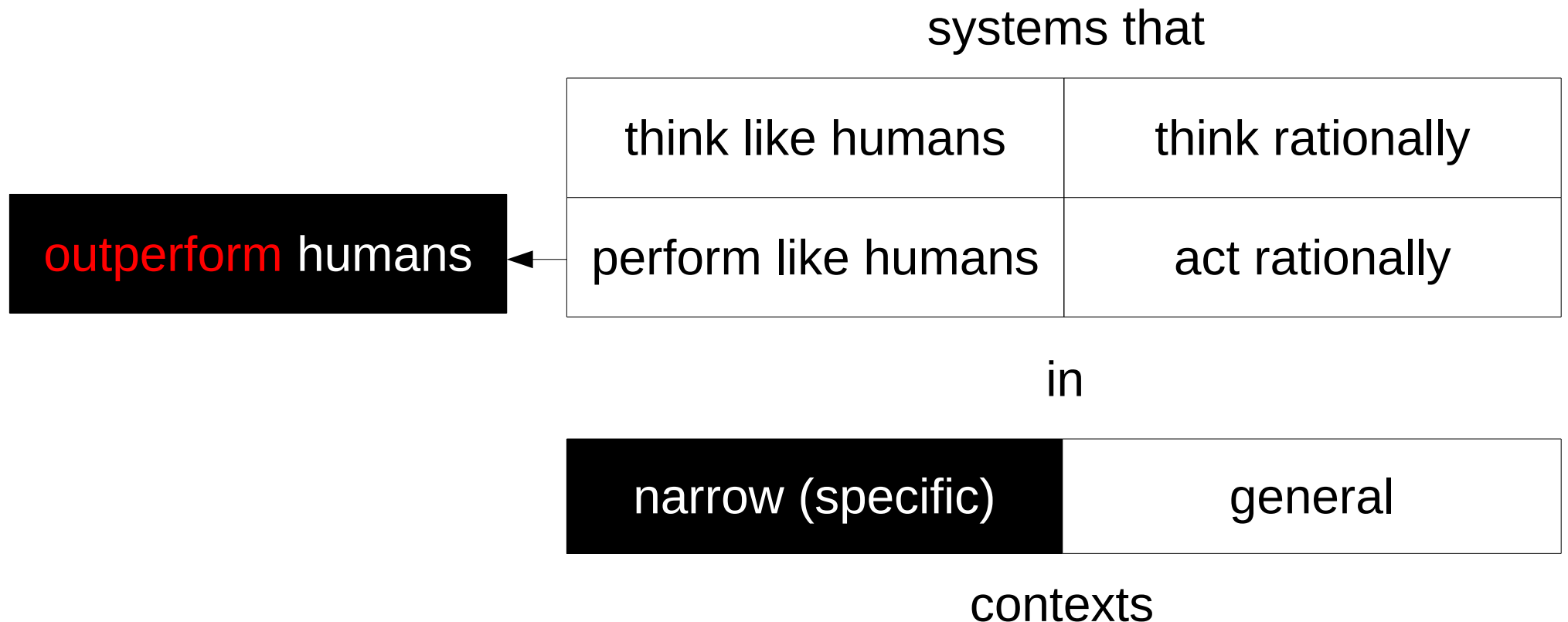
- the agent selects the best choice
- to achieve its **goals**
- given its **beliefs**



autonomous entity

Recent advances

- In specific tasks, performance can be easily measured (quantified).
→ *systems can adapt to perform better than humans.*



AI waves

- This variety of topics has been developed through a cycles of ***springs*** (and ***winters***) centered around different topics.
- Some of the peaks:
 - ad-hoc systems with handcrafted knowledge (**60s/70s**)
 - expert systems/problem solving methods (**80s**)
 - robotics, computer vision, speech recognition (**80s**)
 - evolutionary computing (**90s**)
 - agent-based modeling and multi-agent systems (**90s/00s**)
 - semantic web (**00s**)
 - deep learning (**10s**)

AI waves

- This variety of topics has been developed through a cycles of ***springs*** (and ***winters***) centered around different topics.
- Some of the peaks:
 - ad-hoc systems with handcrafted knowledge (**60s/70s**)
 - expert systems/problem solving methods (**80s**)
 - robotics, computer vision, speech recognition (**80s**)
 - evolutionary computing (**90s**)
 - agent-based modeling and multi-agent systems (**90s/00s**)
 - semantic web (**00s**)
 - deep learning (**10s**)
- Although each time the *mainstream topic* eclipsed all the others, the ***following advances*** were ***enabled only because fundamental research in the others somehow continued.***

AI waves

- This variety of topics has been developed through a cycles of ***springs*** (and ***winters***) centered around different topics.
- Some of the peaks:
 - ad-hoc systems with handcrafted knowledge (**60s/70s**)
 - expert systems/problem solving methods (**80s**)
 - robotics, computer vision, speech recognition (**80s**)
 - evolutionary computing (**90s**)
 - agent-based modeling and multi-agent systems (**90s/00s**)
 - semantic web (**00s**)
 - deep learning (**10s**)
- Although each time the *mainstream topic* eclipsed all the others, the ***following advances*** were ***enabled only because***

To present more in detail this phenomenon, we will now look into the start of AI, or its *first wave*.

The start of AI

The start of Artificial Intelligence

- **Artificial Intelligence** is a research field whose name was decided in a *workshop* at Dartmouth College in **1956**.
- A group of scientists gathered at the Dartmouth campus for a brainstorming long 6-8 weeks on the conception of “***Machines that Think***” and settled the foundations of at least three decades of research.

The start of Artificial Intelligence

- **Artificial Intelligence** is a research field whose name was decided in a *workshop* at Dartmouth College in **1956**.
- A group of scientists gathered at the Dartmouth campus for a brainstorming long 6-8 weeks on the conception of “***Machines that Think***” and settled the foundations of at least three decades of research.
- *But such an exploit rarely occurs by chance.*

Contextualization

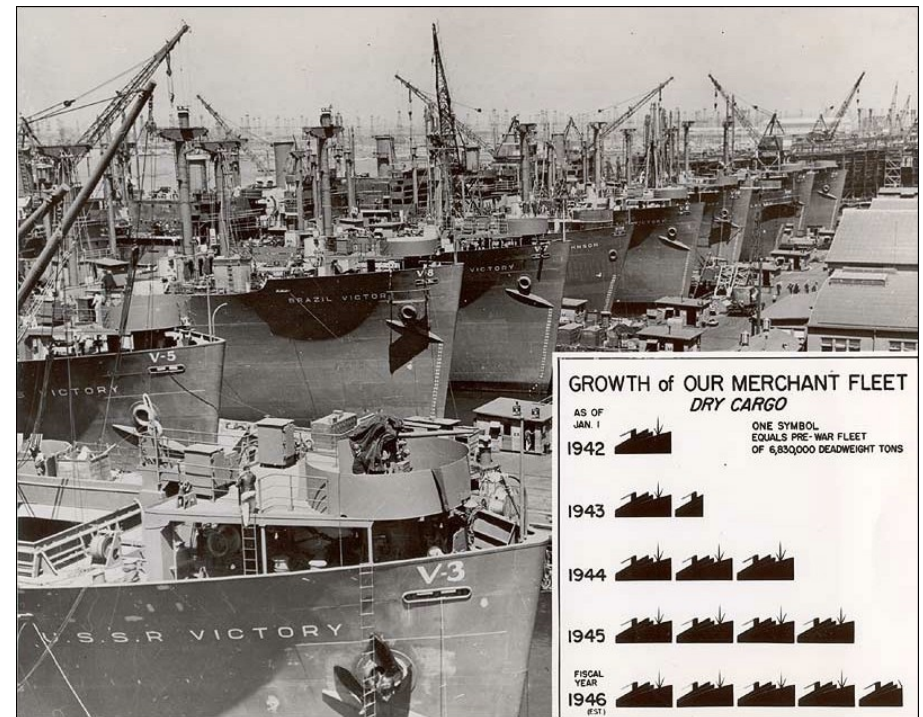
- Operational Research (since ~1930s)**

a sub-field of applied mathematics emerged in the years prior to World War II, when UK prepared to anticipate war.

it focuses on decision-making for operational settings:

- manufacturing
- transportation
- supply chain
- routing
- scheduling
- ...

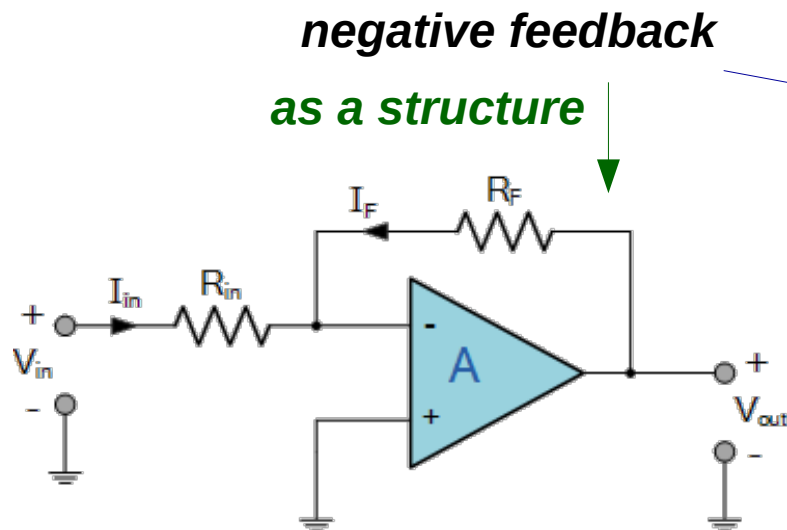
$$\begin{aligned}
 &\text{Maximise } Z = C_1x_1 + C_2x_2 + C_3x_3 + \dots + C_nx_n \\
 &\text{Subject to the conditions} \\
 &a_{11}x_1 + a_{12}x_2 + a_{13}x_3 + \dots + a_{1n}x_n \leq b_1 \\
 &a_{21}x_1 + a_{22}x_2 + a_{23}x_3 + \dots + a_{2n}x_n \leq b_2 \\
 &\vdots \\
 &a_{m1}x_1 + a_{m2}x_2 + a_{m3}x_3 + \dots + a_{mn}x_n \leq b_m \\
 &b_i \geq 0, i=1, 2, 3, \dots, m \\
 &x_j \geq 0, j=1, 2, 3, \dots, n
 \end{aligned}$$



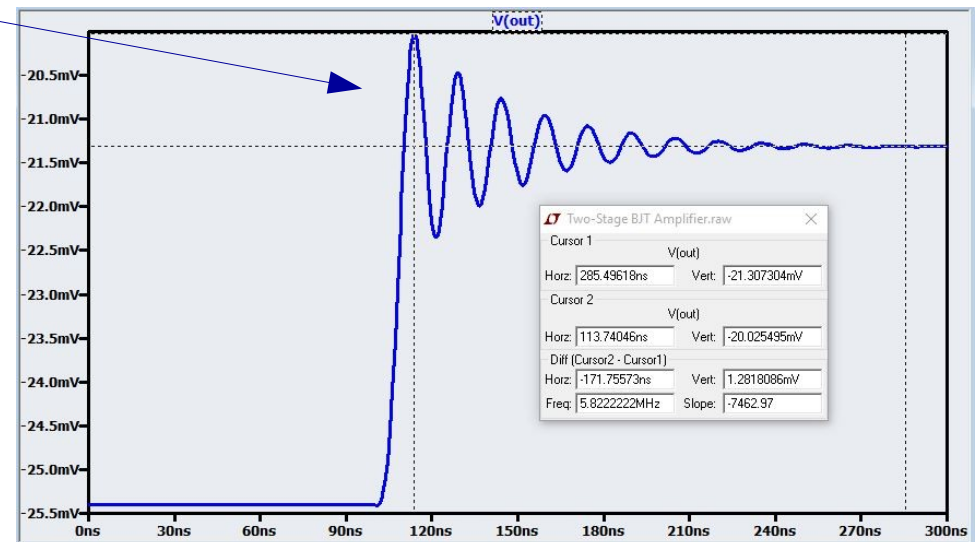
Contextualization

- Cybernetics (~1940s)**

emerged as a transdisciplinary approach to investigate systems of regulation, in fields as diverse as electronics, mechanics, biology and neurosciences. It considers systems holistically and study their internal control structures, constraints and possibilities.



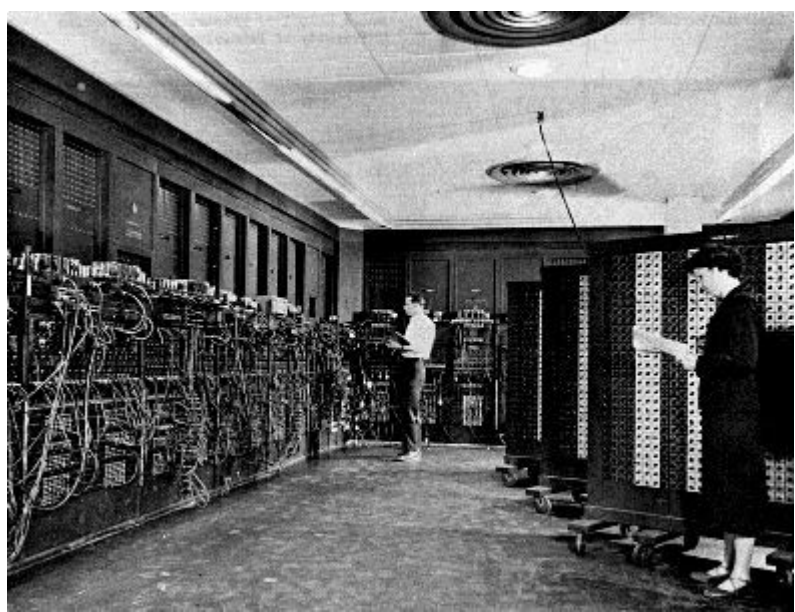
as a process



Contextualization

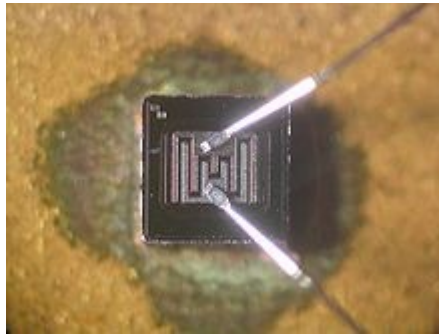
- **Technological advances in Electronics (~1950)**

First generation computers
(vacuum tubes-based)



ENIAC: 30 tons, area of about 1,800 square feet.

Second generation computers
(transistor-based)

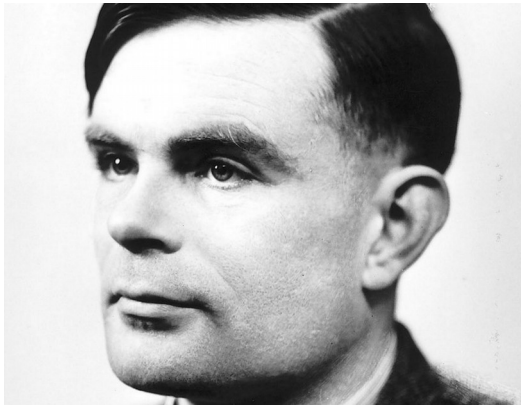


Invention of bipolar
transistor (1947)



Contextualization

- **Theoretical results about Computation and Information**



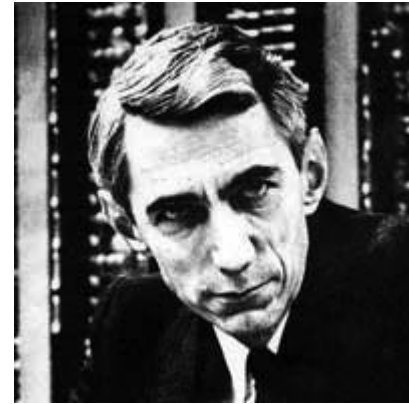
Alan Turing

formal model of computation (1937)

- enabling to write logically all computing processes (*Universal Turing Machine*)

the “Imitation Game” (1950)

- defining an operational standard for intelligence (*Turing Test*)



Claude E. Shannon

Information theory (1948)

- enabling to quantify information (for communication purposes), and so to perform data compression and to identify the limits of signal processing

Contextualization

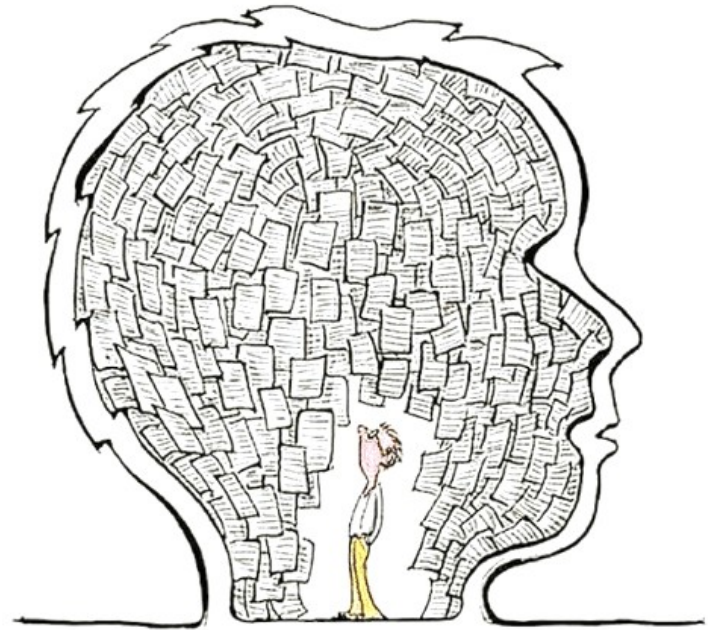
- **Psychology**



behaviorism

B. F. Skinner, "The Behavior of Organisms: An Experimental Analysis" (1938)

- Removal of mental element, Focus on operant conditioning (reward, punishments)



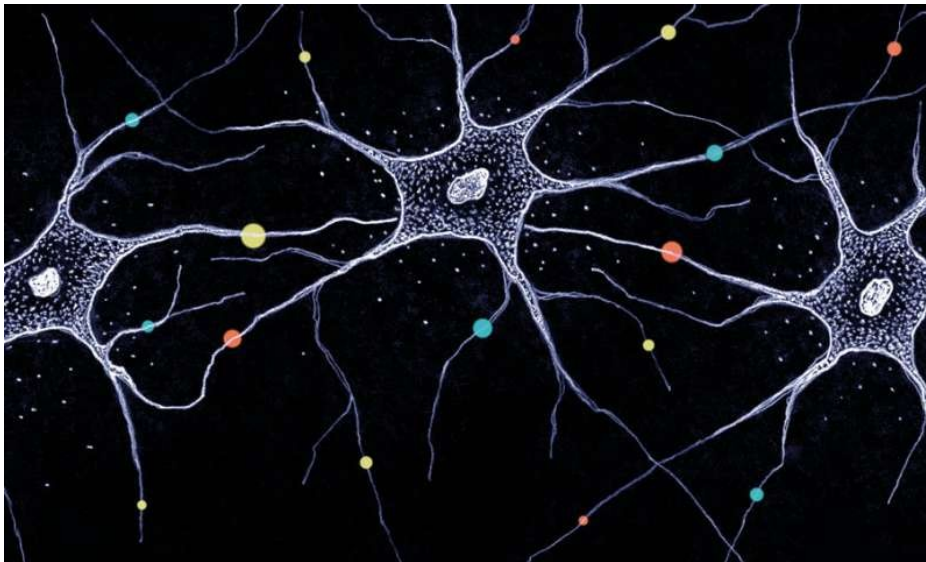
cognitive psychology

K. Craik. "The Nature of Explanation" (1943)

- Recovery of mental element, *folk-psychology*, compatible with *information-processing* view of cognition

Contextualization

- **Psychology (Neural Networks)**



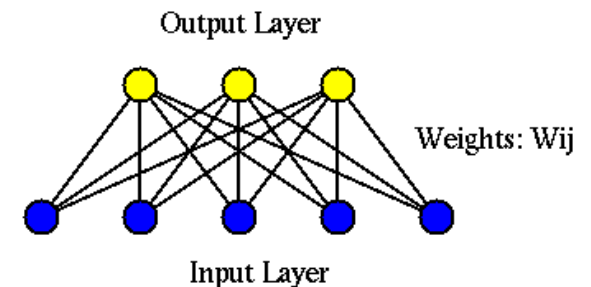
Simultaneous activation of neurons leads to increases in synaptic strength between them.

Donald Hebb (1949)



Thoughts and body activity result from interactions among neurons within the brain.

Alexander Bain (1873), William James (1890).



Presentation of first computational machines simulating neural networks
Farley and Clark (1954), Rochester, Holland, Habit, and Duda (1956).

Who was at the Darmouth Workshop (1956)?

- A remarkable group of ~20 scientist and engineers, including:
 - **John McCarty** (LISP language, situation calculus, non-monotonic logics)
 - **Marvin Minsky** (frames, perceptron, society of minds)
 - **Herbert Simon** (logic theorist, general problem solver, bounded rationality)
 - **Allen Newell** (logic theorist, general problem solver, the knowledge level)
 - **Ray Solomonoff** (father of algorithmic probability, algorithmic information theory)
 - **Arthur Lee Samuel** (first machine learning algorithm for checkers)
 - **W. Ross Ashby** (pioneer in cybernetics, law of requisite variety)
 - **Claude Shannon** (father of information theory)
 - **John Nash** (father of game theory)

future nobel prizes

Who was at the Darmouth Workshop (1956)?

- A remarkable group of ~20 scientist and engineers, including:
 - **John McCarty** (LISP language, situation calculus, non-monotonic logics)
 - **Marvin Minsky** (frames, perceptron, society of minds)
 - **Herbert Simon** (logic theorist, general problem solver, bounded rationality)
 - **Allen Newell** (logic theorist, general problem solver, the knowledge level)
 - **Ray Solomonoff** (father of algorithmic probability, algorithmic information theory)
 - **Arthur Lee Samuel** (first machine learning algorithm for checkers)
 - **W. Ross Ashby** (pioneer in cybernetics, law of requisite variety)
 - **Claude Shannon** (father of information theory)
 - **John Nash** (father of game theory)

The workshop brought no tangible result, but resulted in a shift from semantic approaches to ***symbolic processing***.

Who was at the Darmouth Workshop (1956)?

- A remarkable group of ~20 scientist and engineers, including:
 - **John McCarty** (LISP language, situation calculus, non-monotonic logics)
 - **Marvin Minsky** (frames, perceptron, society of minds)
 - **Herbert Simon** (logic theorist, general problem solver, bounded rationality)
 - **Allen Newell** (logic theorist, general problem solver, the knowledge level)
 - **Ray Solomonoff** (father of algorithmic probability, algorithmic information theory)
 - **Arthur Lee Samuel** (first machine learning algorithm for checkers)
 - **W. Ross Ashby** (pioneer in cybernetics, law of requisite variety)
 - **Claude Shannon** (father of information theory)
 - **John Nash** (father of game theory)

a strong agenda

The workshop brought no tangible results, but resulted in a shift from semantic approaches to ***symbolic processing***.

logician ▲

reasoning and decision-making

AI AS ENGINEERING OF THE “MIND”

induction of functions from data

empiricist

monolithic
systems

logic

logician ▲

reasoning and decision-making

AI AS ENGINEERING OF THE “MIND”

▼
heterogeneous
systems

induction of functions from data

homogeneous
systems

**artificial neural
networks (ANNs)**

monolithic
systems

probability

empiricist

monolithical
systems

logician

“Neats”

elegant solutions, provably correct

reasoning and decision-making

“Scruffies”

ad-hoc solutions,
empirical evaluation

**AI AS ENGINEERING
OF THE “MIND”**

heterogeneous
systems

induction of functions from data

homogeneous
systems

*characteristics of most people
at the Darmouth workshop*

monolithical
systems

empiricist

monolithical
systems

logician ▲

“Neats”

elegant solutions, provably correct

reasoning and decision-making

“Scruffies”

ad-hoc solutions,
empirical evaluation

**AI AS ENGINEERING
OF THE “MIND”**

heterogeneous
systems

induction of functions from data

homogeneous
systems

*characteristics of most people
at the Dartmouth workshop*

There were few researchers working on neural networks, and more in general ***learning*** was not brought to the foreground.

What/who stayed in the background

- In the words of another remarkable researcher (who was invited but could not go):
 - **John Holland** (neural networks, pioneer of complex adaptive systems and genetic algorithms)

[It resulted that] “there was very little interest in learning. In my honest opinion, this held up AI in quite a few ways. It would have been much better if **Rosenblatt’s Perceptron** work, or in particular **Samuels’ checkers playing system**, or some of the other early machine learning work, had had more of an impact. In particular, I think there would have been less of this notion that **you can just put it all in as expertise**” [..]

“it’s still not absolutely clear to me why the other approaches fell away. Perhaps there was no forceful advocate.”

Ingredients for many stories of shining
success and dramatic fall in AI

Ingredients for many stories of shining success and dramatic fall in AI

- societal needs
- strong advocates
- initial unexpected successes
- adequate computational technologies

Ingredients for many stories of shining success and dramatic fall in AI

- societal needs
- strong advocates
- initial unexpected successes
- adequate computational technologies



- financial resources



Ingredients for many stories of shining success and dramatic fall in AI

- societal needs
- strong advocates
- initial unexpected successes
- adequate computational technologies

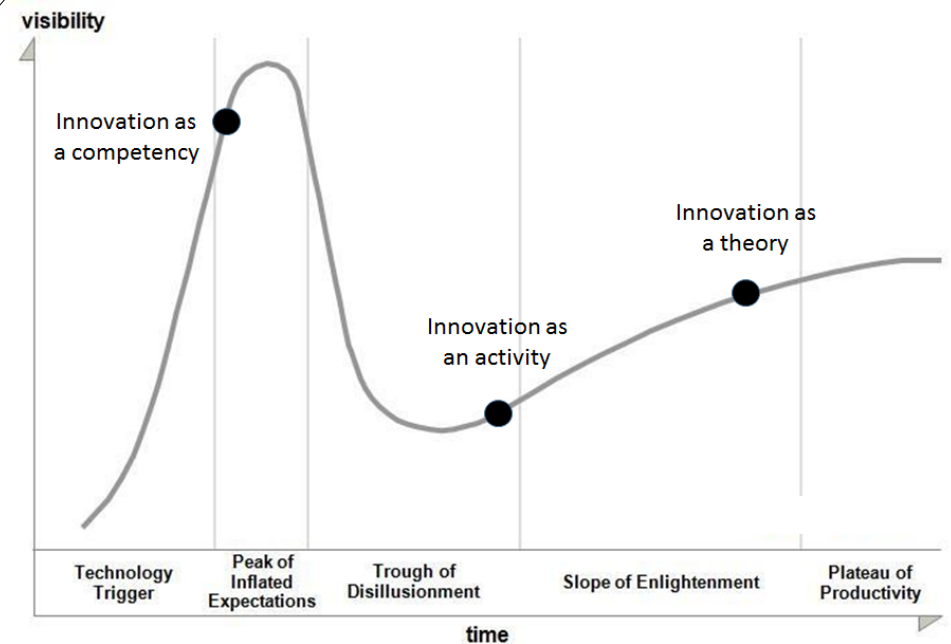


raising expectations
illusions
and then *delusions*



- financial resources

APPROVED



Ingredients for many stories of shining success and dramatic fall in AI

- societal needs
- strong advocates
- initial unexpected successes
- adequate computational technologies



raising expectations
illusions
and then *delusions*



- financial resources

APPROVED

visibility

Innovation as
a competency

Innovation as
a theory

Innovation as
an activity

but still (most of the times) there are concrete achievements.
They just become **infrastructure**: invisible, but necessary.

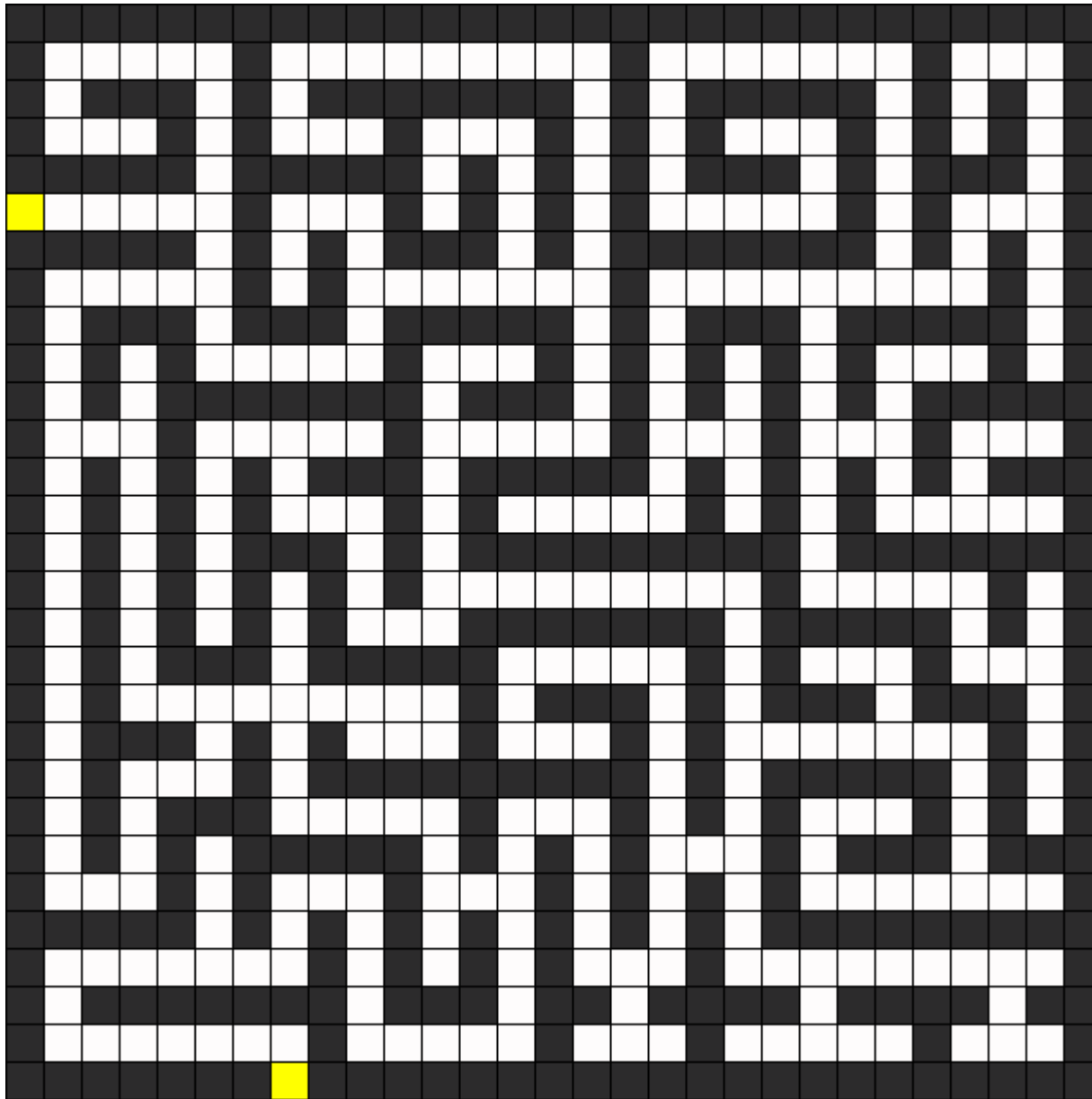
Computational intelligence

Algorithm = Logic + Control

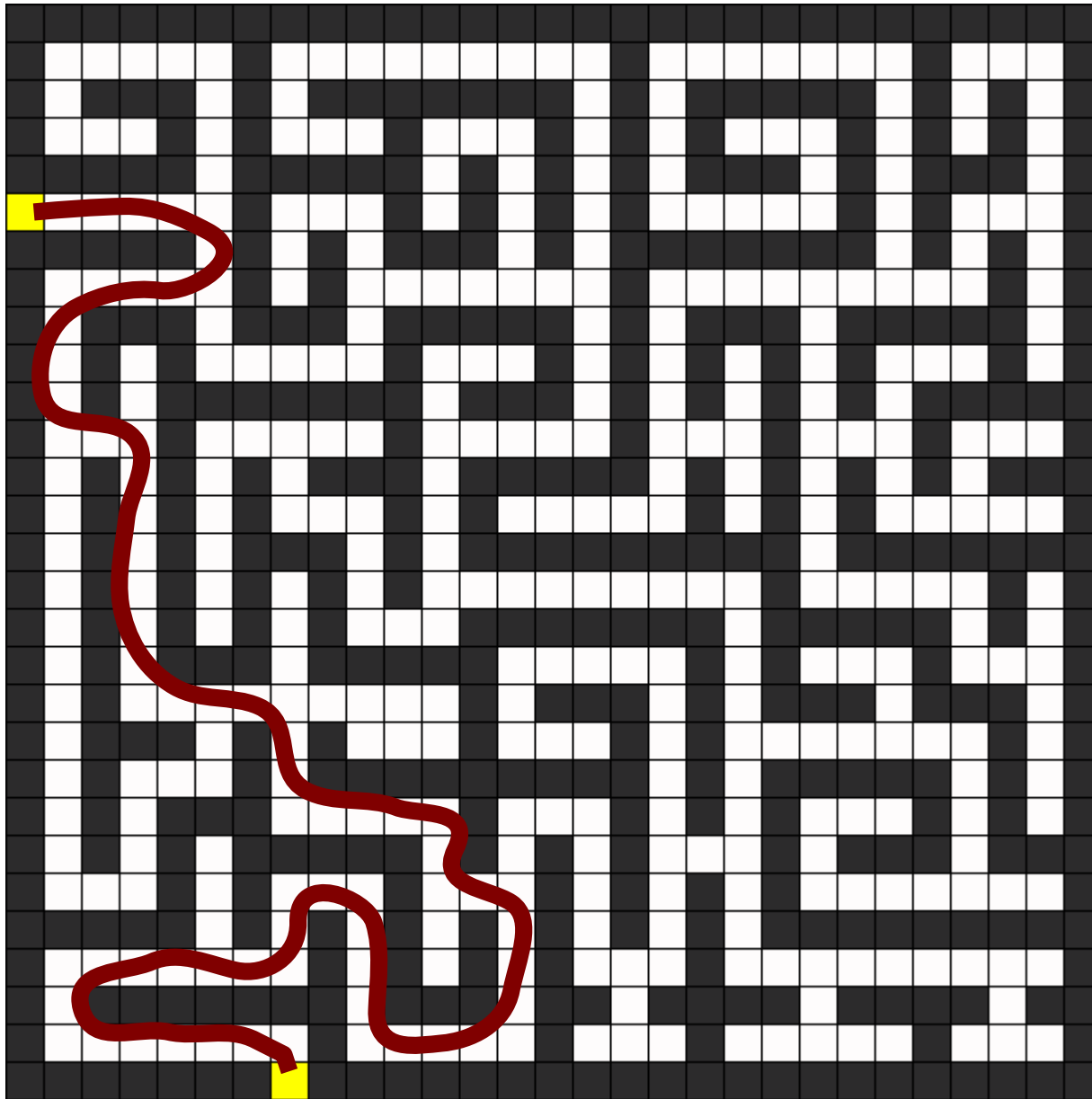
“An algorithm can be regarded as consisting of

- a **logic component**, which specifies the *knowledge* to be used in solving problems, and
- a **control component**, which determines the *problem-solving strategies* by means of which that knowledge is used.

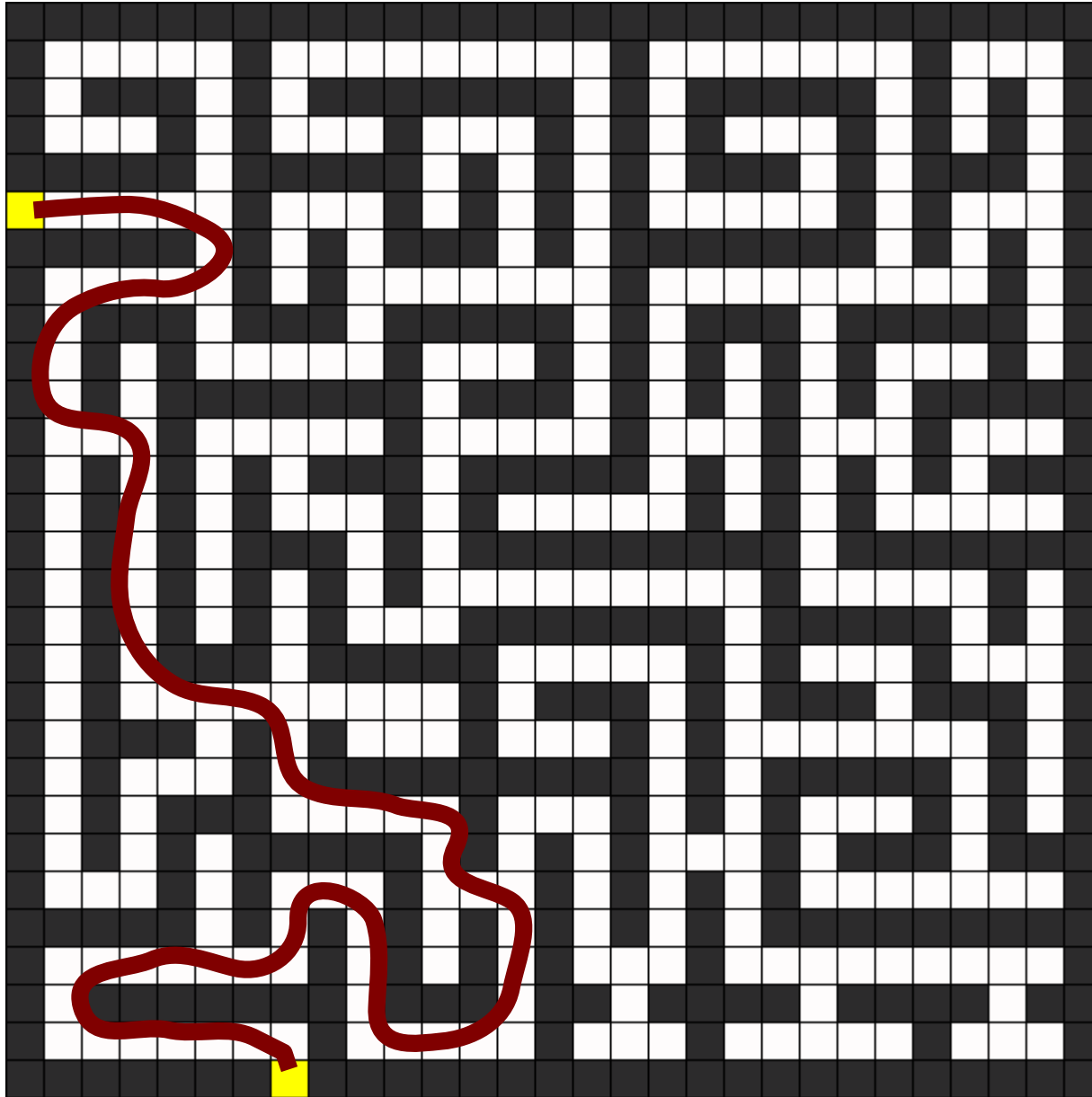
The logic component determines the meaning of the algorithm whereas the control component only affects its efficiency.”



**Imperative style of
programming:**
you command the
directions

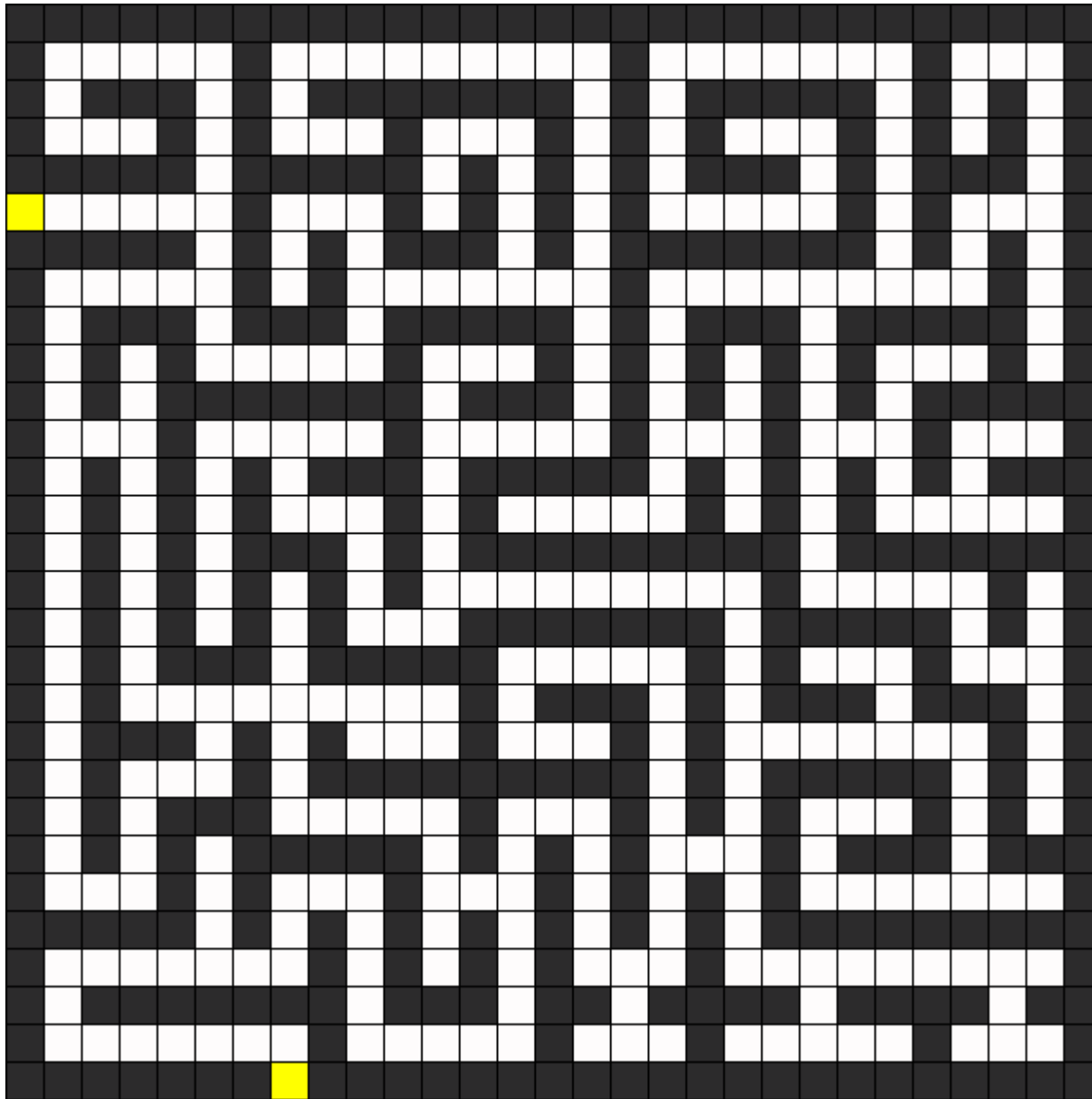


Imperative style of programming:
you command the directions



Imperative style of programming:
you command the directions

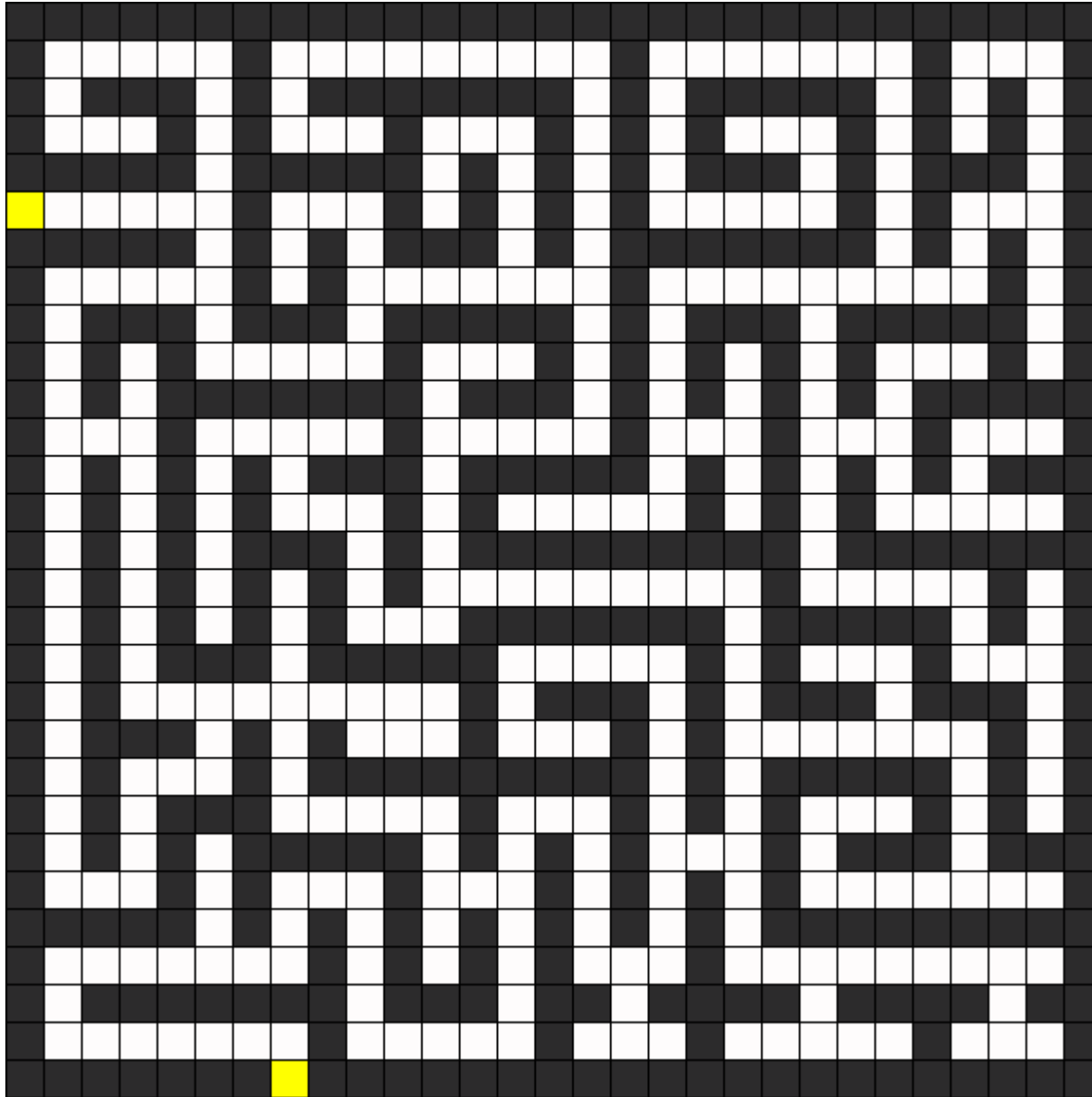
- What if the labyrinth changes?



Declarative style of programming:

you give just the labyrinth.

the computer finds the way.

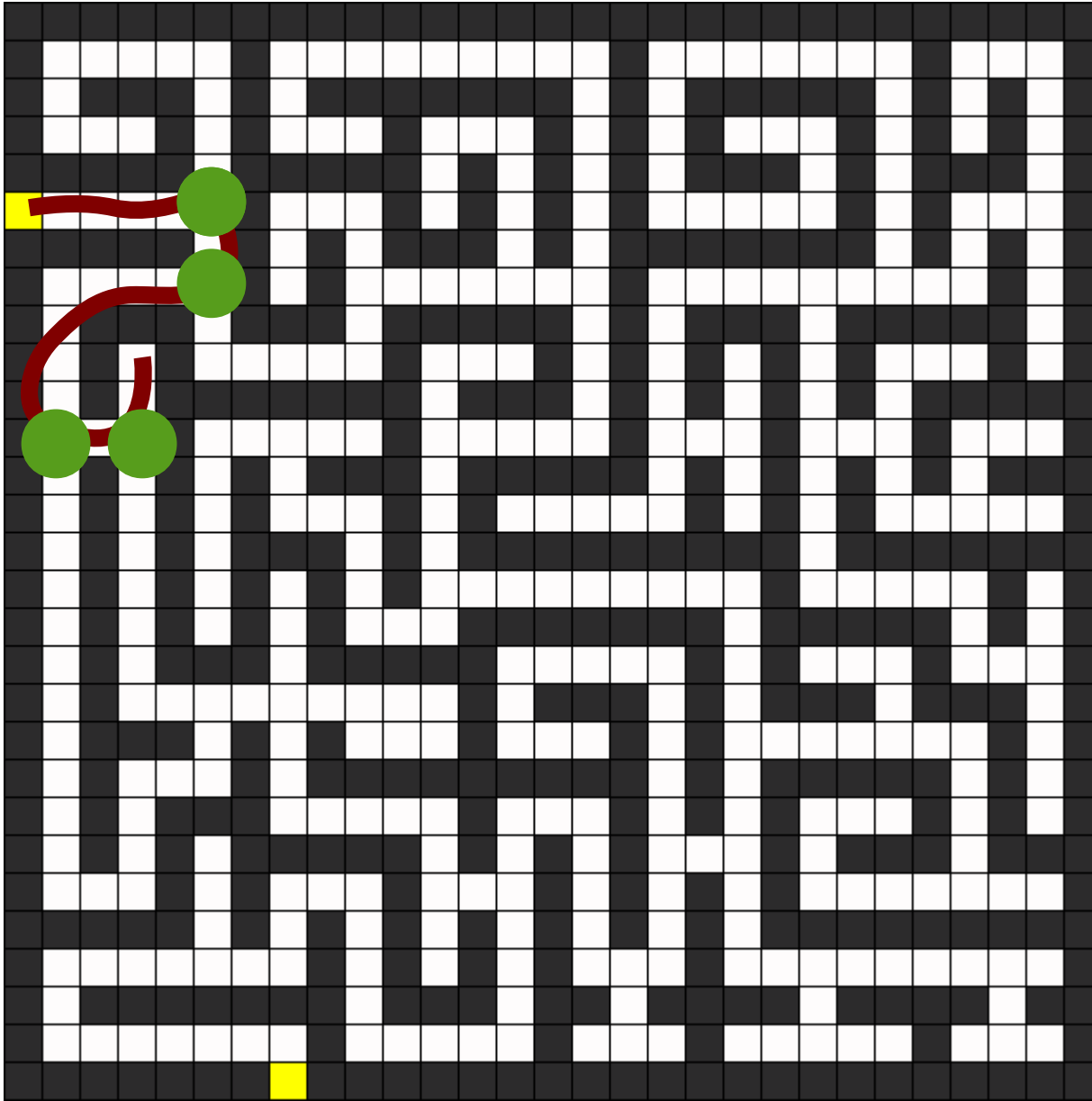


Declarative style of programming:

you give just the labyrinth.

the computer finds the way.

- For instance, via *trial*, *error* and *backtracking*.

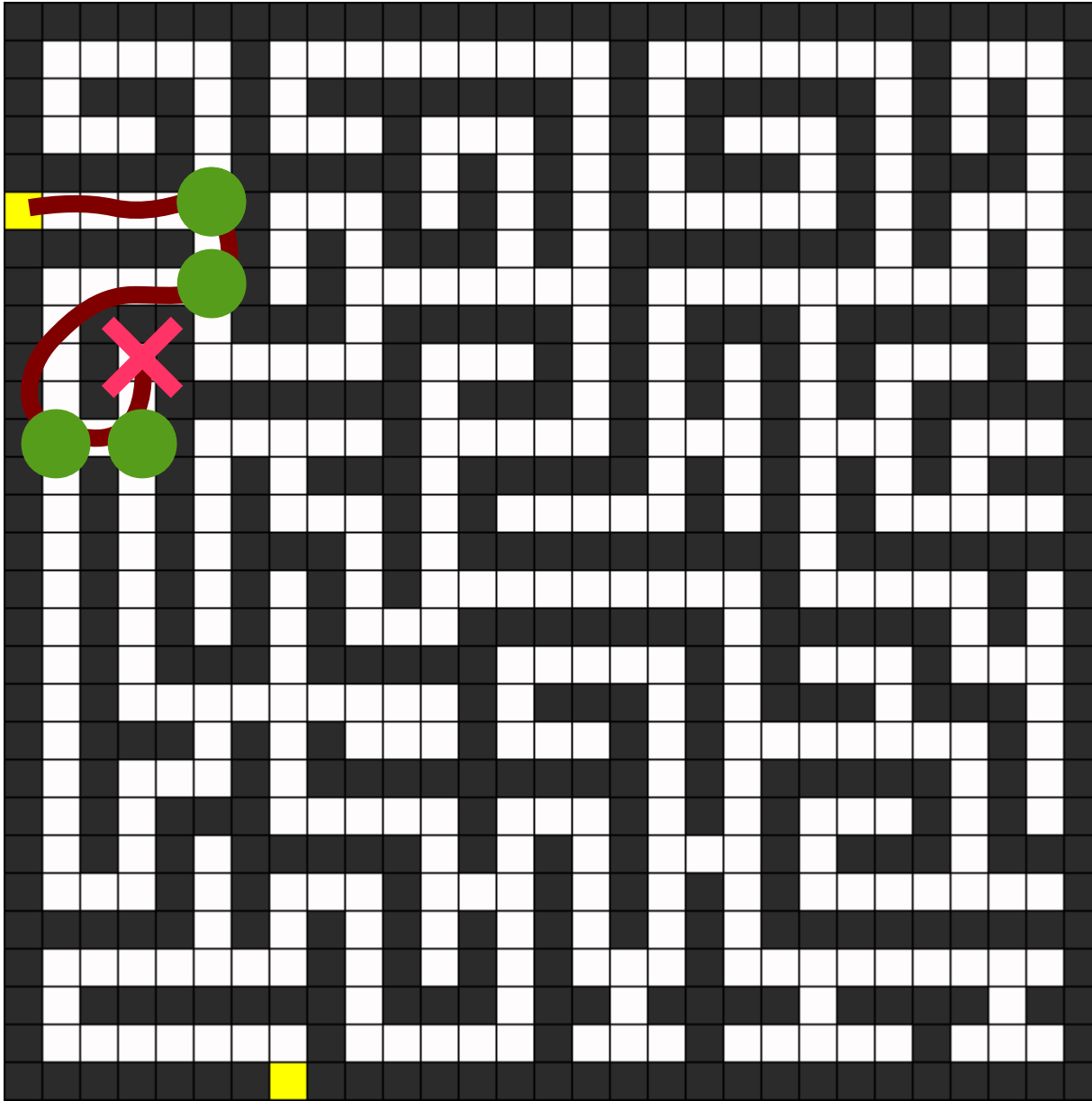


Declarative style of programming:

you give just the
labyrinth.

the computer finds
the way.

- For instance, via *trial*, *error* and *backtracking*.

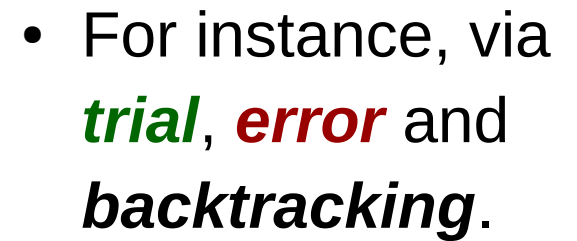


Declarative style of programming:

you give just the
labyrinth.

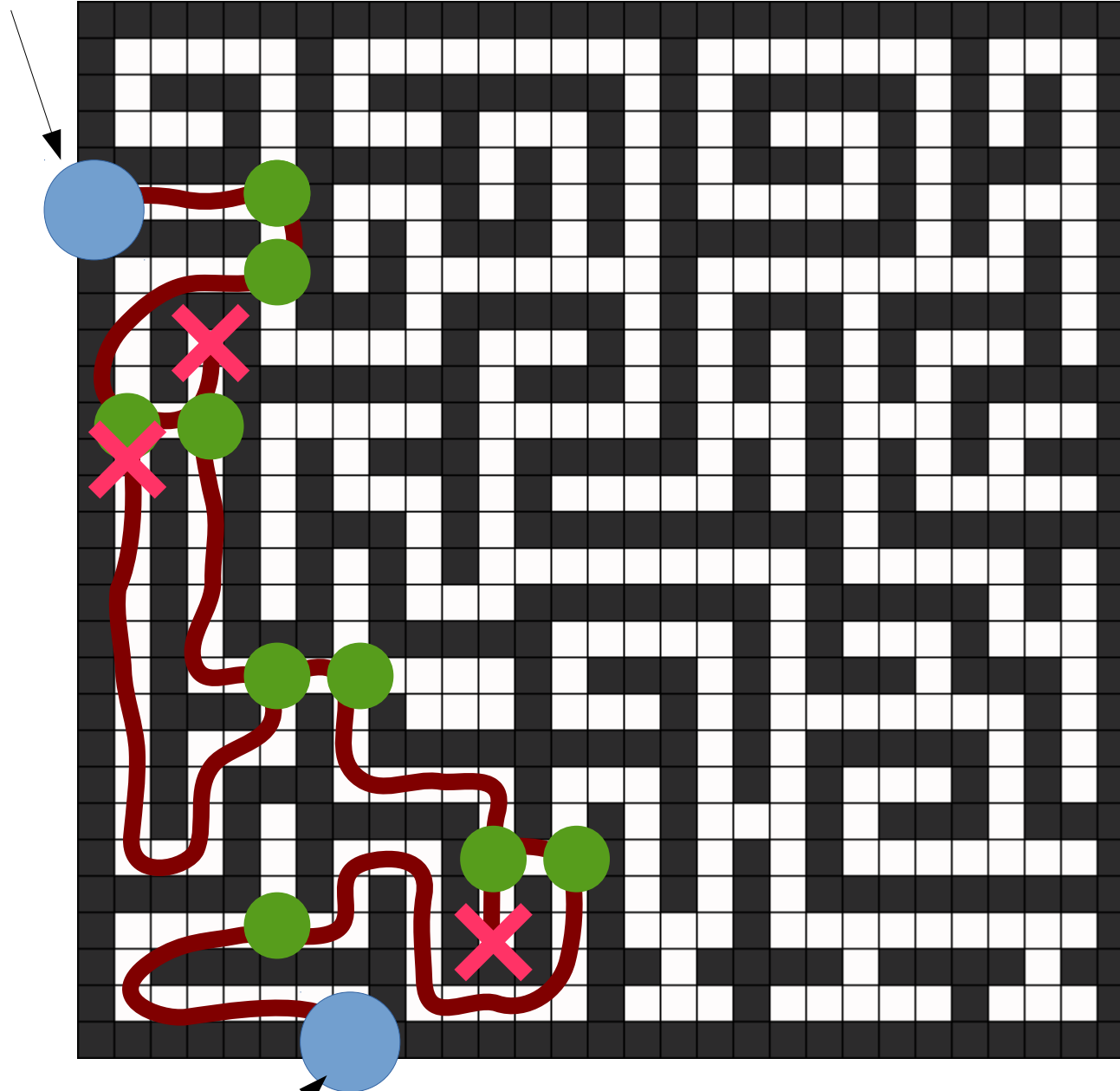
the computer finds
the way.

- For instance, via *trial*, *error* and *backtracking*.



Initial state

WELL-DEFINED PROBLEM



Goal state

KNOWLEDGE

Declarative style of programming:

you give just the labyrinth.

the computer finds the way.

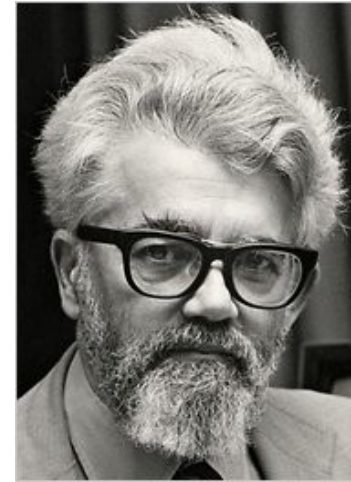
- For instance, via *trial*, *error* and *backtracking*.

PROBLEM-SOLVING METHOD

Well-defined problems & problem spaces

Problems are *well-defined* when there is a simple test to conclude whether a solution is a solution.

J. McCarthy (1956) The inversion of functions defined by Turing machines. Automata Studies, Annals of Mathematical Studies, 34:177 – 181.



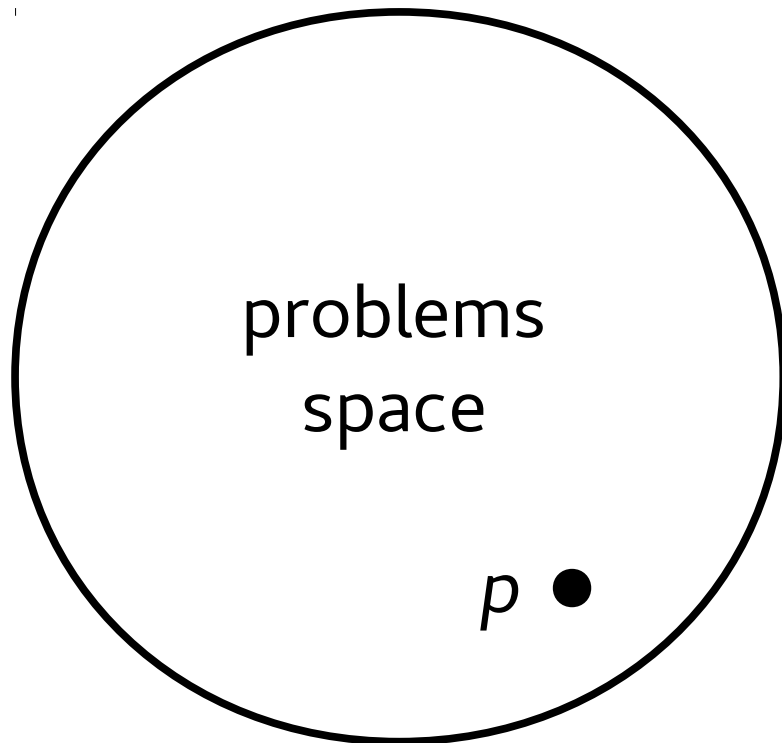
People solve problems by *searching* through a problem space, consisting of the *initial state*, the *goal state*, and *all possible states in between*.

Newell, A., & Simon, H. A. (1972). Human problem solving.

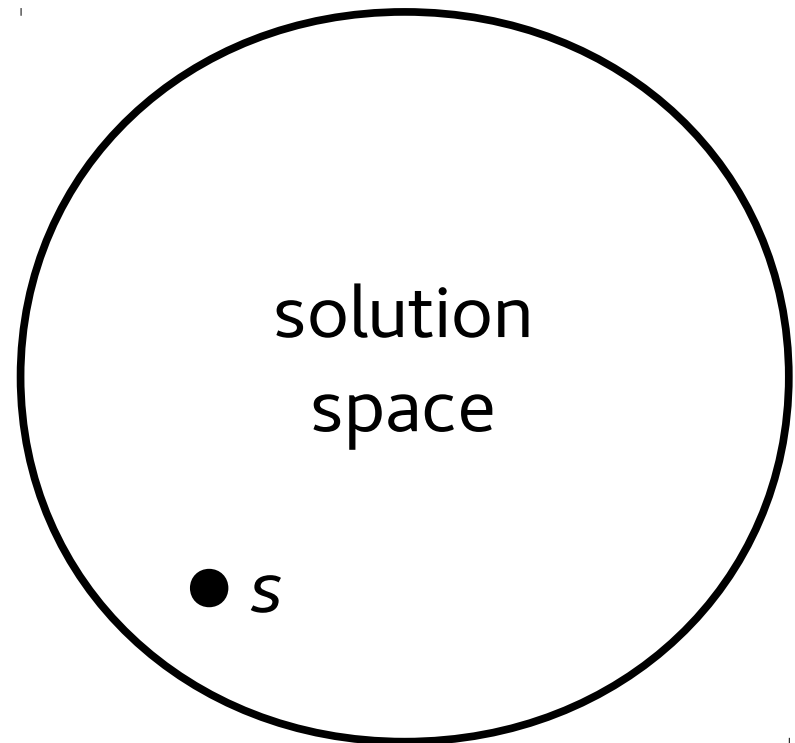


Problem and solution spaces

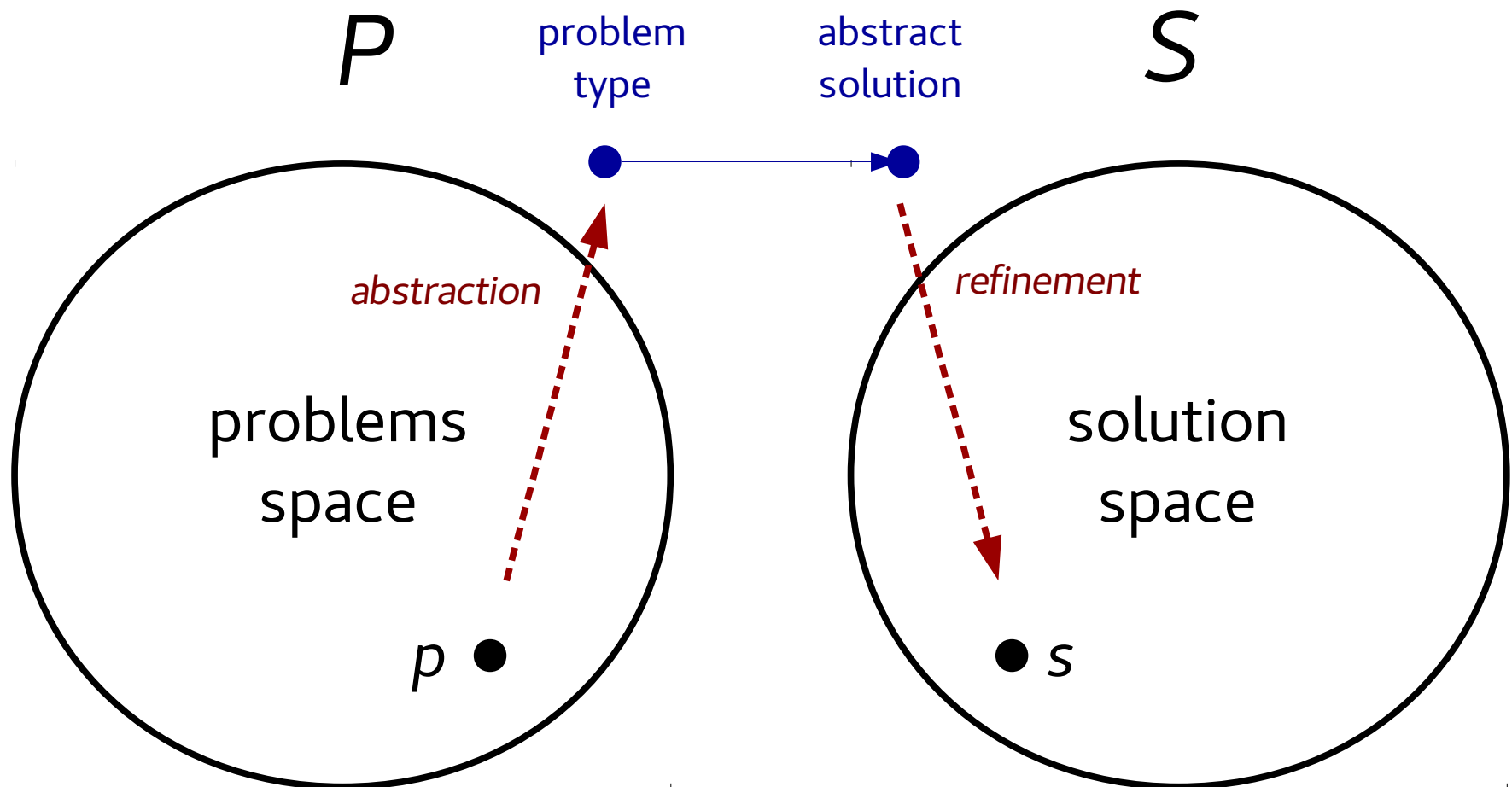
P



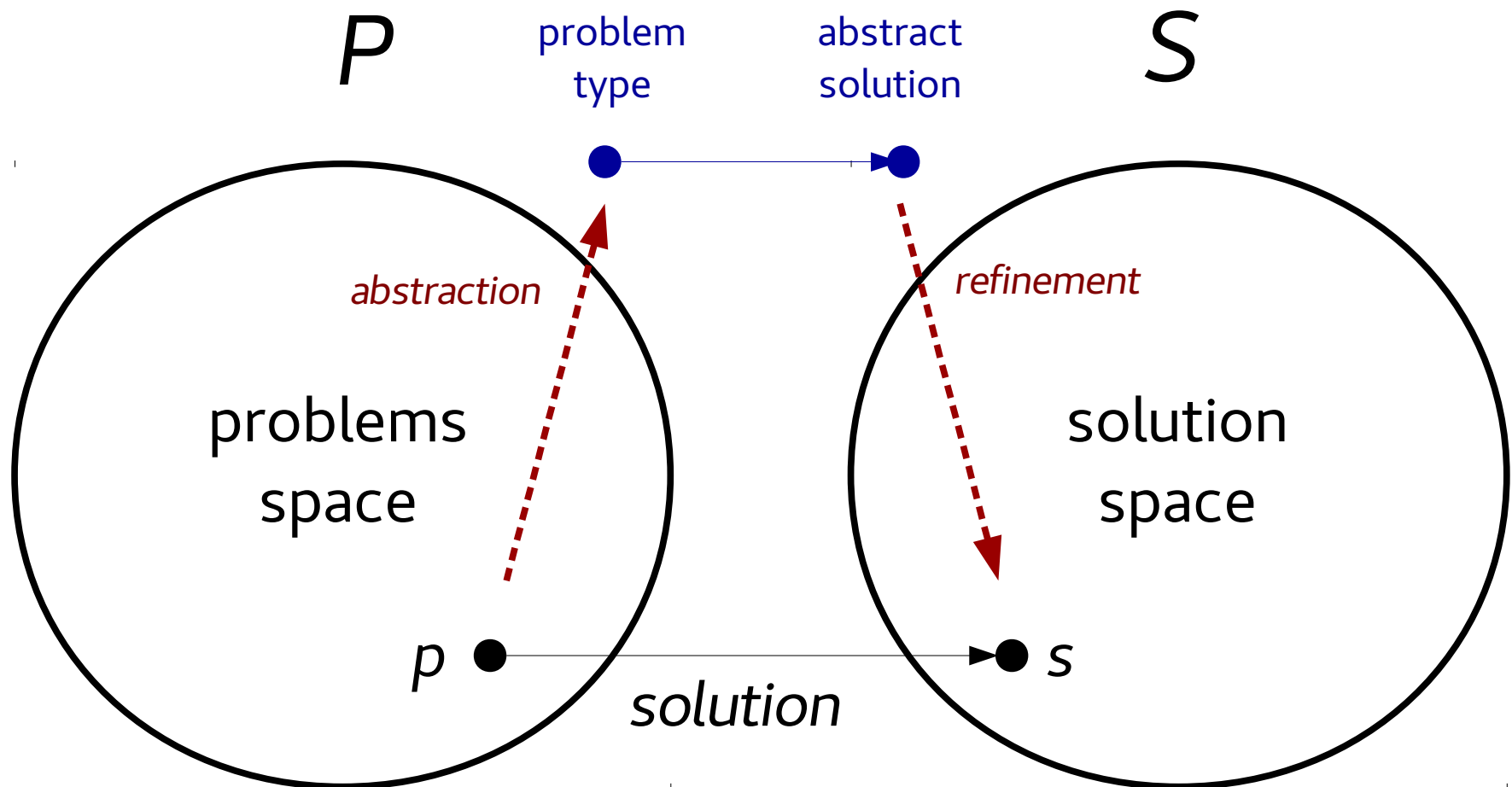
S



Problem and solution spaces



Problem and solution spaces



Defining the problem...

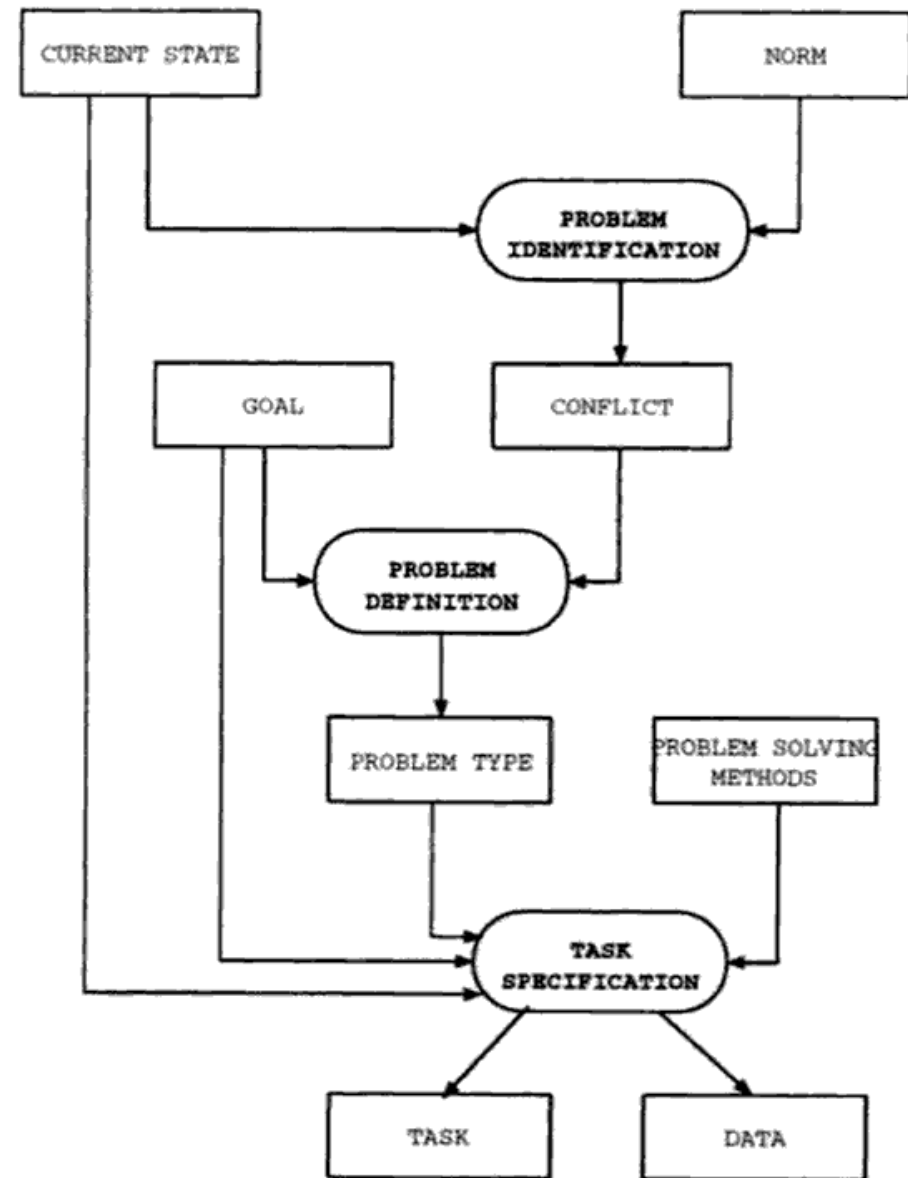
An old lady wants to visit her friend in a neighbouring village. She takes her car, but halfway the engine stops after some hesitations. On the side of the road she tries to restart the engine, but to no avail.



Which is the problem here?

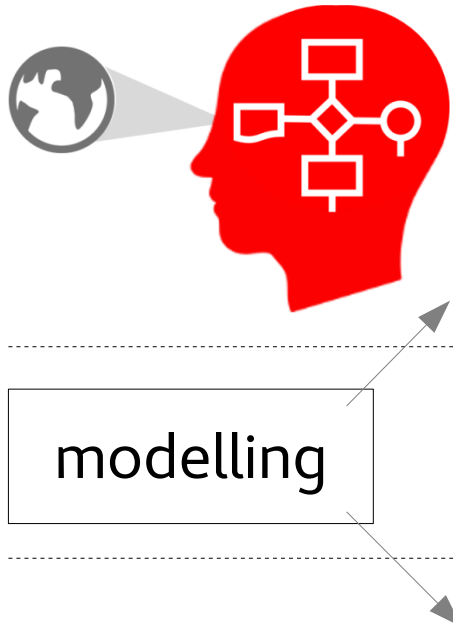
Breuker, J. (1994). Components of problem solving and types of problems. A Future for Knowledge Acquisition, 867, 118–136.

from ill-defined
to well-defined
problems...



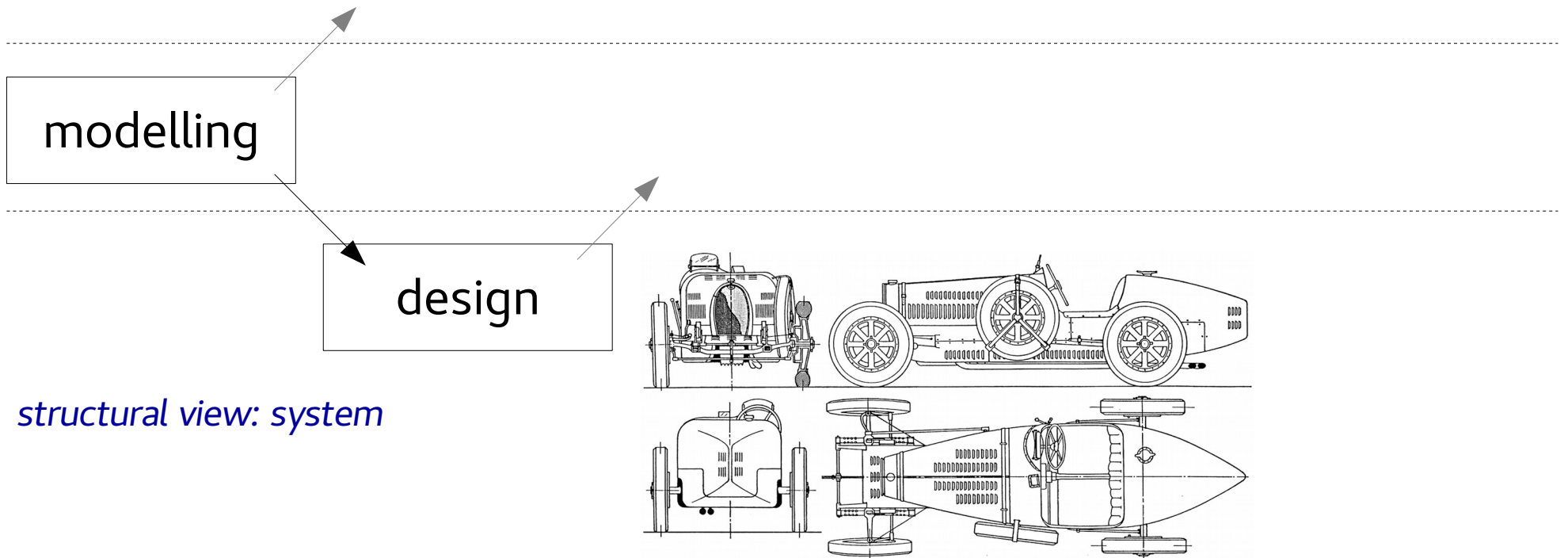
Breuker, J. (1994). Components of problem solving and types of problems. A Future for Knowledge Acquisition, 867, 118-136.

Suite of problem types



Breuker, J. (1994). Components of problem solving and types of problems. A Future for Knowledge Acquisition, 867, 118–136.

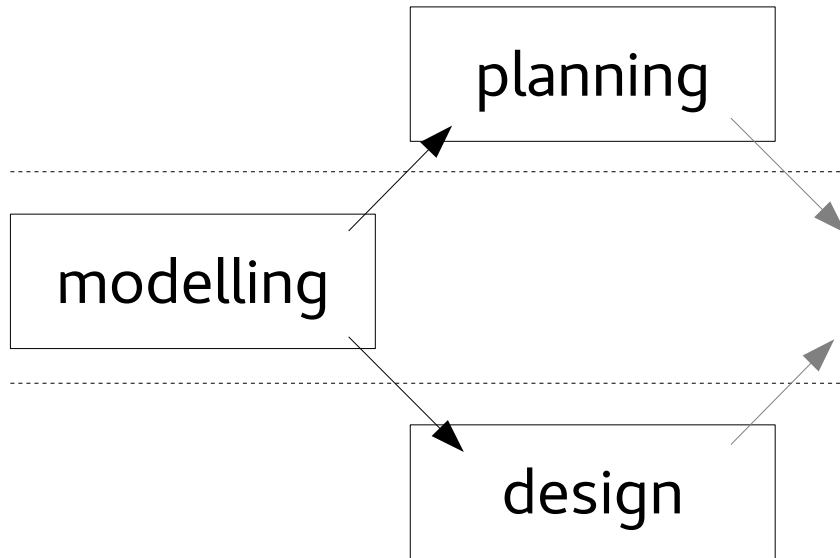
Suite of problem types



Breuker, J. (1994). Components of problem solving and types of problems. A Future for Knowledge Acquisition, 867, 118–136.

Suite of problem types

behavioural view: system + environment

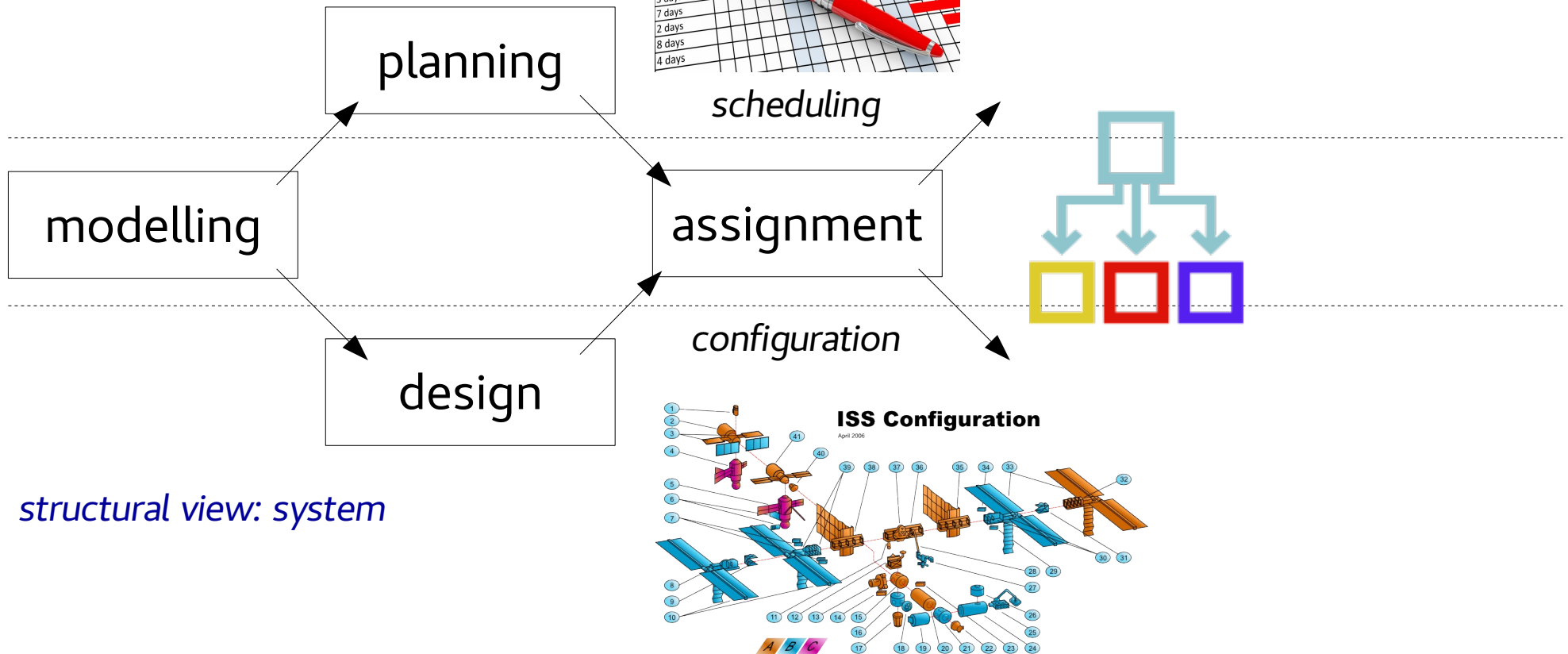


structural view: system

Breuker, J. (1994). Components of problem solving and types of problems. A Future for Knowledge Acquisition, 867, 118–136.

Suite of problem types

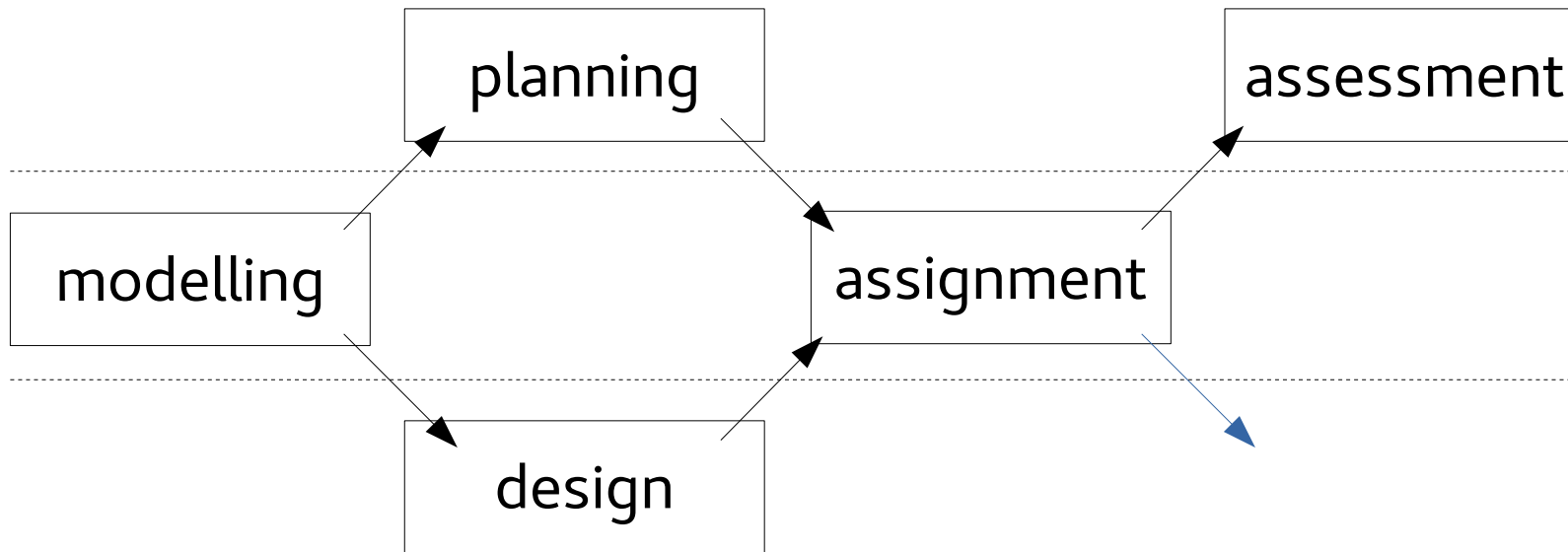
behavioural view: system + environment



Breuker, J. (1994). Components of problem solving and types of problems. A Future for Knowledge Acquisition, 867, 118–136.

Suite of problem types

behavioural view: system + environment

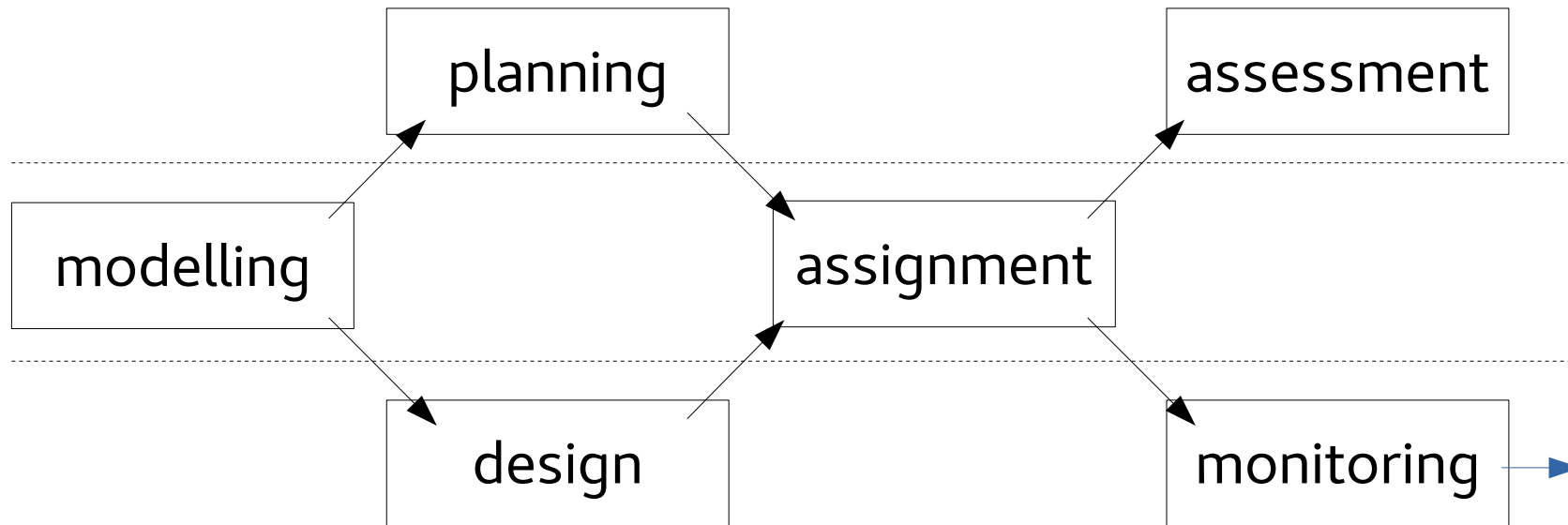


structural view: system

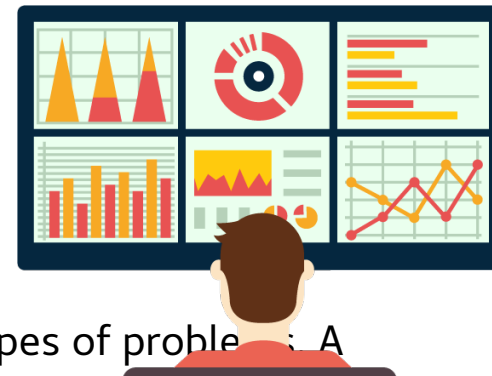
Breuker, J. (1994). Components of problem solving and types of problems. A Future for Knowledge Acquisition, 867, 118-136.

Suite of problem types

behavioural view: system + environment



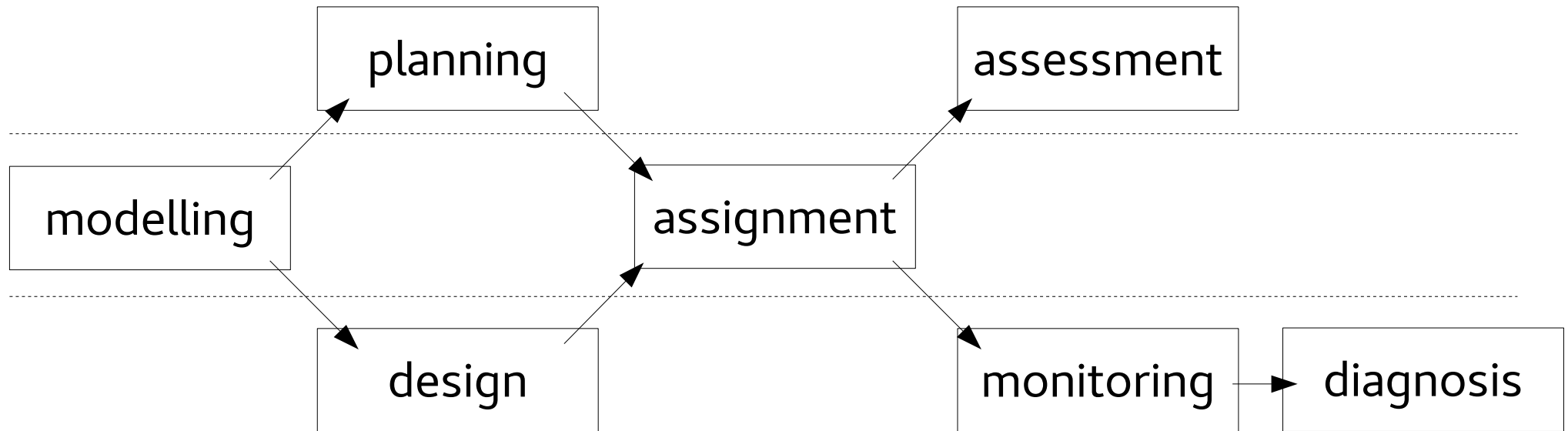
structural view: system



Breuker, J. (1994). Components of problem solving and types of problem : A Future for Knowledge Acquisition, 867, 118–136.

Suite of problem types

behavioural view: system + environment



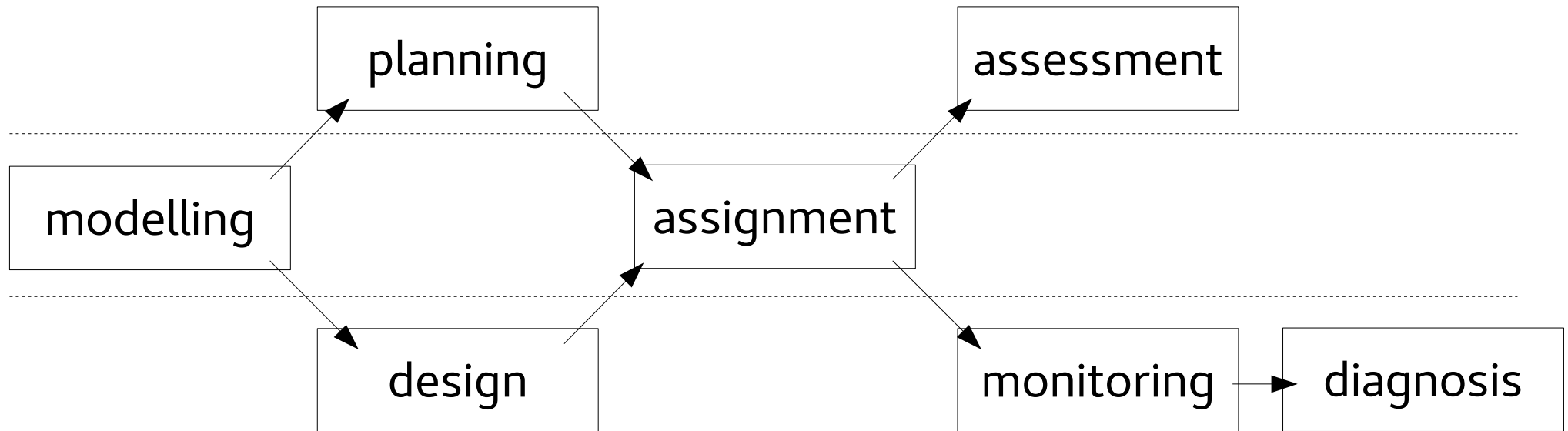
structural view: system



Breuker, J. (1994). Components of problem solving and types of problems. A Future for Knowledge Acquisition, 867, 118–136.

Suite of problem types

behavioural view: system + environment



structural view: system

AI researchers studied **problem-solving methods** and associated **knowledge structures** for each problem type.

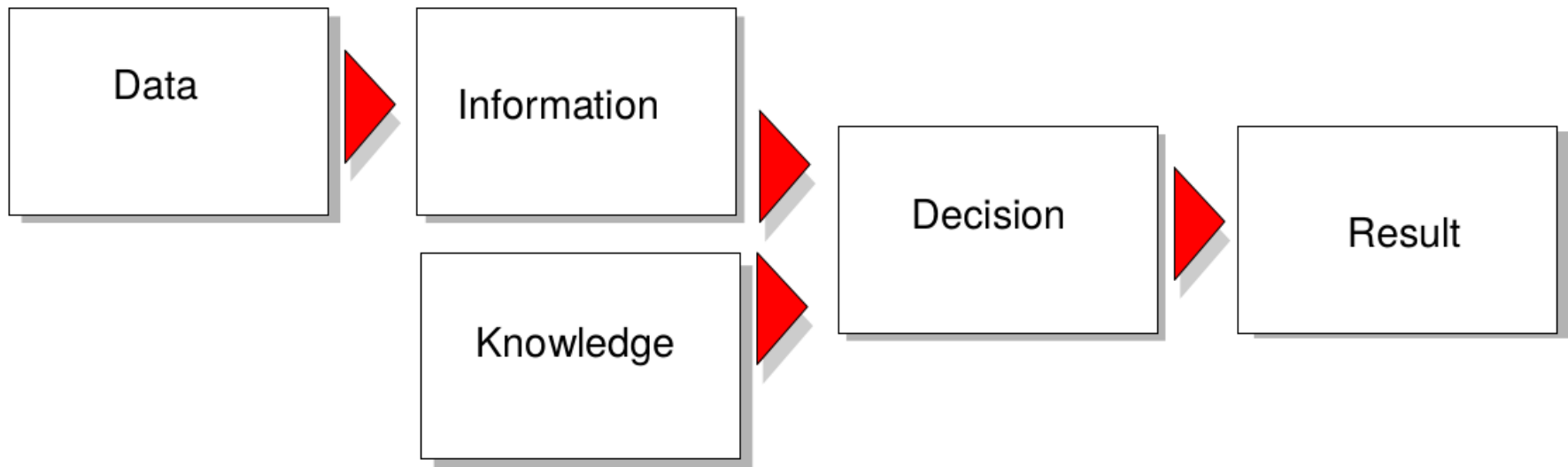
What is Knowledge in AI?

- **Knowledge** is what we ascribe to an agent to predict its behaviour following principles of rationality.

Note: this **knowledge representation** is not intended to be an accurate, physical model.

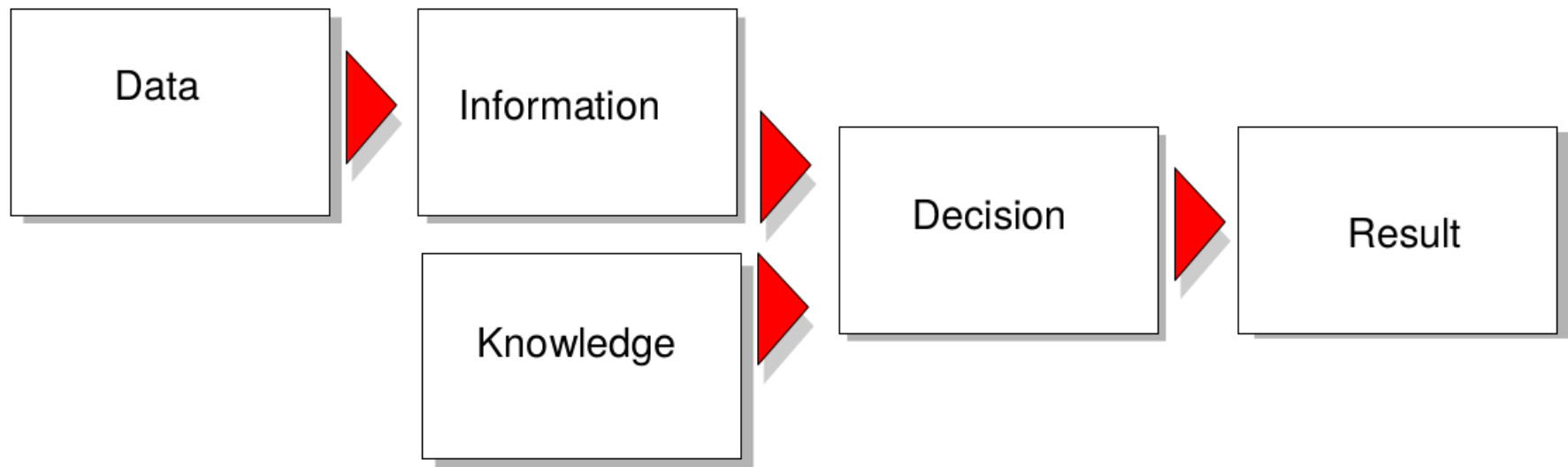
Data, Information, Knowledge

- **Data:** uninterpreted signals or symbols



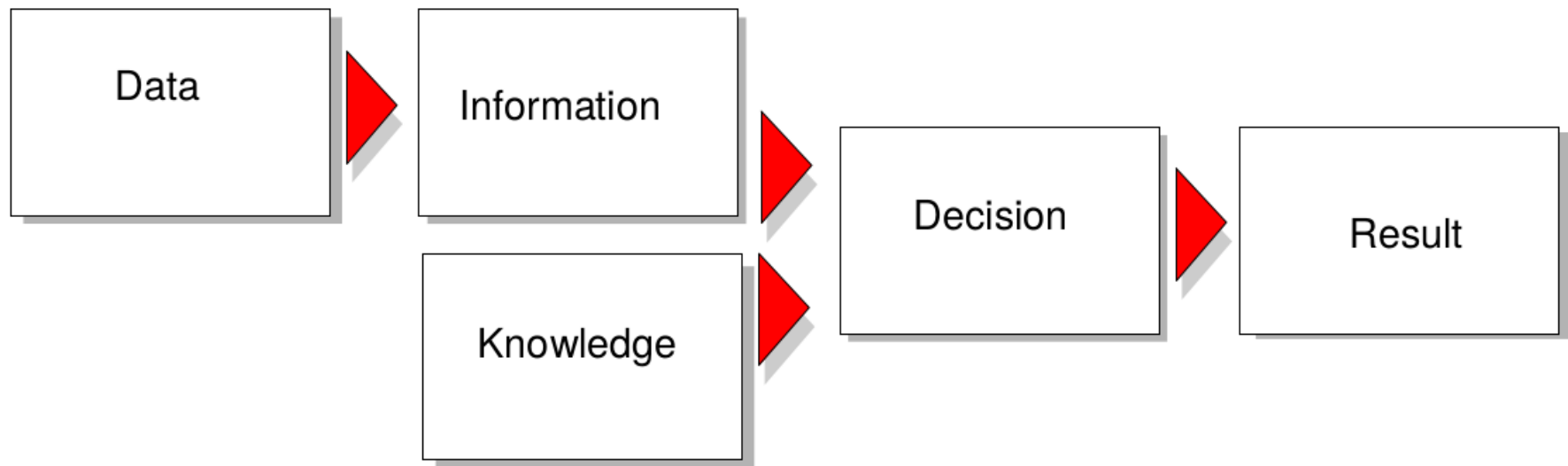
Data, Information, Knowledge

- **Data:** uninterpreted signals or symbols
- **Information:** data with added meaning



Data, Information, Knowledge

- **Data:** uninterpreted signals or symbols
- **Information:** data with added meaning
- **Knowledge:** all data and information that people use to *act, accomplish tasks* and to *create new information* (e.g. know-how, -why, -who, -where and -when).



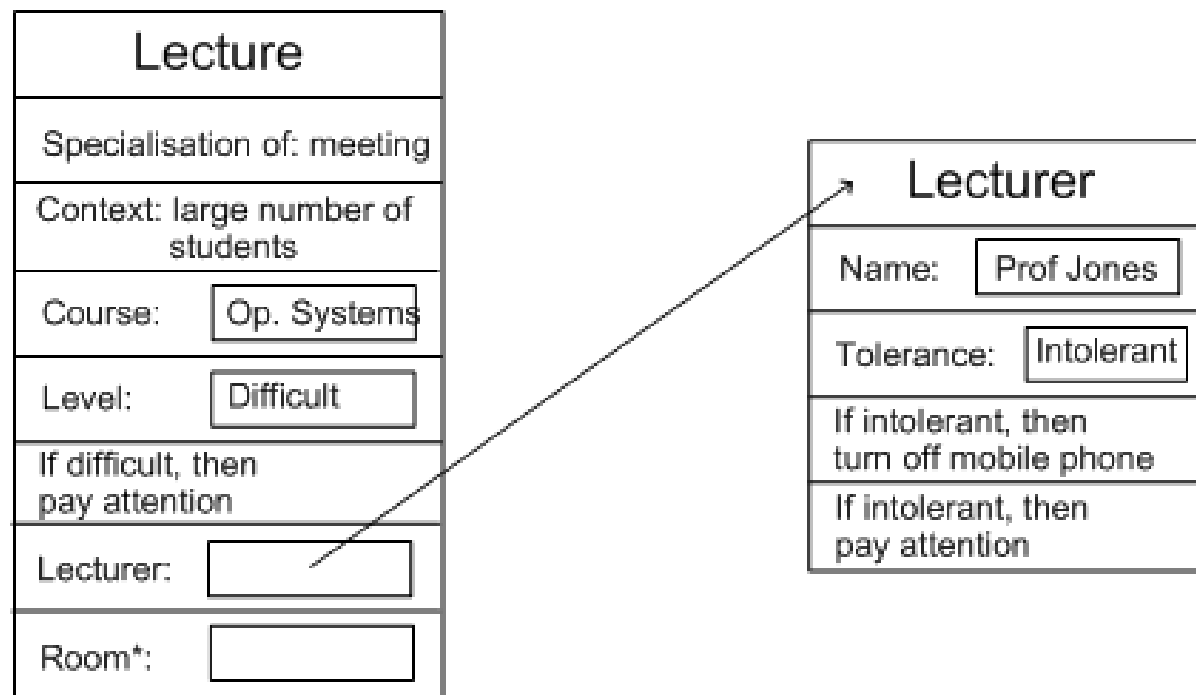
Expert system (rule base)

```
if flower and seed then phanerogam
if phanerogam and bare-seed then fir
if phanerogam and 1-cotyledon then monocotyledonous
if phanerogam and 2-cotyledon then dicotyledonous
if monocotyledon and rhizome then thrush
if dicotyledon then anemone
if monocotyledon and ¬rhizome then lilac
if leaf and flower then cryptogamous
if cryptogamous and ¬root then foam
if cryptogamous and root then fern
if ¬leaf and plant then thallophyte
if thallophyte and chlorophyll then algae
if thallophyte and ¬ chlorophyll then fungus
if ¬leaf and ¬flower and ¬plant then colibacille
```

rhizome + flower + seed + 1-cotyledon ?

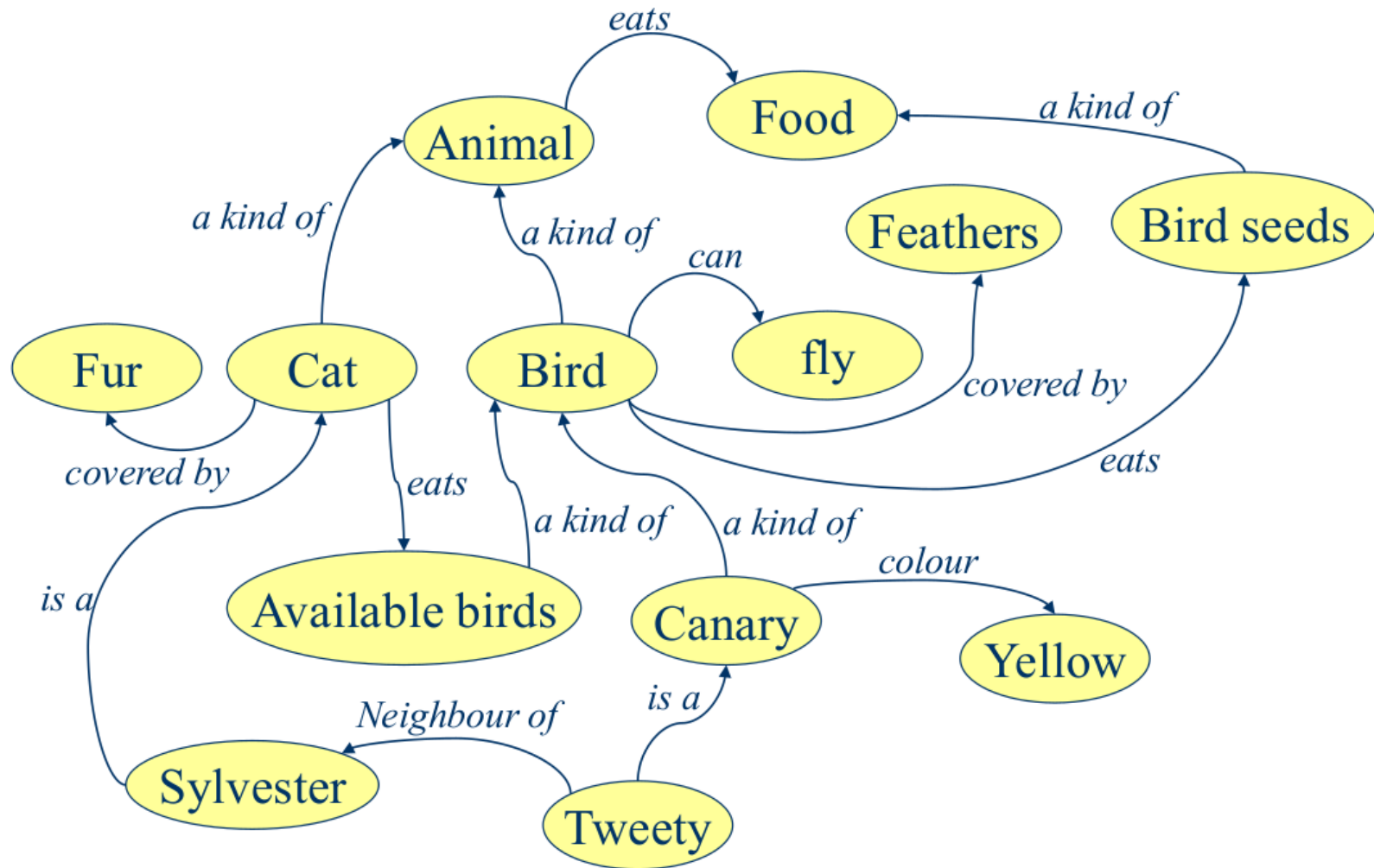
Frames

- Frames are "stereotyped" knowledge units representing situations, objects or events or (classes) sets of such entities.



(base for the **Object-Oriented Programming** paradigm)

Semantic Networks

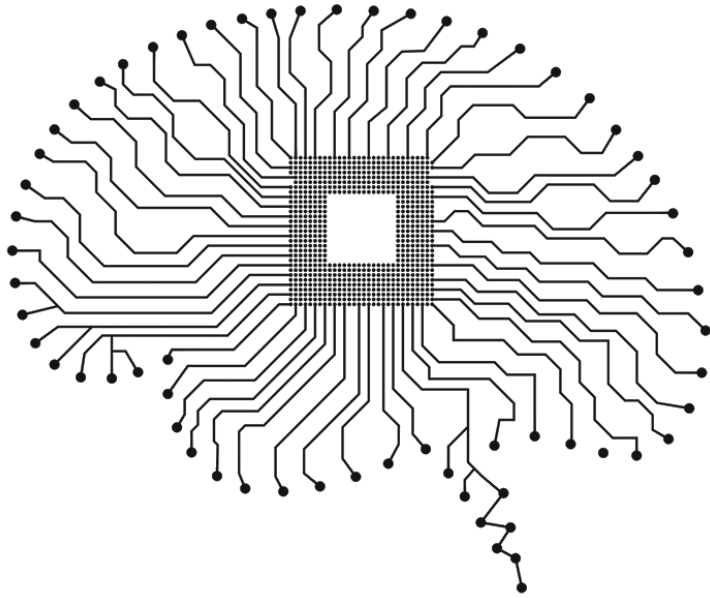


(used in contemporary **Semantic Web** technologies)

In sum

- Symbolic AI presents ***transparent*** techniques to effectively model and solve problems **that can be described in symbolic terms** (*where expertise can be verbalized*).
- All IT systems of organizations today rely on some of the technologies introduced or emerged during the first AI wave.
- ***But these results are much inferior than what promised..*** (*even more in the 70s*).





A physical symbol system has the
necessary and sufficient means for
general intelligent action

Allen Newell and Herbert A. Simon

Computer Science as Empirical Inquiry: Symbols and Search (1976)

Acknowledged limitations

- **knowledge acquisition bottleneck**
- **scaling or modularity**
- **tractability** (e.g. *ramification problem*)
- **symbol grounding**

Acknowledged limitations

- **knowledge acquisition bottleneck**
- **scaling or modularity**
- **tractability** (e.g. *ramification problem*)
- **symbol grounding**
- **natural language**
- **sensory-motor tasks**
 - *computer vision,*
 - *speech recognition,*
 - *actuator control*

Acknowledged limitations

- **knowledge acquisition bottleneck**
 - **scaling or modularity**
 - **tractability** (e.g. *ramification problem*)
 - **symbol grounding**
 - natural language
 - sensory-motor tasks
 - *computer vision*,
 - *speech recognition*,
 - *actuator control*
-

Hacking solutions

- Scruffies never believed the mind was a monolithical system, so they tinkered with heuristics, *ad-hoc* methods, and opportunistically with logic (“neat shells for scruffy approaches”).

(the first chatbot)

ELIZA

Weizenbaum ~1965

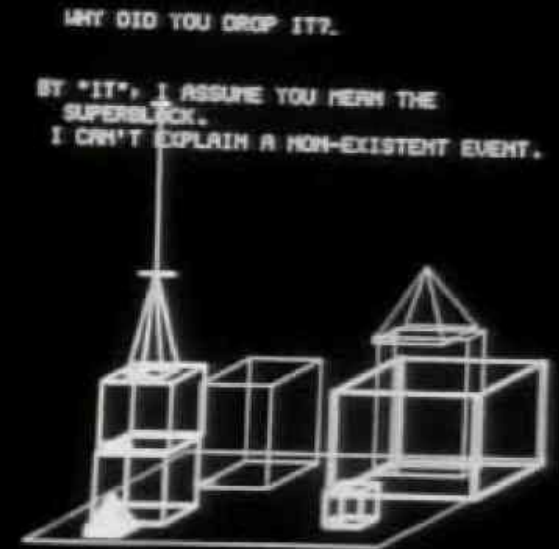
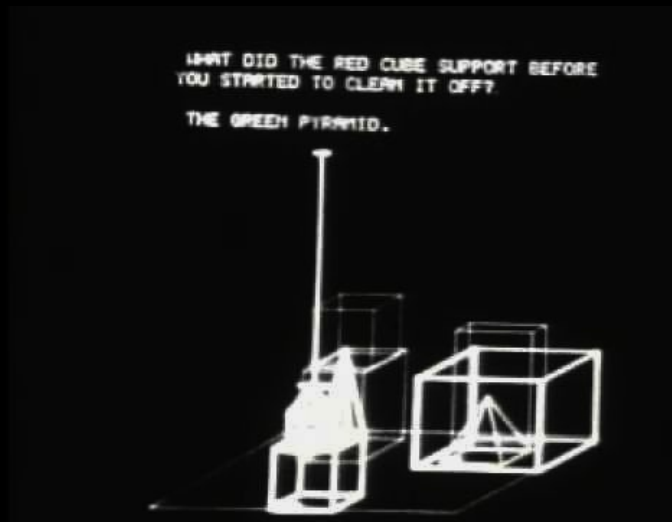
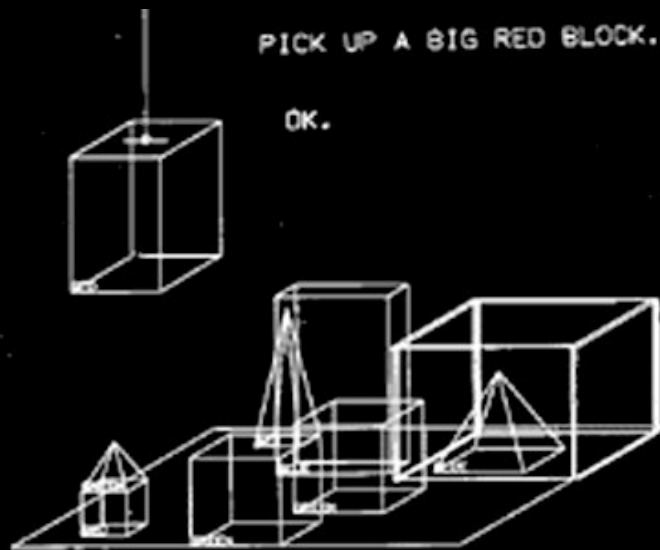
```
ELIZA: Is something troubling you ?
YOU: I am bored
ELIZA: Do you know anyone else who is bored ?
YOU: Maybe
ELIZA: You aren't sure ?
YOU: I don't really know
ELIZA: Why don't you really know ?
YOU: Well, I haven't really asked
ELIZA: Do you say you haven't really asked for some special reason ?
YOU: I'm feeling blue
ELIZA: Do you believe it is normal to be feeling blue ?
YOU: What's normal?
ELIZA: What is it you really want to know ?
YOU: Why don't I have more energy in the morning?
ELIZA: That's quite interesting.
YOU: Not really
ELIZA: Do you feel strongly about discussing such things ?
YOU: Yea, I'm LONELY
ELIZA: Is it because you are lonely that you came to me ?
YOU: Duh
ELIZA: Does talking about this bother you ?
YOU: It is SO NICE to TALK TO YOU
ELIZA: We were discussing you -- not me.
YOU:
```

```
=====
EEEEEEEE L      IIIIII ZZZZZZZ AAA
E         L      I      Z      A  A
E         L      I      Z      A  A
EEEEEE   L      I      Z      A  A
E         L      I      Z      AAAAAA
E         L      I      Z      A    A
EEEEEEEE LLLLLLL IIIIII ZZZZZZ A    A
=====
ELIZA > (PLEASE TYPE IN ALL CAPS) WHAT'S YOUR NAME DEAR ?
AMIT
ELIZA > HI, I'M ELIZA. WHAT DO YOU WANT TO TALK ABOUT ?
AMIT > CAN I TALK ABOUT MY PROBLEM ?
ELIZA > SURE... !
AMIT > DUE TO ON GOING EXAMS IN COLLEGE, I HAVE A LOT OF STRESS.
ELIZA > PLEASE RELAX A LITTLE. SLEEP WELL.
AMIT > THANKS FOR YOUR ADVICE.
ELIZA > NO MENTION.
AMIT > BYE
ELIZA > BYE AND KEEP IN TOUCH...
=====
```

Still running e.g. on: <https://www.masswerk.at/elizabot/eliza.html>

SHRDLU

Winograd ~1969



- Deeper linguistic understanding
- but limited to simple *blocks* worlds

Acknowledged limitations

- **knowledge acquisition bottleneck**
 - **scaling or modularity**
 - **tractability** (e.g. *ramification problem*)
 - **symbol grounding**
 - natural language
 - sensory-motor tasks
 - *computer vision,*
 - *speech recognition,*
 - *actuator control*
-

Hacking solutions

Scruffies never believed the mind was a monolithical system, so they tinkered with heuristics, *ad-hoc* methods, and opportunistically with logic (“neat shells for scruffy approaches”).

- **but these successes were impossible to generalize.**

AI Winter (early 70s/80s)

- After a series of critical reports, funding to AI projects reduced massively. Researchers started to seek other names for their own research fields.



- Facing overwhelming difficulties to go beyond from toy problems, radically different paradigms started to be (re)considered, **renouncing** to symbolic representations.
- As Rodney Brooks famously put it:



“Elephants don't
play chess”

The revenge of machine learning

Machine learning

Machine learning is a process that enables artificial systems to improve from experience.

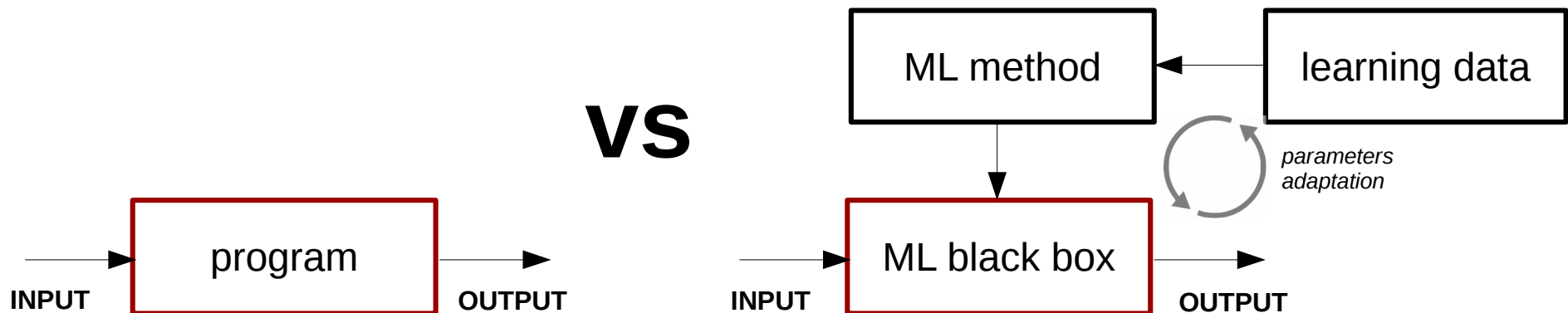


according to well-defined criteria

Machine learning

Machine learning is a process that enables artificial systems to improve from experience.

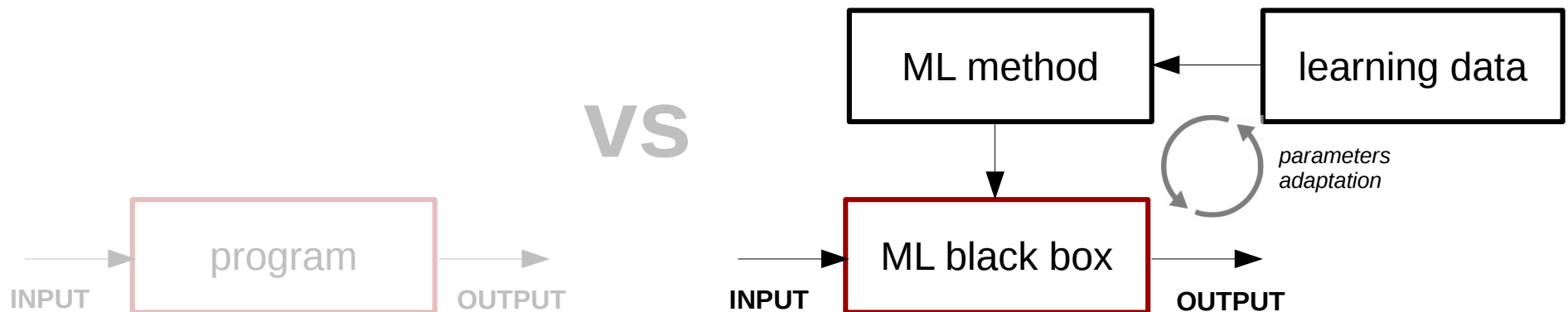
- Rather than writing a program, here the developer has to collect adequate training data and decide a ML method.



Machine learning

Machine learning is a process that enables artificial systems to improve from experience.

- Rather than writing a program, here the developer has to collect adequate training data and decide a ML method.



- Unfortunately, an **adequate parameter adaptation can be highly data-demanding**, especially for rich inputs.

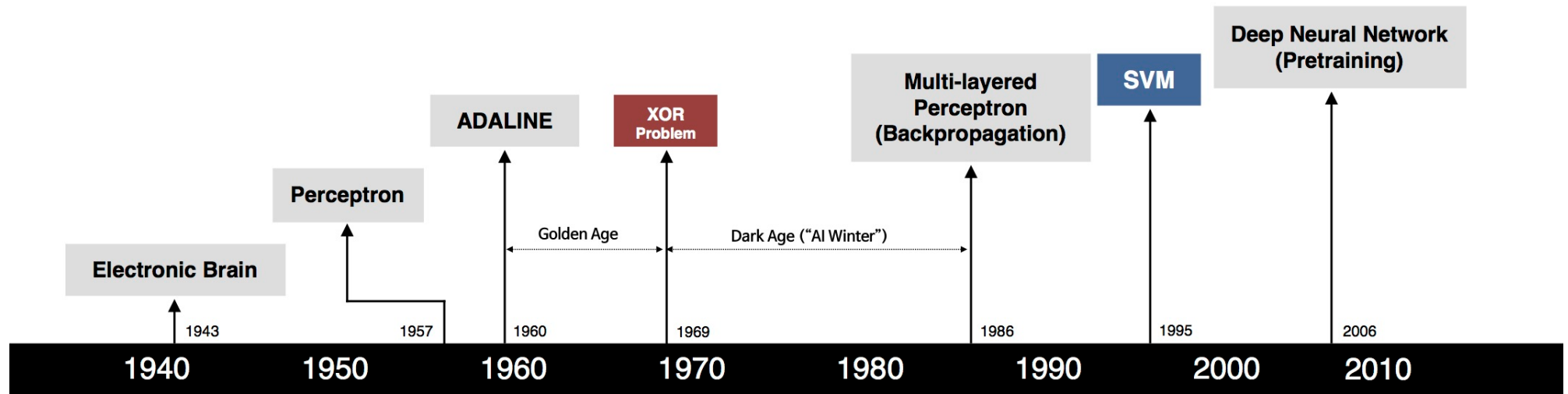
Machine learning & co.

Machine learning is a process that enables artificial systems to improve from experience.

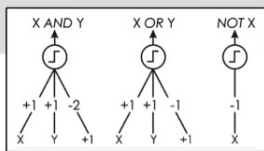
- Many learning methods are available, but studied and used by different communities!
 - **Neural networks** are only one among many.
- (In some situations, *evolutionary algorithms* can also be of use for this task)

Nice video applying evolutionary algorithms: <https://www.youtube.com/watch?v=pgaEE27nsQw>

Neural Networks timeline



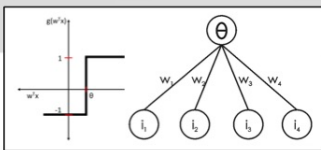
S. McCulloch - W. Pitts



- Adjustable Weights
- Weights are not Learned



F. Rosenblatt



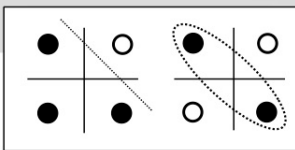
- Learnable Weights and Threshold



B. Widrow - M. Hoff



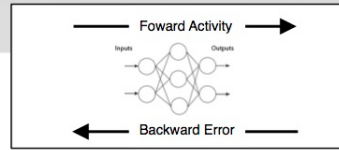
M. Minsky - S. Papert



- XOR Problem



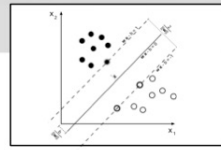
D. Rumelhart - G. Hinton - R. Williams



- Solution to nonlinearly separable problems
- Big computation, local optima and overfitting



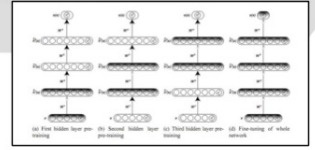
V. Vapnik - C. Cortes



- Limitations of learning prior knowledge
- Kernel function: Human Intervention



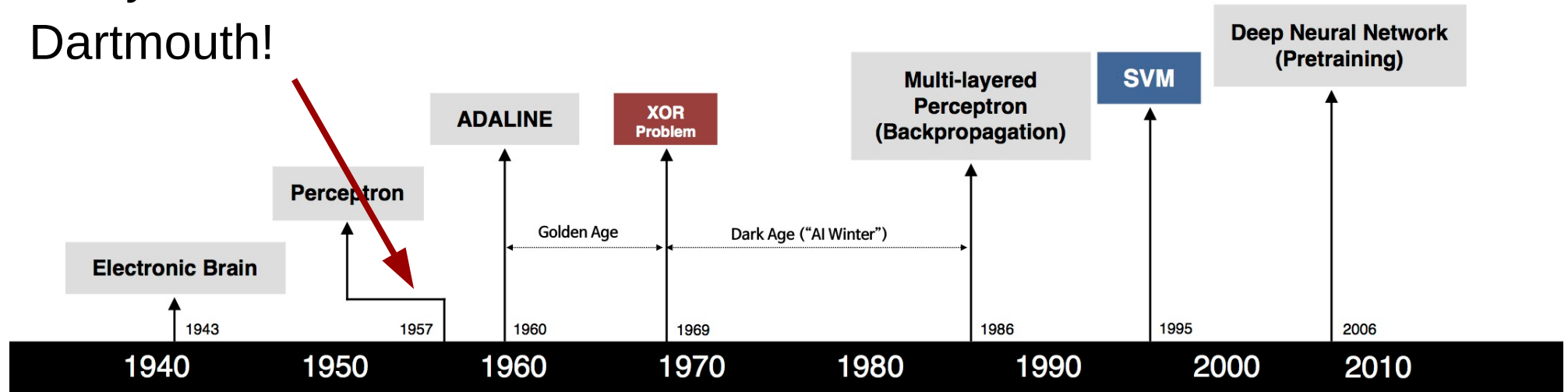
G. Hinton - S. Ruslan



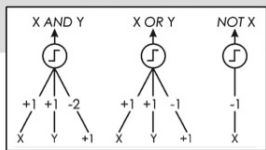
- Hierarchical feature Learning

Neural Networks timeline

one year after
Dartmouth!



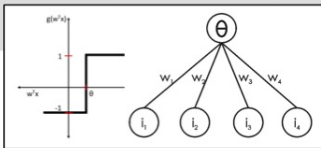
S. McCulloch - W. Pitts



- Adjustable Weights
- Weights are not Learned



F. Rosenblatt



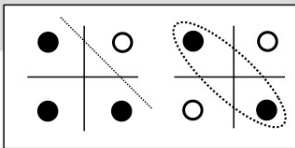
- Learnable Weights and Threshold



B. Widrow - M. Hoff



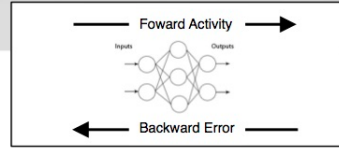
M. Minsky - S. Papert



- XOR Problem



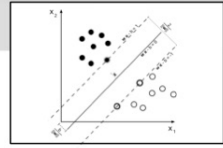
D. Rumelhart - G. Hinton - R. Williams



- Solution to nonlinearly separable problems
- Big computation, local optima and overfitting



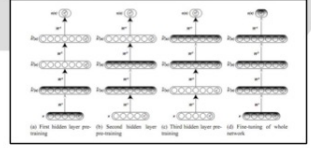
V. Vapnik - C. Cortes



- Limitations of learning prior knowledge
- Kernel function: Human Intervention

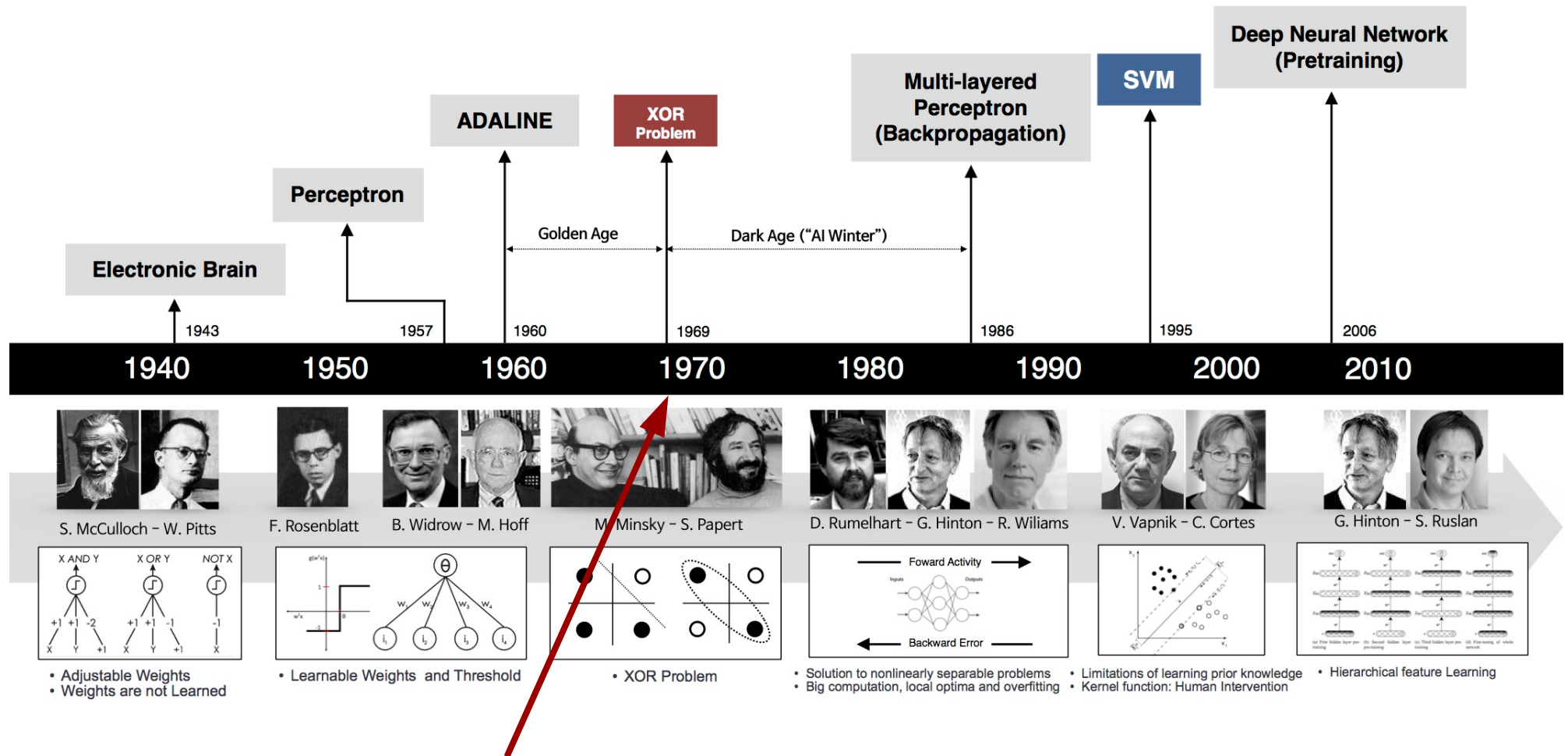


G. Hinton - S. Ruslan



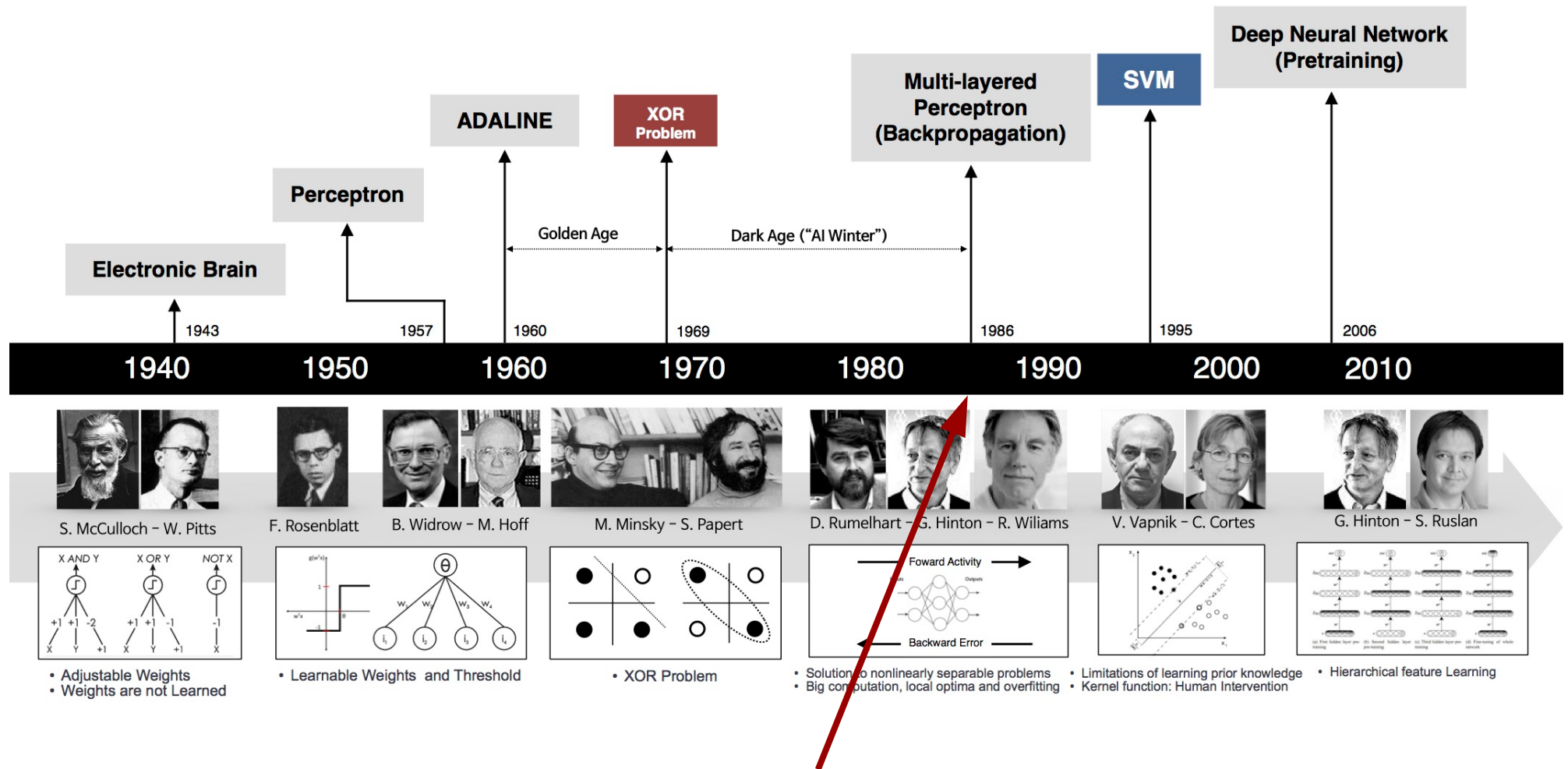
- Hierarchical feature Learning

Neural Networks timeline



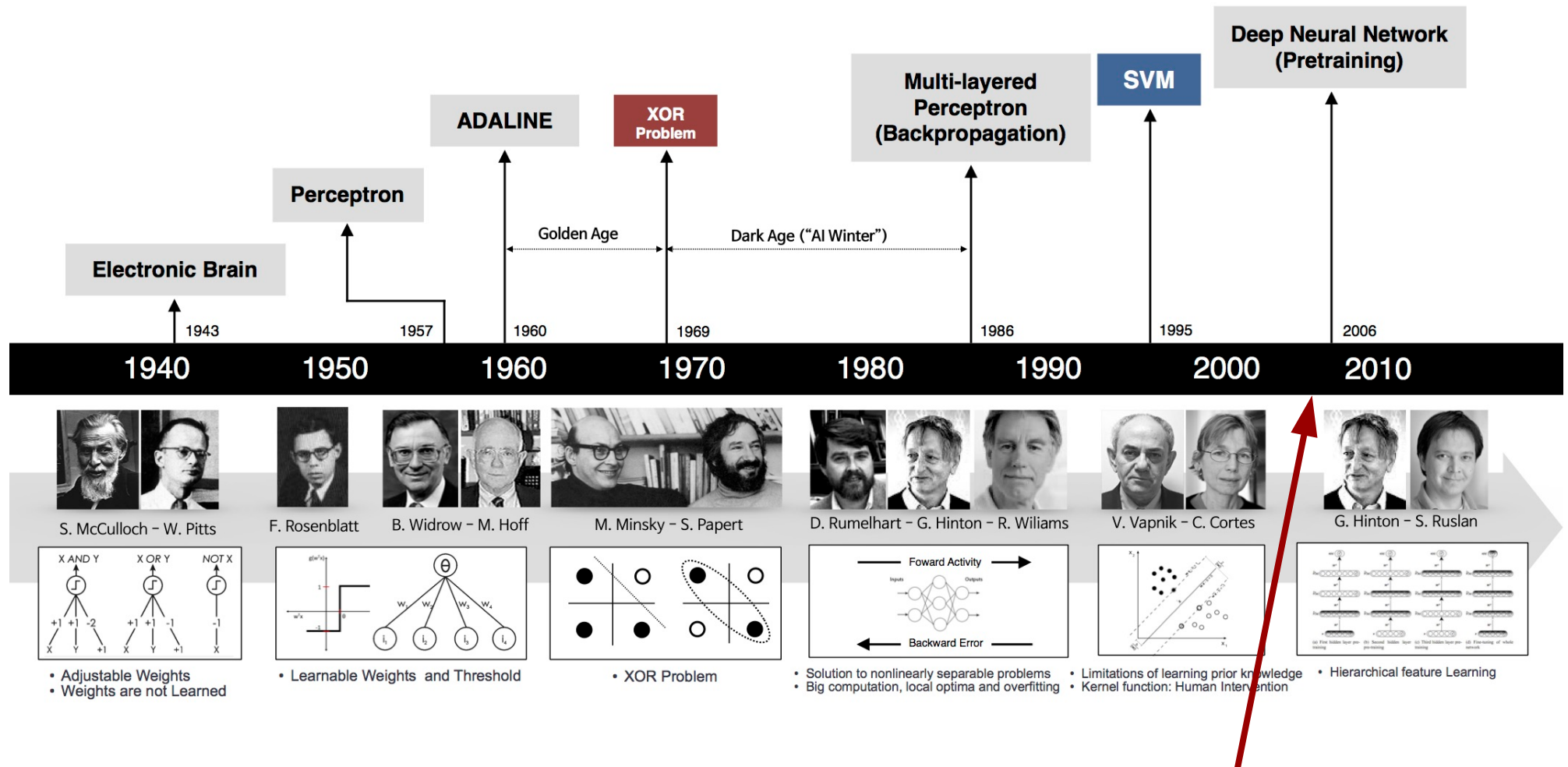
Minsky and Papert mathematically prove that the Perceptron could not model "exclusive or"

Neural Networks timeline



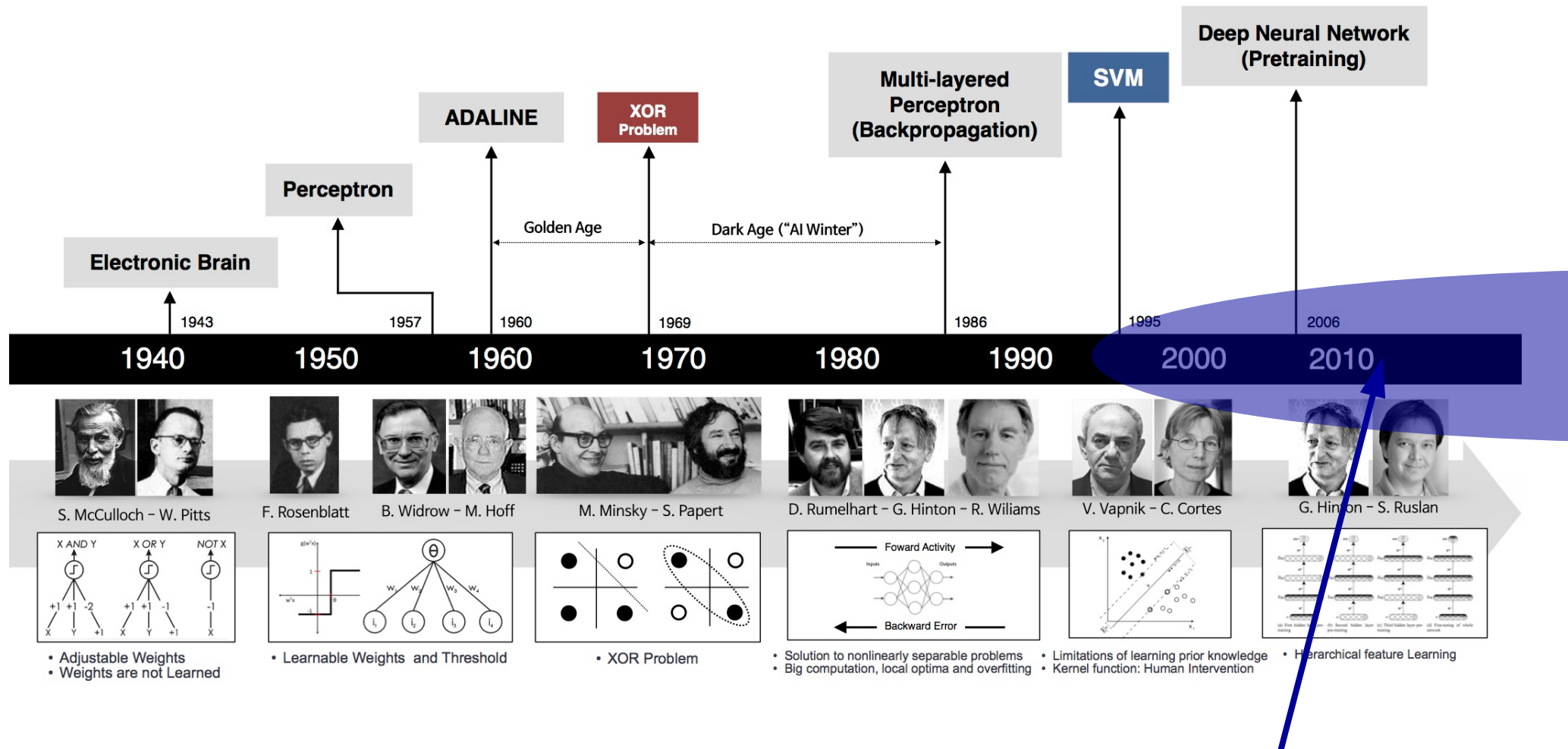
Backpropagation and the addition of layers solved the problem

Neural Networks timeline



reuse of previous training
possible → fine-tuning

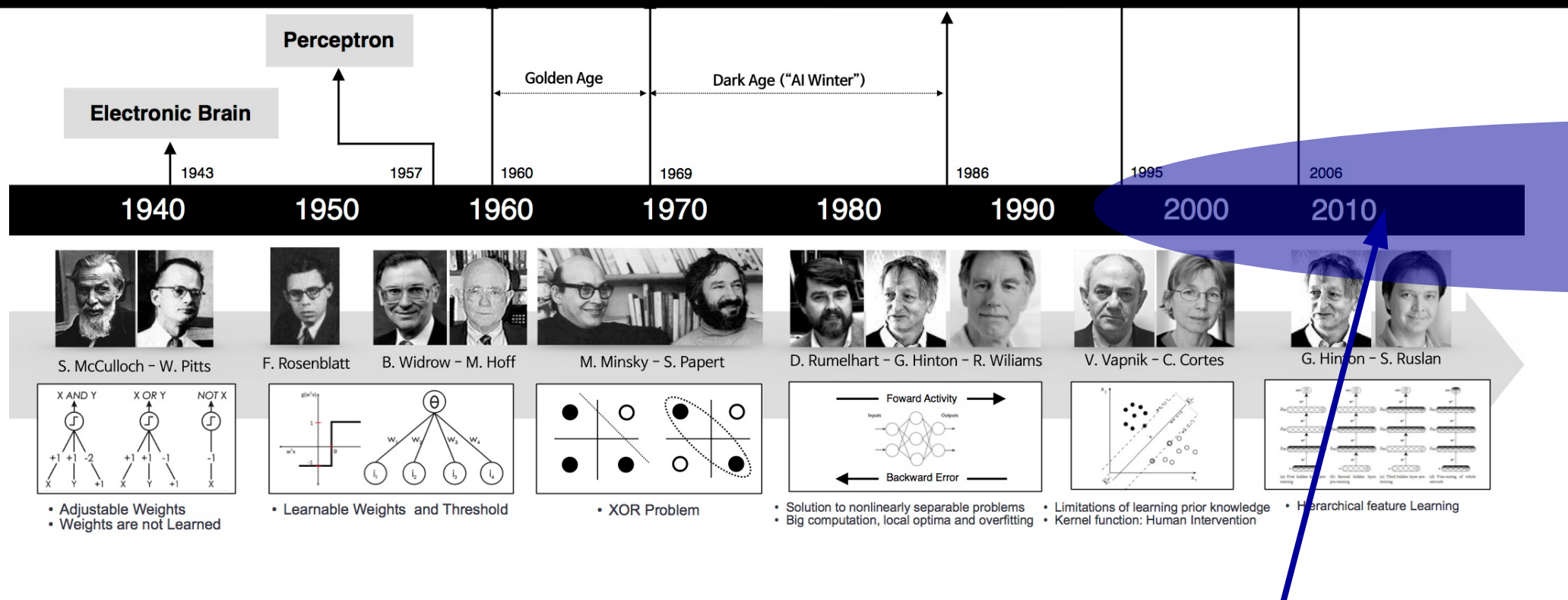
Neural Networks timeline



pervasive introduction of internet, smart devices, global IT corporations → **big data era**

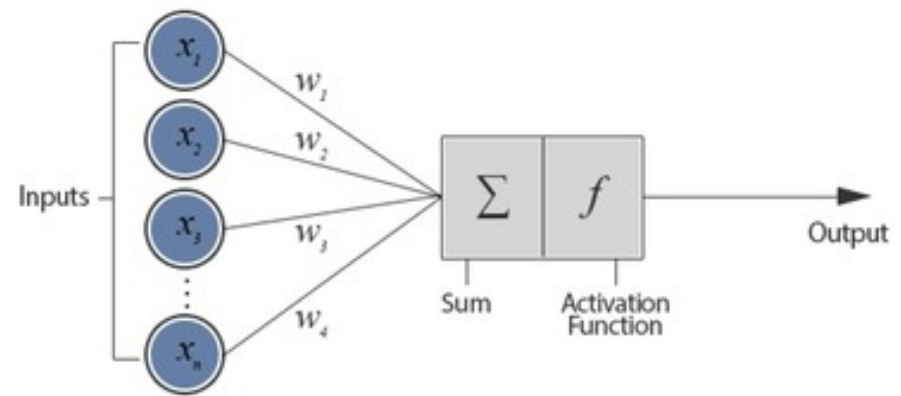
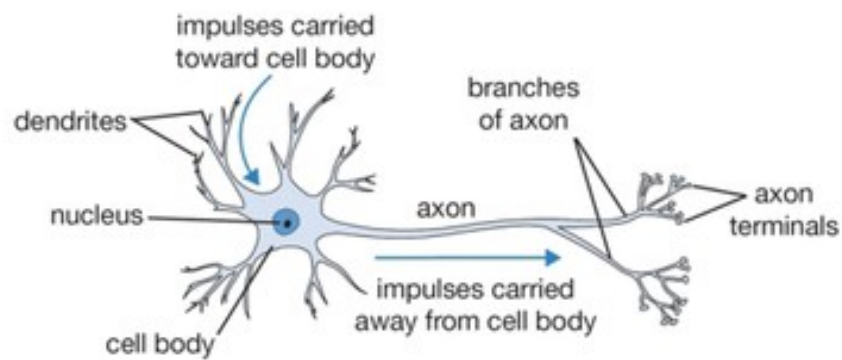
Neural Networks timeline

All ingredients to start another AI wave are there!!!

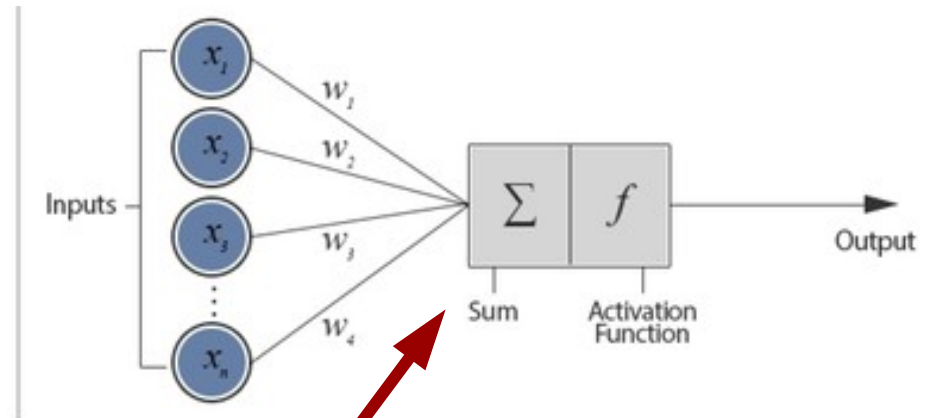
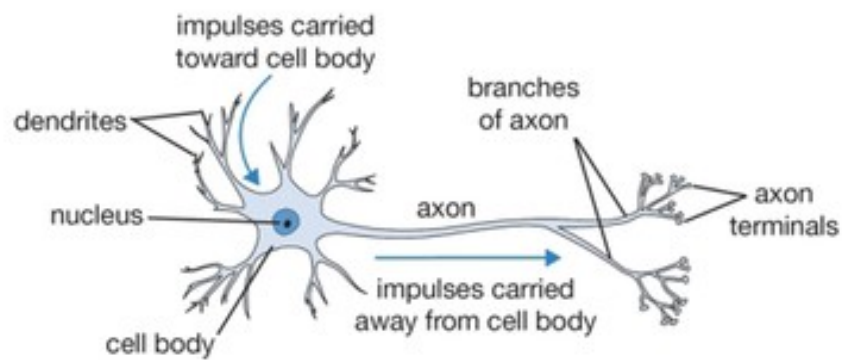


pervasive introduction of internet, smart devices, global IT corporations → **big data era**

Biological neurons vs ANN nodes

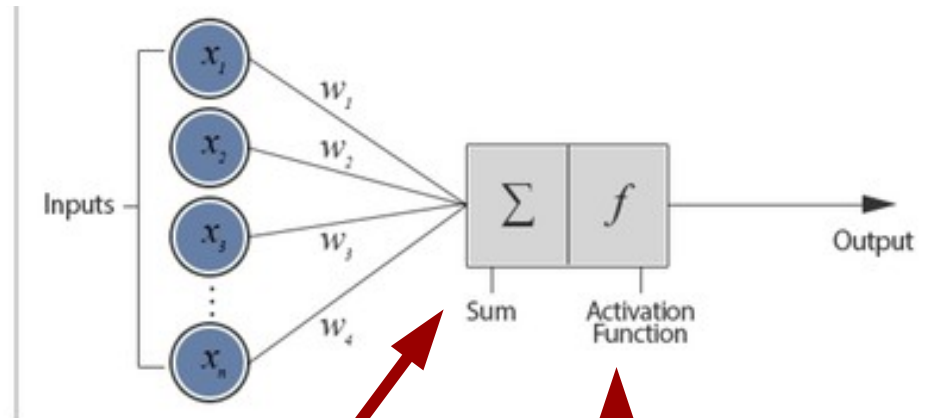
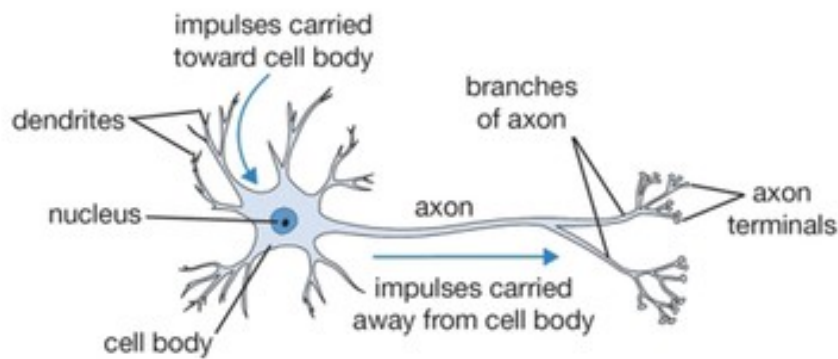


Biological neurons vs ANN nodes



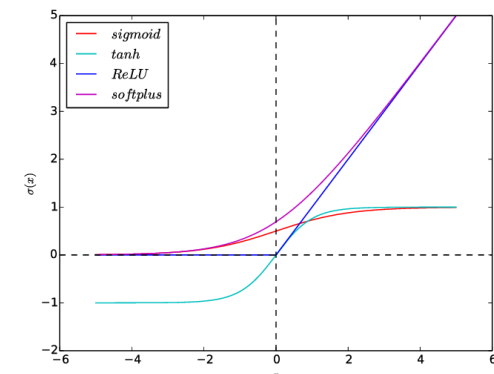
*weighted
accumulation*

Biological neurons vs ANN nodes

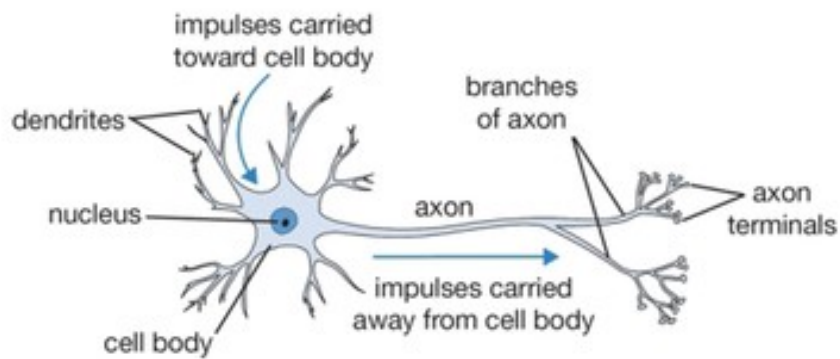


*weighted
accumulation*

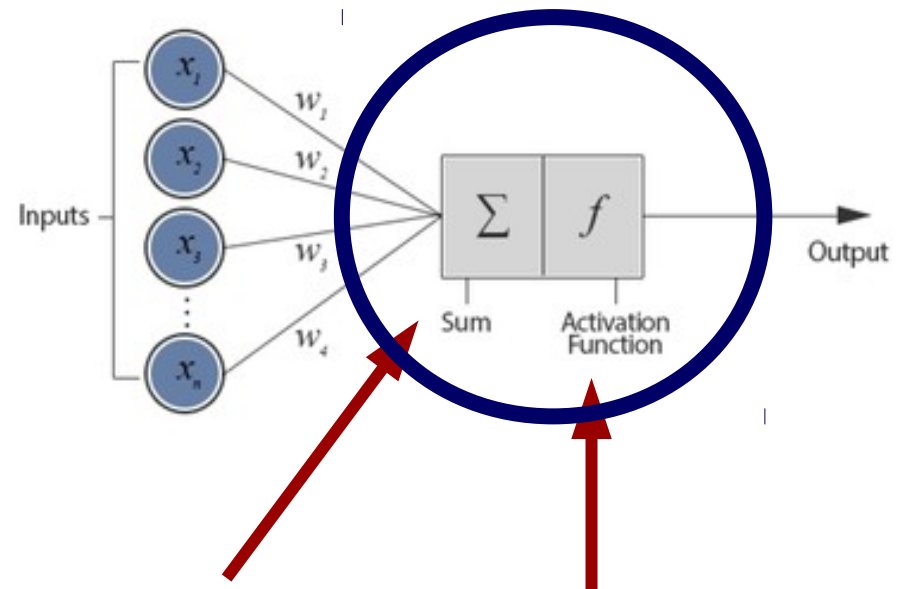
non-linearization



Biological neurons vs ANN nodes

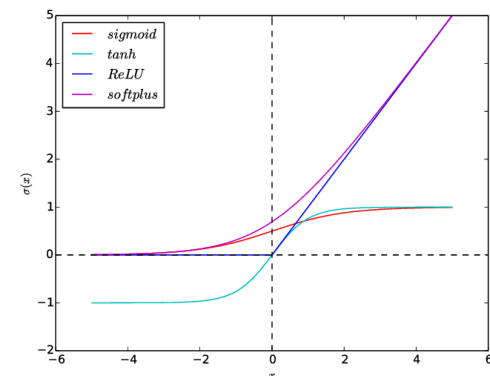


a sort of informational filter

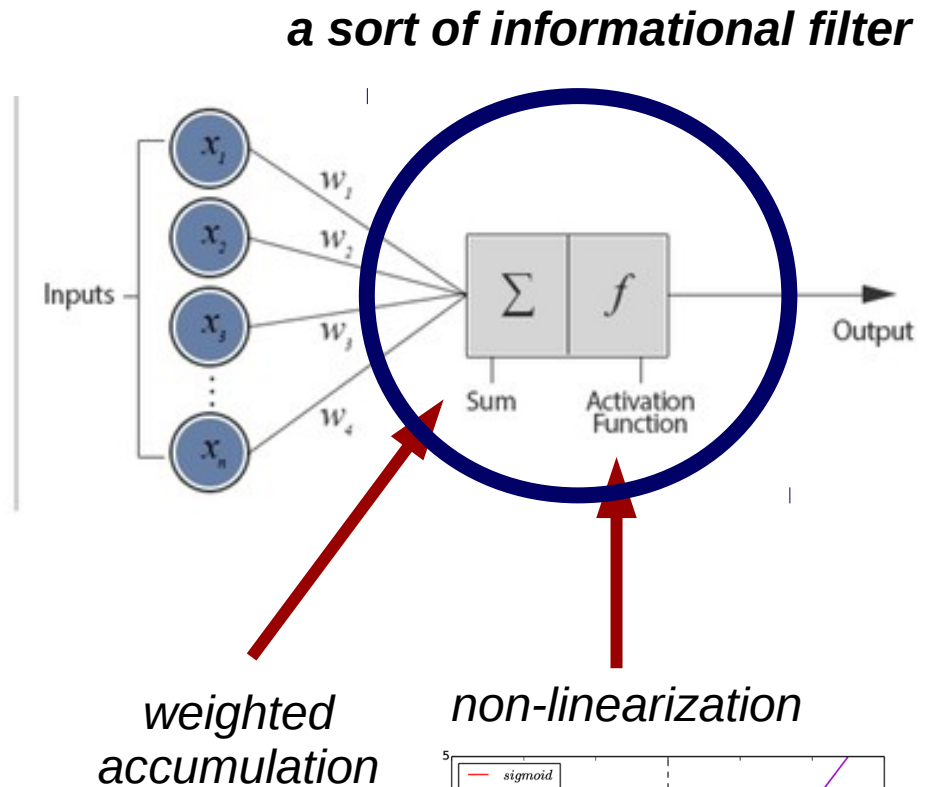
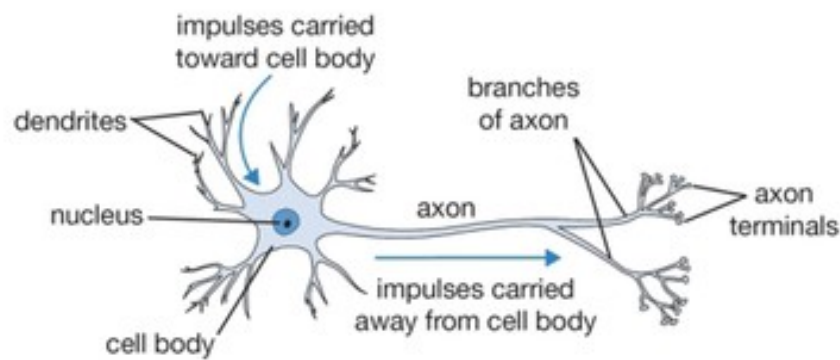


*weighted
accumulation*

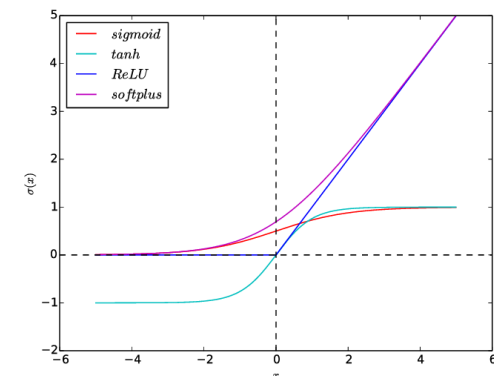
non-linearization



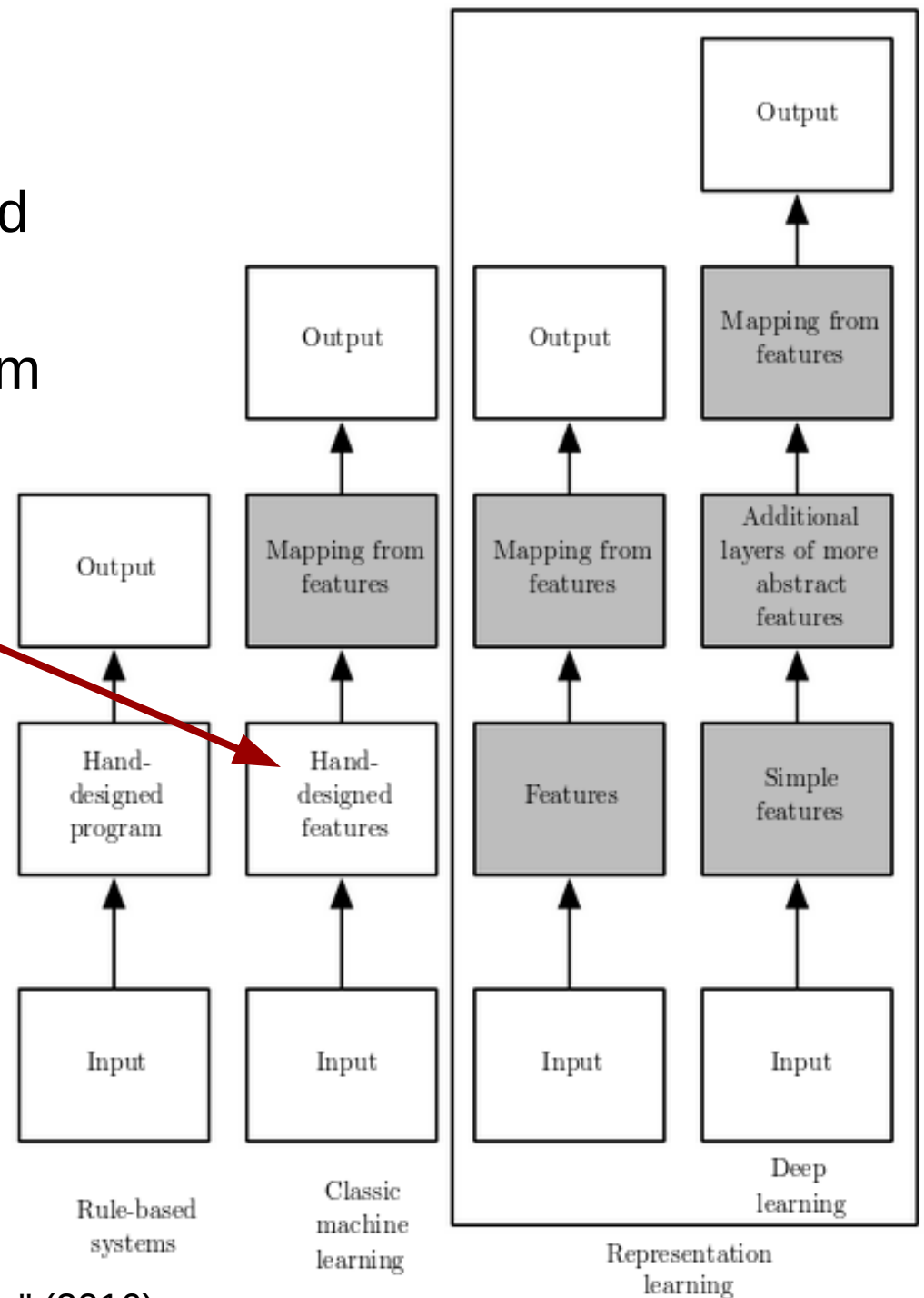
Biological neurons vs ANN nodes



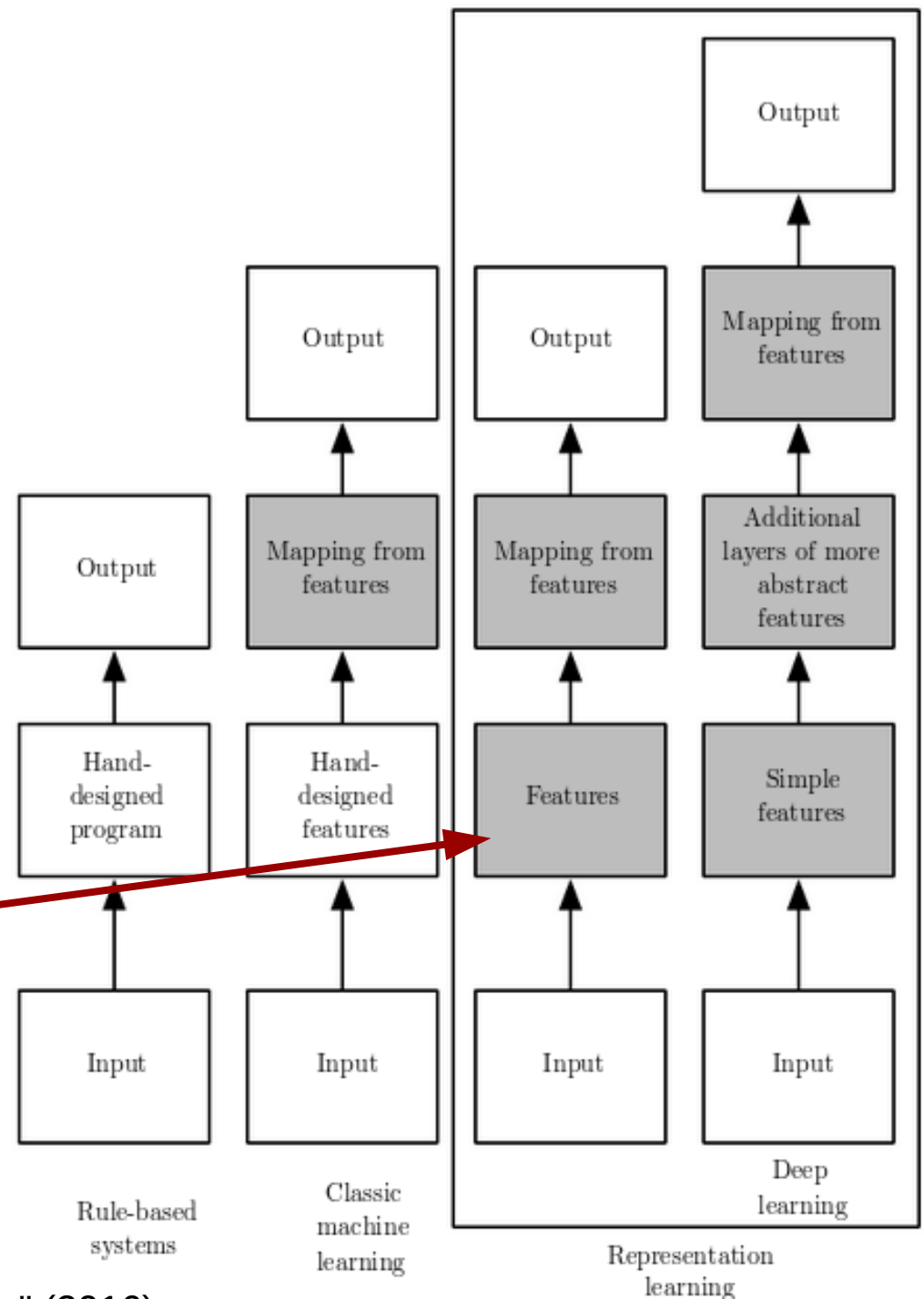
- A multi-layered artificial neural network is similar to a **cascade of filters**, that can be used to extract what is relevant and transform it adequately.



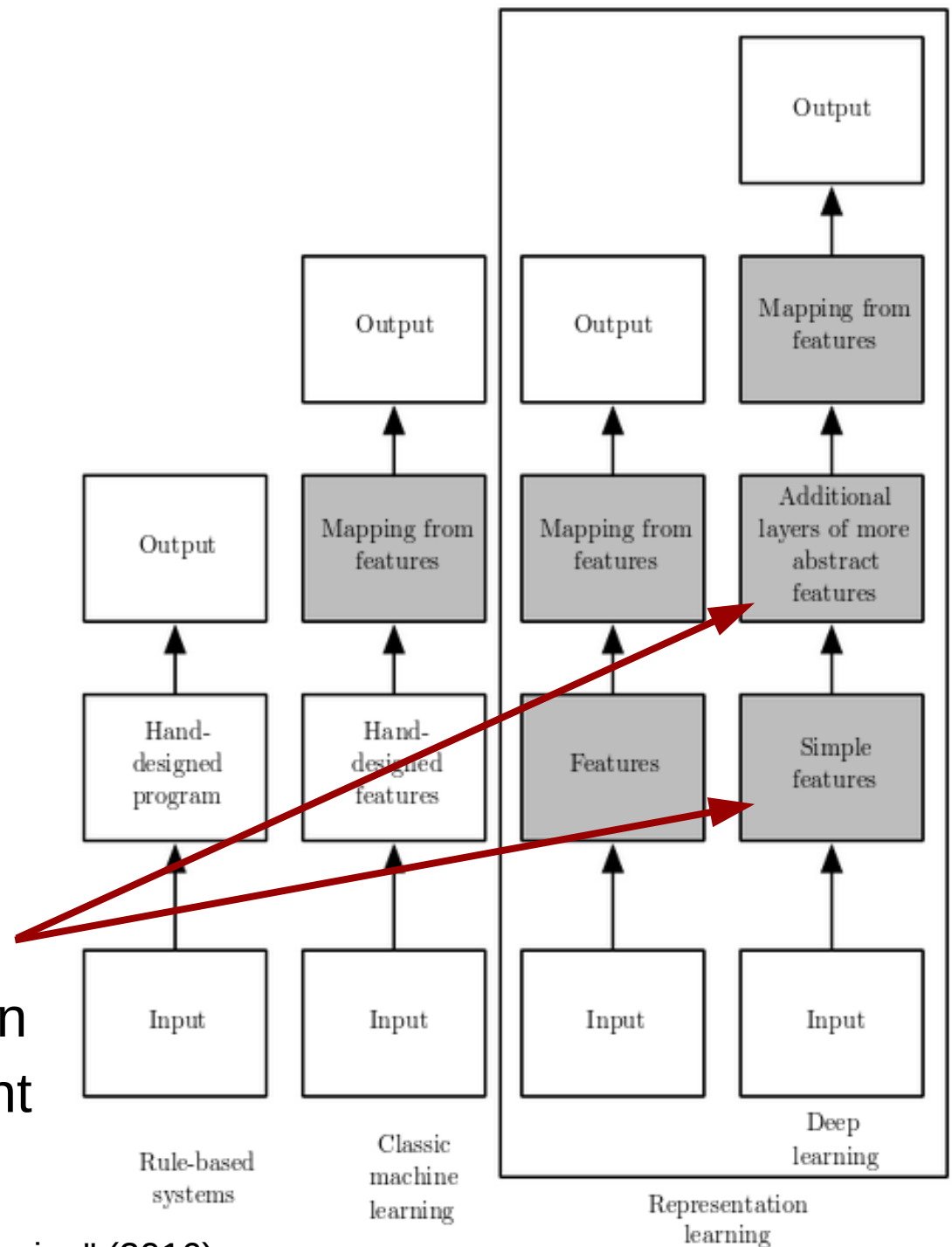
- To reduce data requirements, in classic ML features deemed to be relevant are manually selected by the developer from the available input.

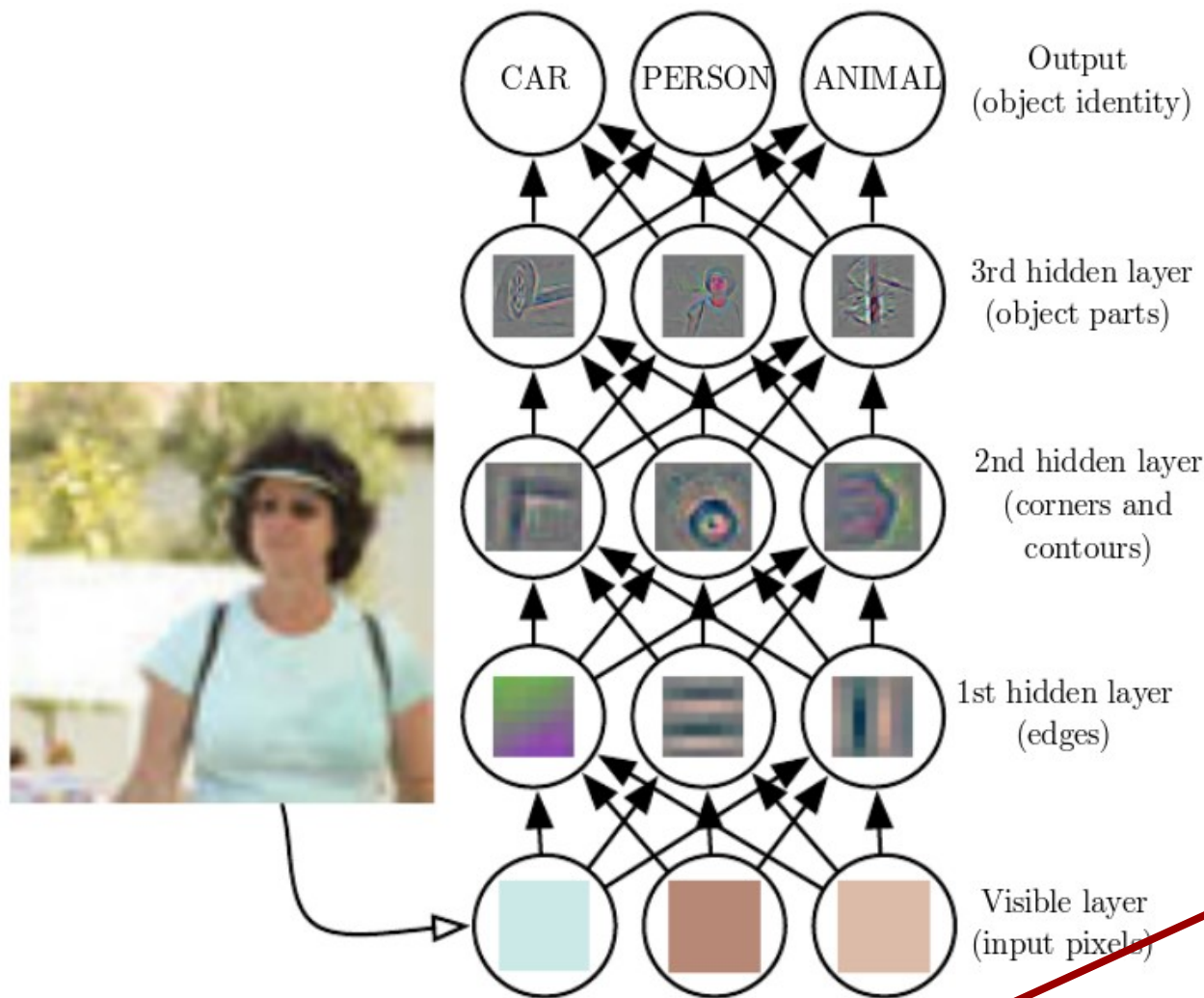


- When this is not possible, features have to be extracted as well, through some ***representation learning***.

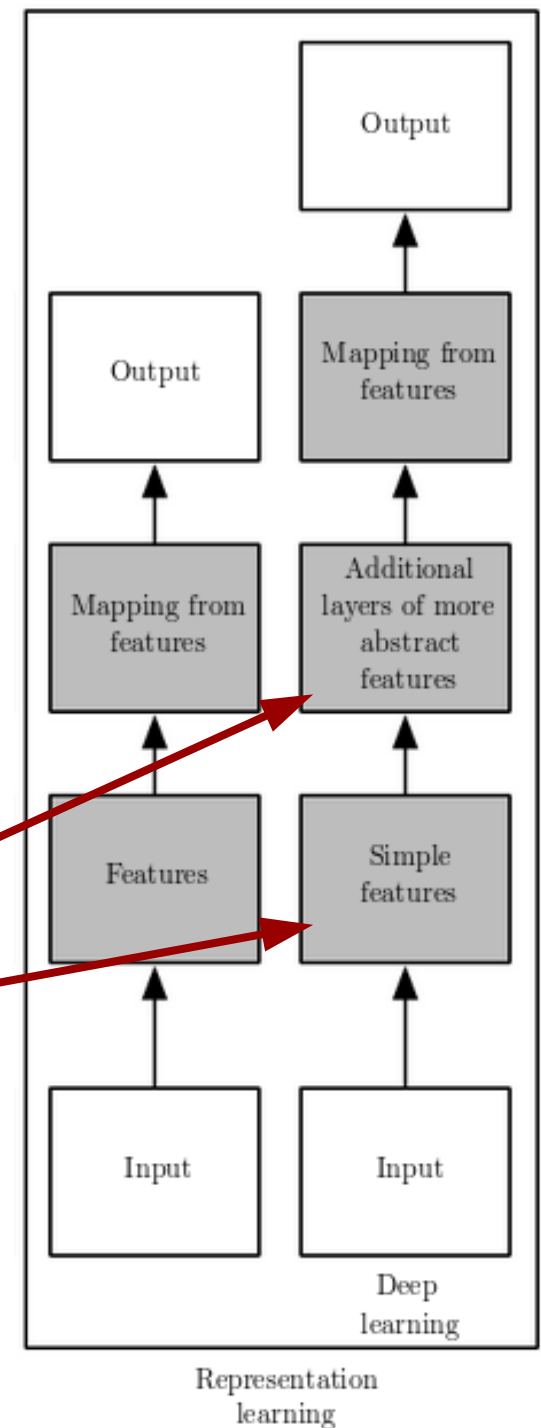


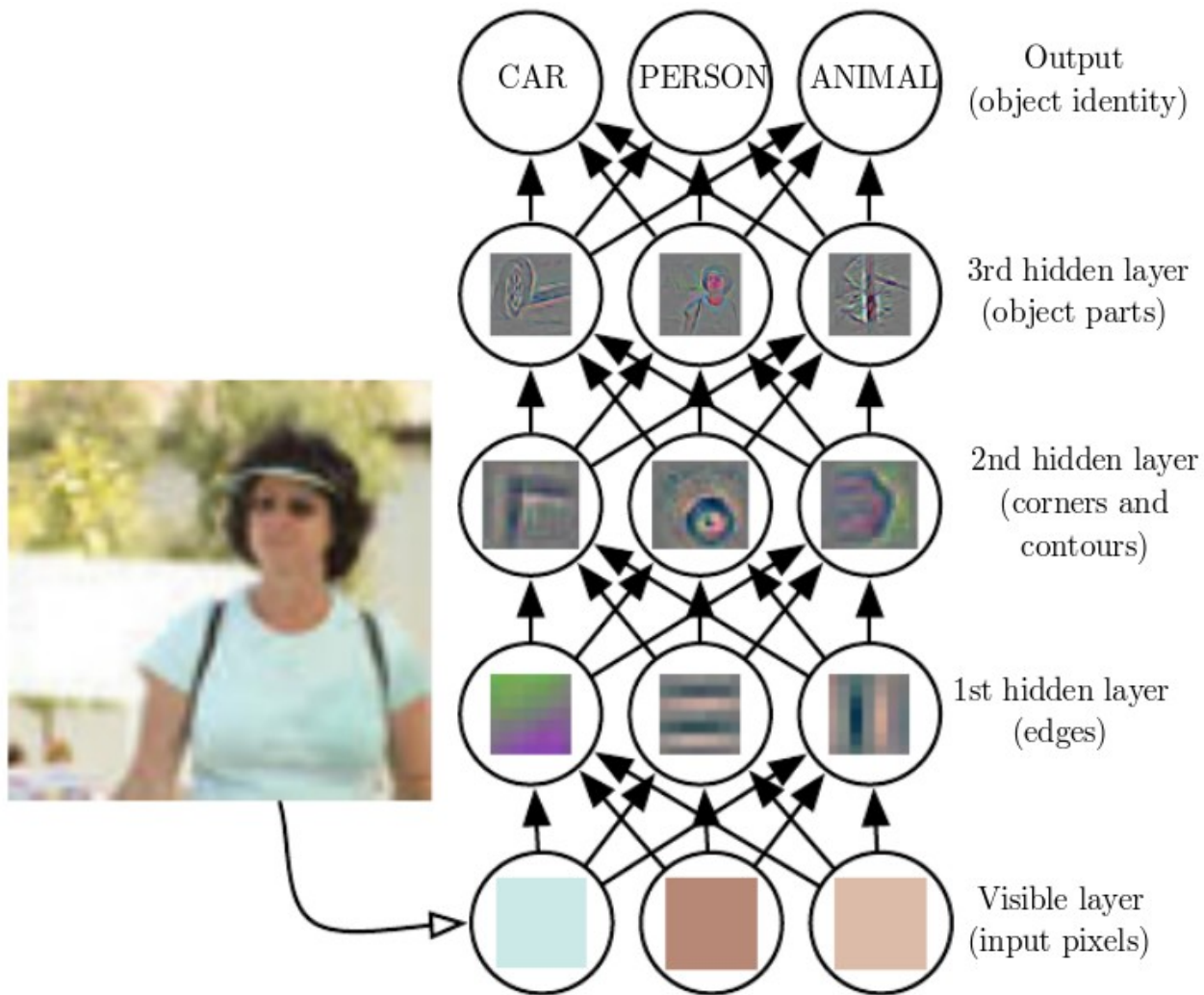
- **Deep learning** relies on a **hierarchy** of representation learning, producing different level of abstractions



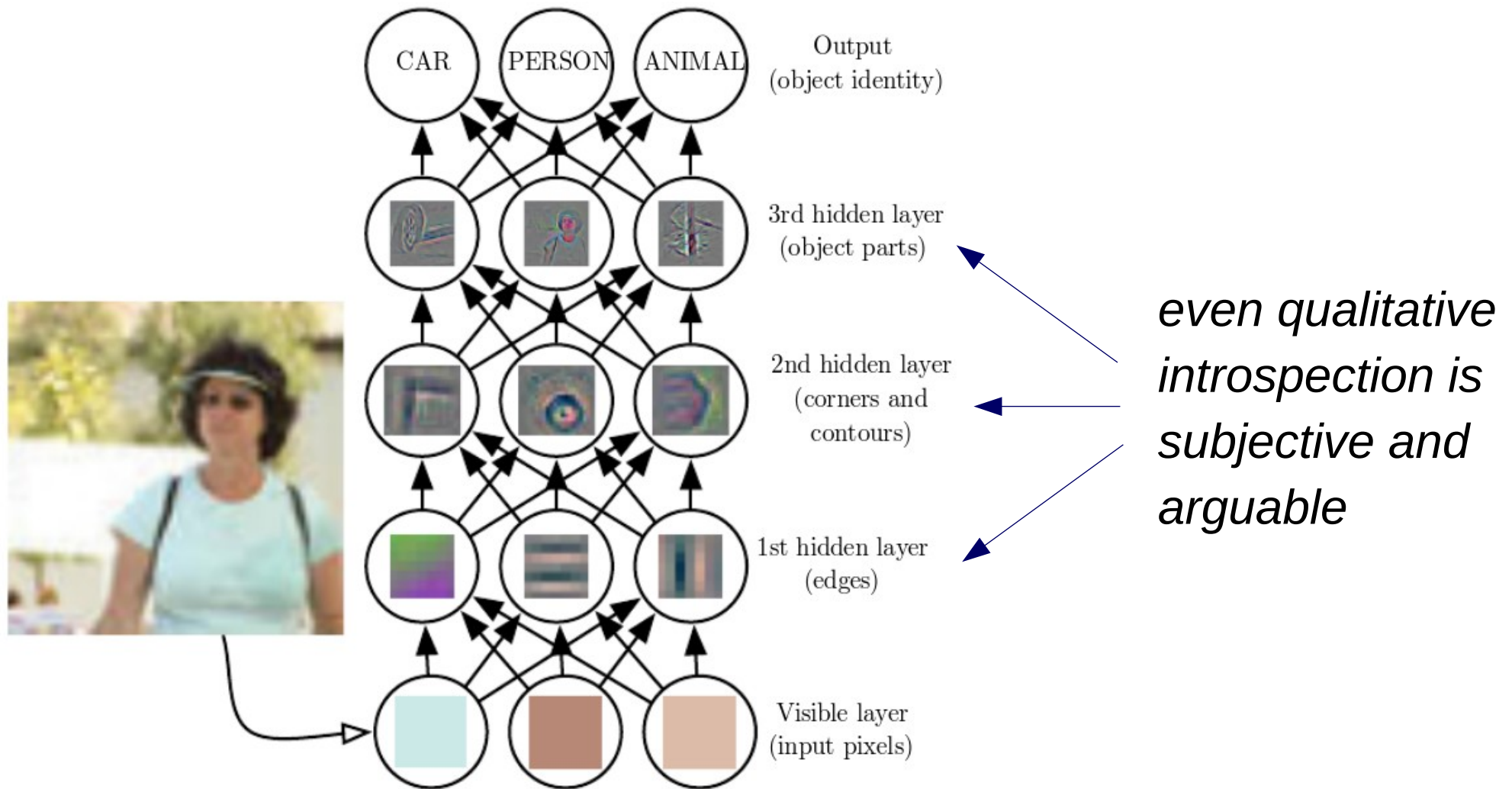


- **Deep learning** relies on a **hierarchy** of representation learning, producing different level of abstractions



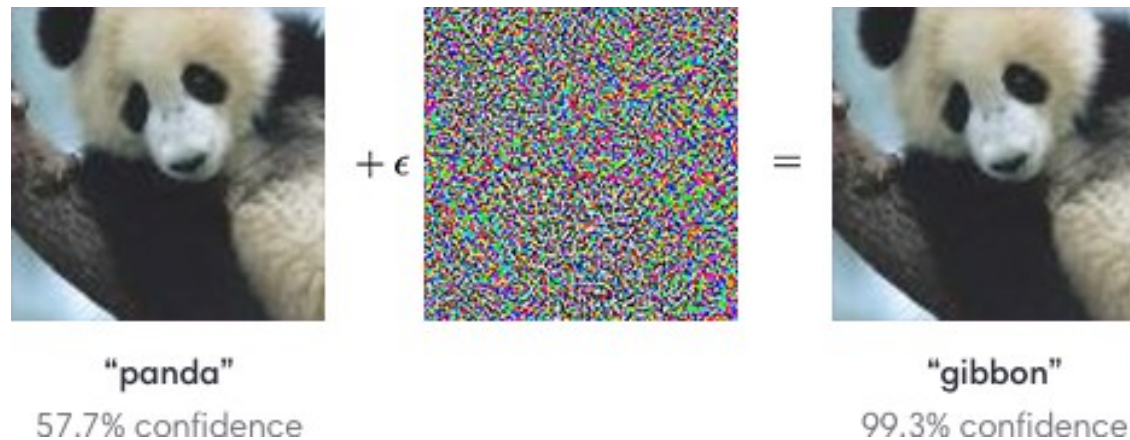


- **Problem:** the developer does not have direct control on which features are considered to be **relevant** to the task.



- **Problem:** the developer does not have direct control on which features are considered to be **relevant** to the task.

Adversarial attacks



- Knowing what is deemed of attention by the machine can be exploited by an attacker can produce targeted “optical illusions” for the machine, but not for us.

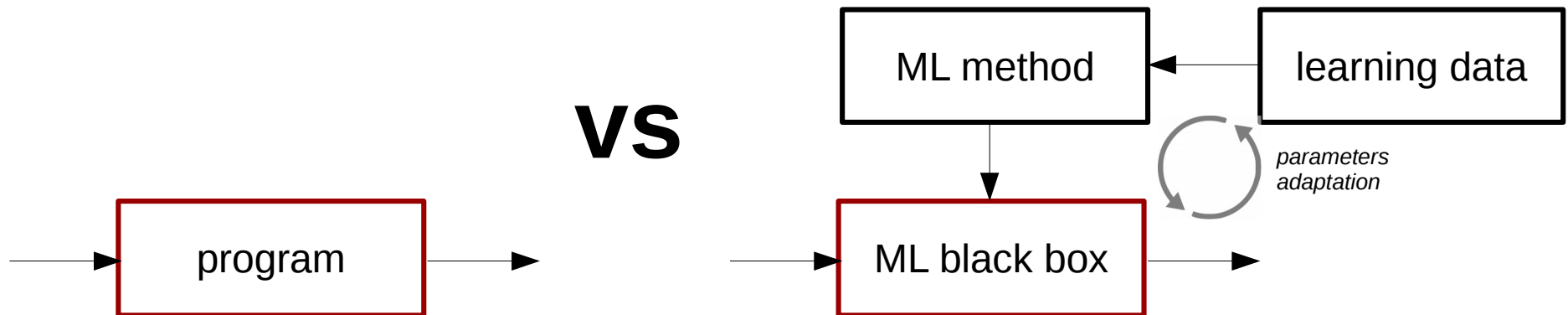
Using encoding/decoding abilities of deep learning

- On the other hand, knowing what is relevant to *our vision*, someone can play dirty tricks.



Face to face: <https://www.youtube.com/watch?v=ohmajJTcpNk>
Voice to lips: <https://www.youtube.com/watch?v=9Yq67CjDqvw>

From software/knowledge engineering to data engineering



- Clearly, the outcome of applying a ML method critically depends on the training data.

From software/knowledge engineering to data engineering



- Country A's army demands a classifier to recognize whether a tanks is from country A or country B. It provides the developers with a series of photos of tanks from both countries.

From software/knowledge engineering to data engineering

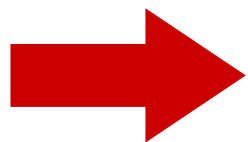


- Country A's army demands a classifier to recognize whether a tank is from country A or country B. It provides the developers with a series of photos of tanks from both countries.
- After the training, the developers investigate by introspection the activation patterns. They discover that “**daylight**” is a major factor supporting a B-tank classification. Returning on the source data, the developers discovered that there was *no photo of B-tanks at night*.

From software/knowledge engineering to data engineering



- Country A's army demands a classifier to recognize whether a tank is from country A or country B. It provides the developers with a series of photos of tanks from both countries.
- After the training, the developers investigate by introspection the activation patterns. They discover that “**daylight**” is a major factor supporting a B-tank classification. Returning on the source data, the developers discovered that there was *no photo of B-tanks at night*.

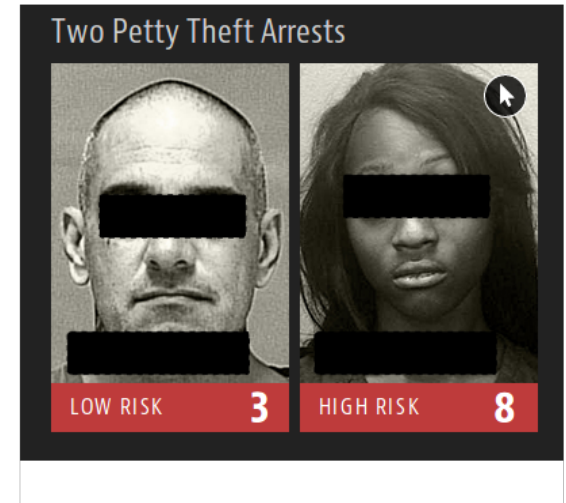


statistical biases endanger ML predictive abilities

On the “artificial prejudice”

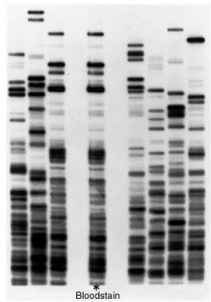
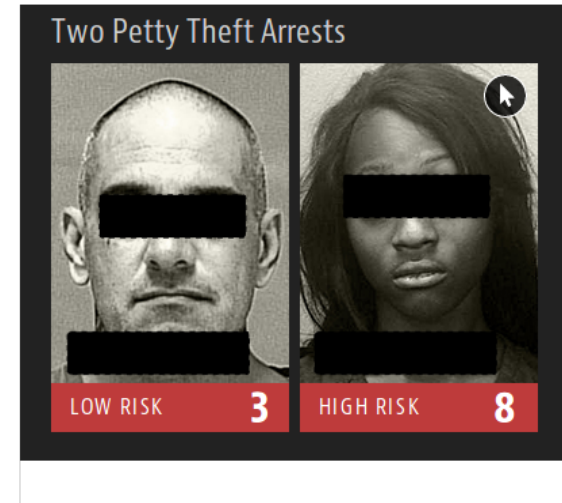
On the “artificial prejudice”

- Software used across the US predicting future crimes and criminals argued to be biased against African Americans (2016)

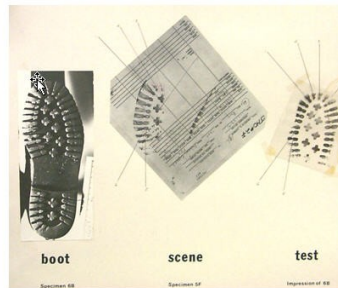


On the “artificial prejudice”

- Software used across the US predicting future crimes and criminals argued to be biased against African Americans (2016)
- **Core problem:** role of *circumstantial evidence*, how to integrate statistical inference in judgment?



DNA



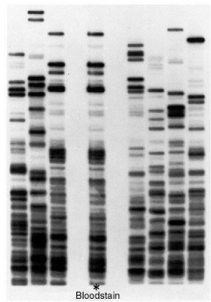
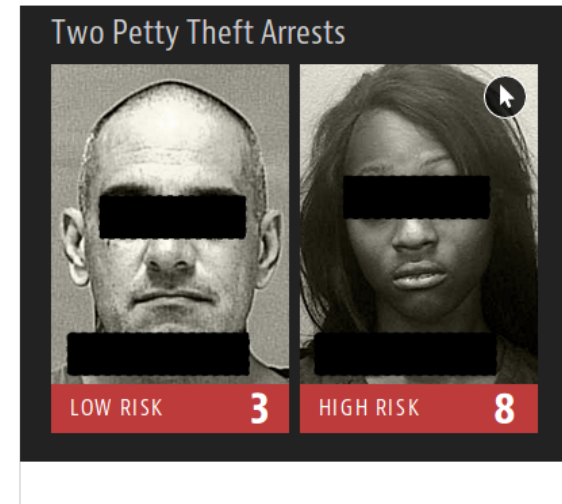
footwear

...

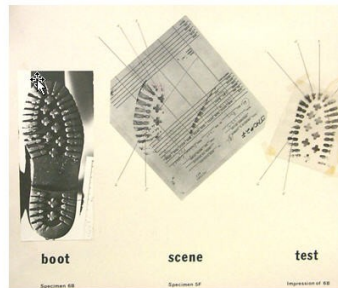
origin, gender,
ethnicity, wealth, ...

On the “artificial prejudice”

- Software used across the US predicting future crimes and criminals argued to be biased against African Americans (2016)
- **Core problem:** role of *circumstantial evidence*, how to integrate statistical inference in judgment?



DNA



footwear

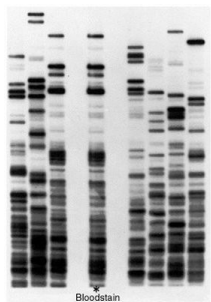
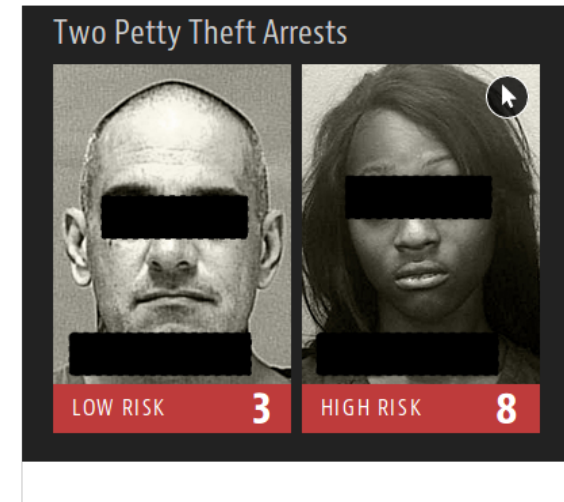
...

origin, gender,
ethnicity, wealth, ...

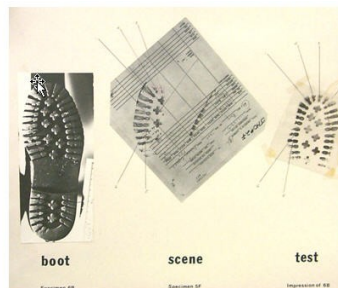
*improper
profiling?*

On the “artificial prejudice”

- Software used across the US predicting future crimes and criminals argued to be biased against African Americans (2016)
- **Core problem:** role of *circumstantial evidence*, how to integrate statistical inference in judgment?



DNA



footwear

*improper
because it causes
unfair judgment*

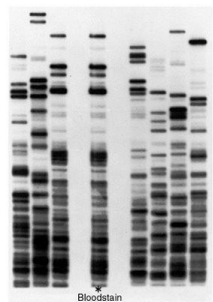
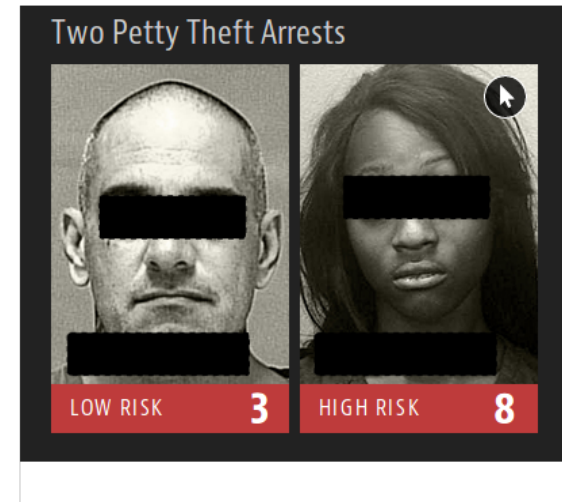
...

origin, gender,
ethnicity, wealth, ...

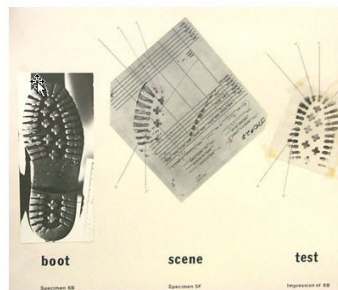


On the “artificial prejudice”

- Software used across the US predicting future crimes and criminals argued to be biased against African Americans (2016)
- **Core problem:** role of *circumstantial evidence*, how to integrate statistical inference in judgment?



DNA



footwear

...

origin, gender,
ethnicity, wealth, ...

*improper
because it causes
unfair judgment*



Norms determine which factors are acceptable or not.

The present

- By using a mixture of ML techniques, several human or super-human performances are achieved every year in specific tasks (*mostly by corporation-driven research*).

The AI index publishes reports on these records: <https://aiindex.org>

The present

- By using a mixture of ML techniques, several human or super-human performances are achieved every year in specific tasks (*mostly by corporation-driven research*).



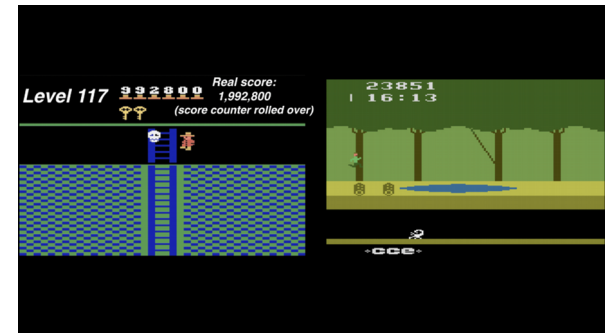
Google DeepMind (2016)

Chinese - English



中文 - 英文

Microsoft (2018)



Uber (Feb 2019)

The AI index publishes reports on these records: <https://aiindex.org>

The present

- By using a mixture of ML techniques, several human or super-human performances are achieved every year in specific tasks (*mostly by corporation-driven research*).



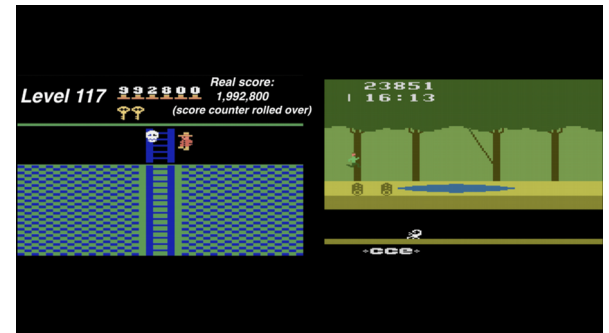
Google DeepMind (2016)

Chinese - English



中文 - 英文

Microsoft (2018)



Uber (Feb 2019)

- But the problems of **generalization**, **explainability**, **transparency**, **responsibility**, **fairness**, etc. are still there.

The AI index publishes reports on these records: <https://aiindex.org>

The present

- New research trends are emerging to face these issues, trying a variety of different approaches.



The present

- New research trends are emerging to face these issues, trying a variety of different approaches.



- Still unclear which one will achieve the intent.

Prospective trajectories

Refocus on interaction

- **Intelligence** can be rephrased in terms of adequate performance within a certain *interactional niche*:



Refocus on interaction

- **Intelligence** can be rephrased in terms of adequate performance within a certain *interactional niche*:

i.e. the ability of one agent:

contextualization

- to select or create a ***script*** that can be **ascribed** to to the other agent (including the environment)



Refocus on interaction

- **Intelligence** can be rephrased in terms of adequate performance within a certain *interactional niche*:

i.e. the ability of one agent:

- to select or create a ***script*** that can be **ascribed** to the other agent (including the environment)
- to select or create a ***script*** that **drives** rewarding *interactions* with the other agent

contextualization

fitting to
given context



Challenges

- Today, AI and decision-making capitalize too much on *optimization* working on the “**fitting to given context**” phase.

Challenges

- Today, AI and decision-making capitalize too much on *optimization* working on the “**fitting to given context**” phase.
- But the “**contextualization**” phase is particularly problematic w.r.t. the **social environment**, for its high variability.

Challenges

- Today, AI and decision-making capitalize too much on *optimization* working on the “**fitting to given context**” phase.
- But the “**contextualization**” phase is particularly problematic w.r.t. the **social environment**, for its high variability.
 - e.g. the evolution of *dress codes* along history

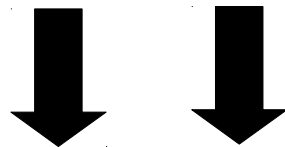


Challenges

- Today, AI and decision-making capitalize too much on *optimization* working on the “**fitting to given context**” phase.
- But the “**contextualization**” phase is particularly problematic w.r.t. the **social environment**, for its high variability.
- The social structure adds upon the physical structure in *indicating* and then *establishing* rewards to agents, via
 - **explicit norms**
 - informal and tacit norms: **social practices**

Challenges

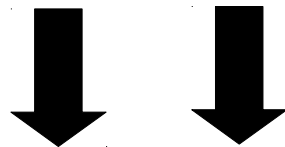
- Today, AI and decision-making capitalize too much on *optimization* working on the “**fitting to given context**” phase.
- But the “**contextualization**” phase is particularly problematic w.r.t. the **social environment**, for its high variability.
- The social structure adds upon the physical structure in *indicating* and then *establishing* rewards to agents, via
 - **explicit norms**
 - informal and tacit norms: **social practices**



Norms are crucial for intelligent (social) behaviour

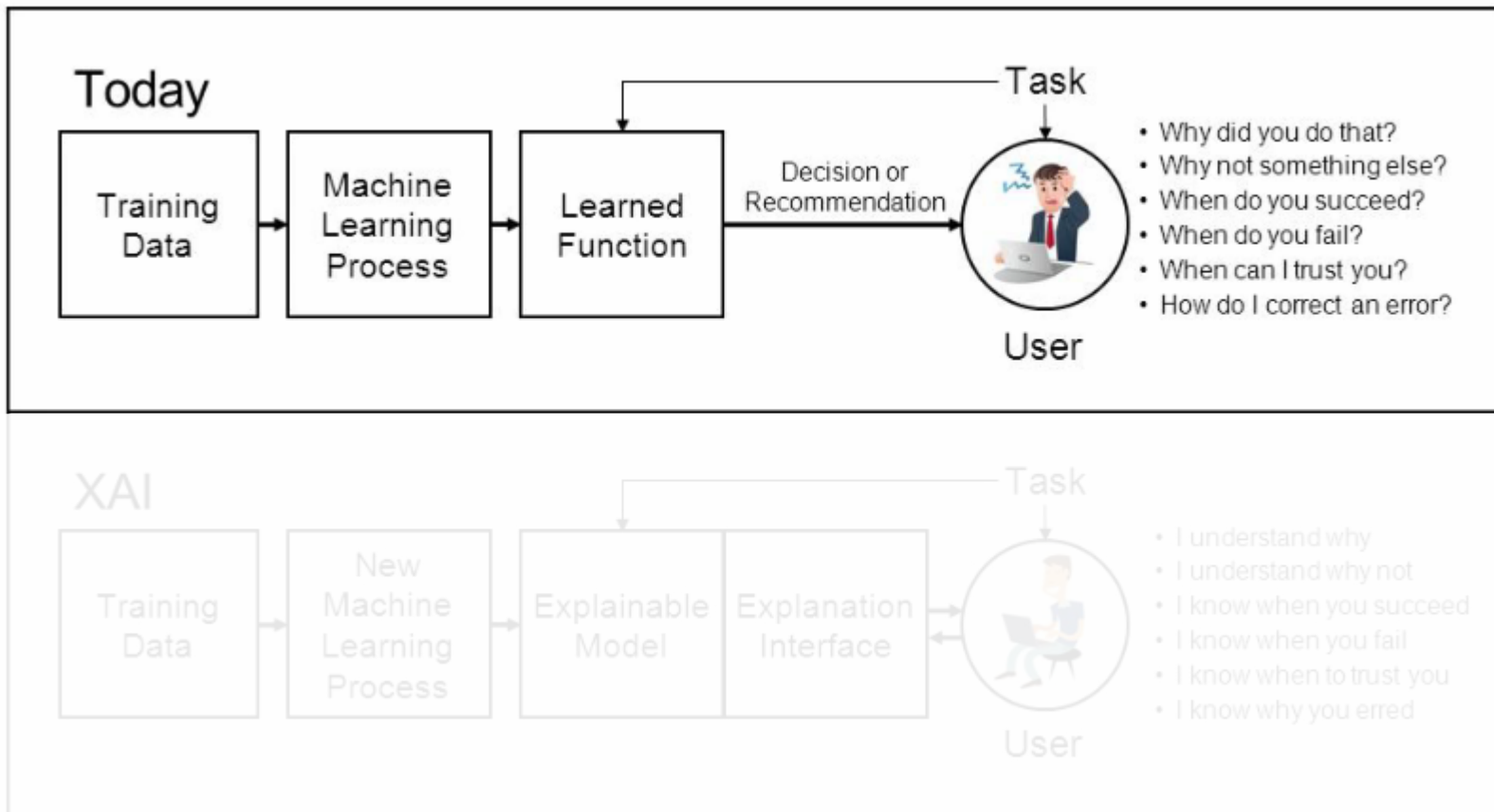
Challenges

- Today, AI and decision-making capitalize too much on *optimization* working on the “**fitting to given context**” phase.
- But the “**contextualization**” phase is particularly problematic w.r.t. the **social environment**, for its high variability.
- The social structure adds upon the physical structure in *indicating* and then *establishing* rewards to agents, via
 - **explicit norms**
 - informal and tacit norms: **social practices**



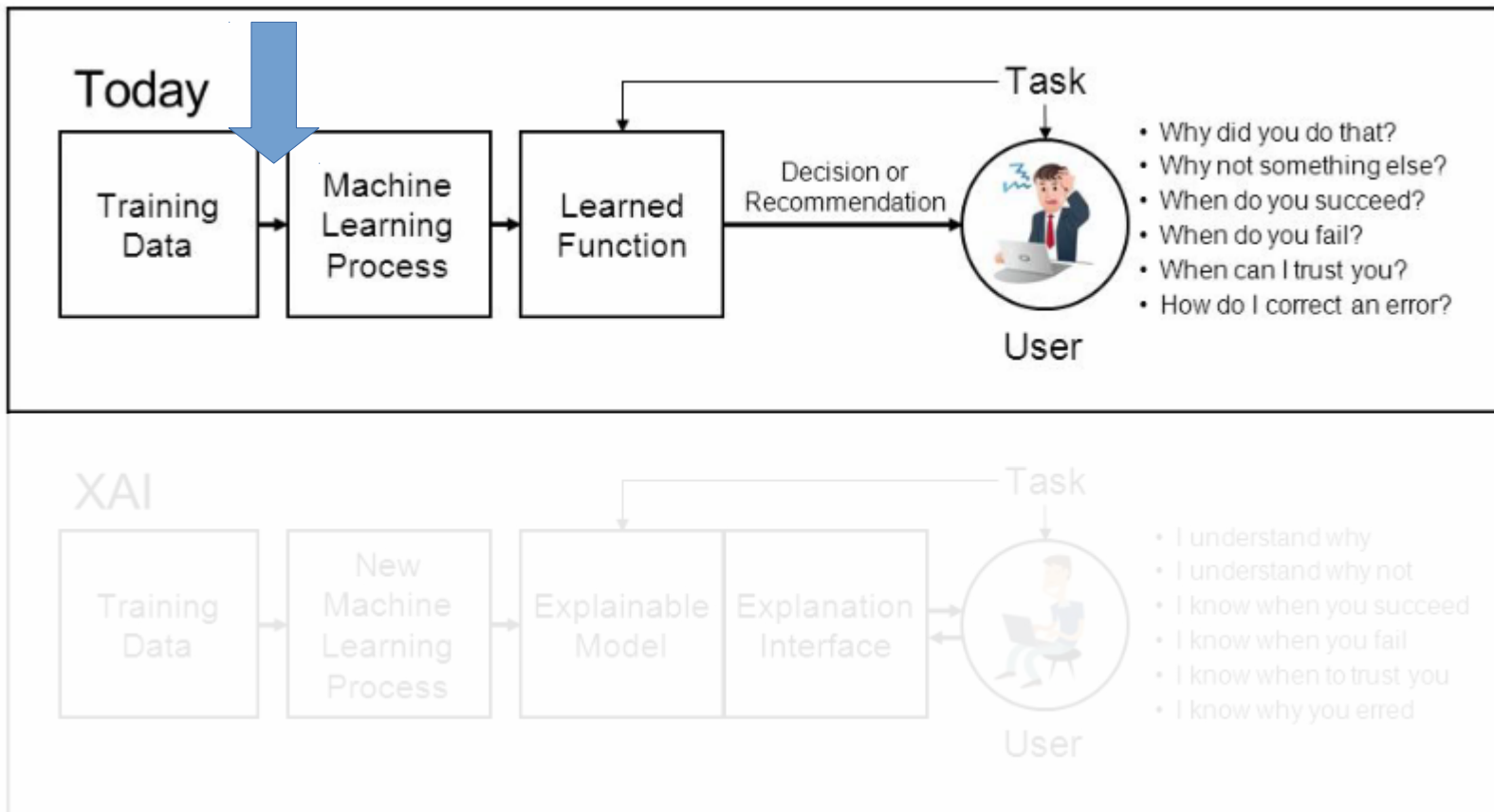
Norms are crucial for intelligent (social) behaviour

The call for *Explanable AI* (XAI)



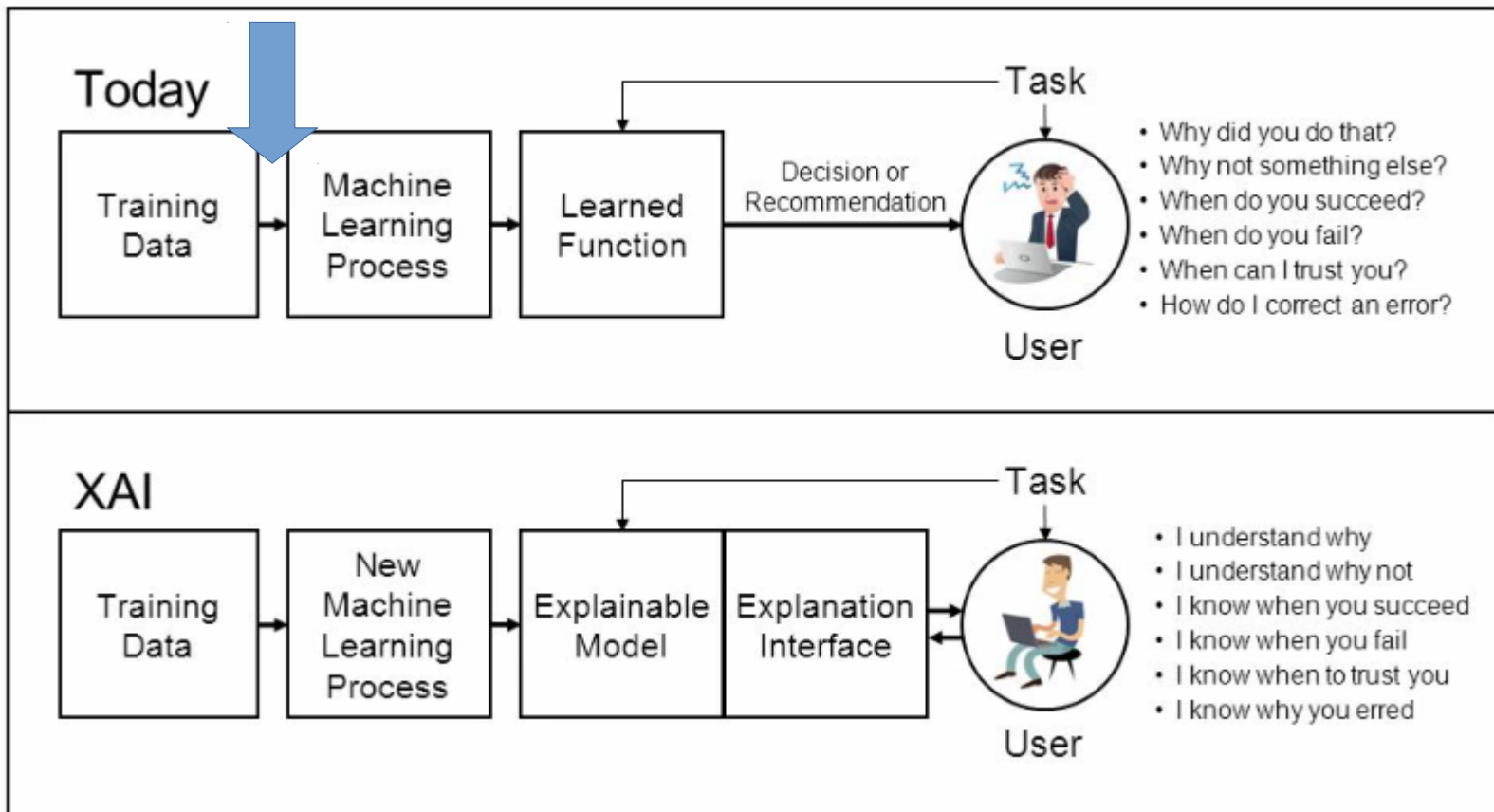
The call for *Explanable AI* (XAI)

***statistical
alignment***



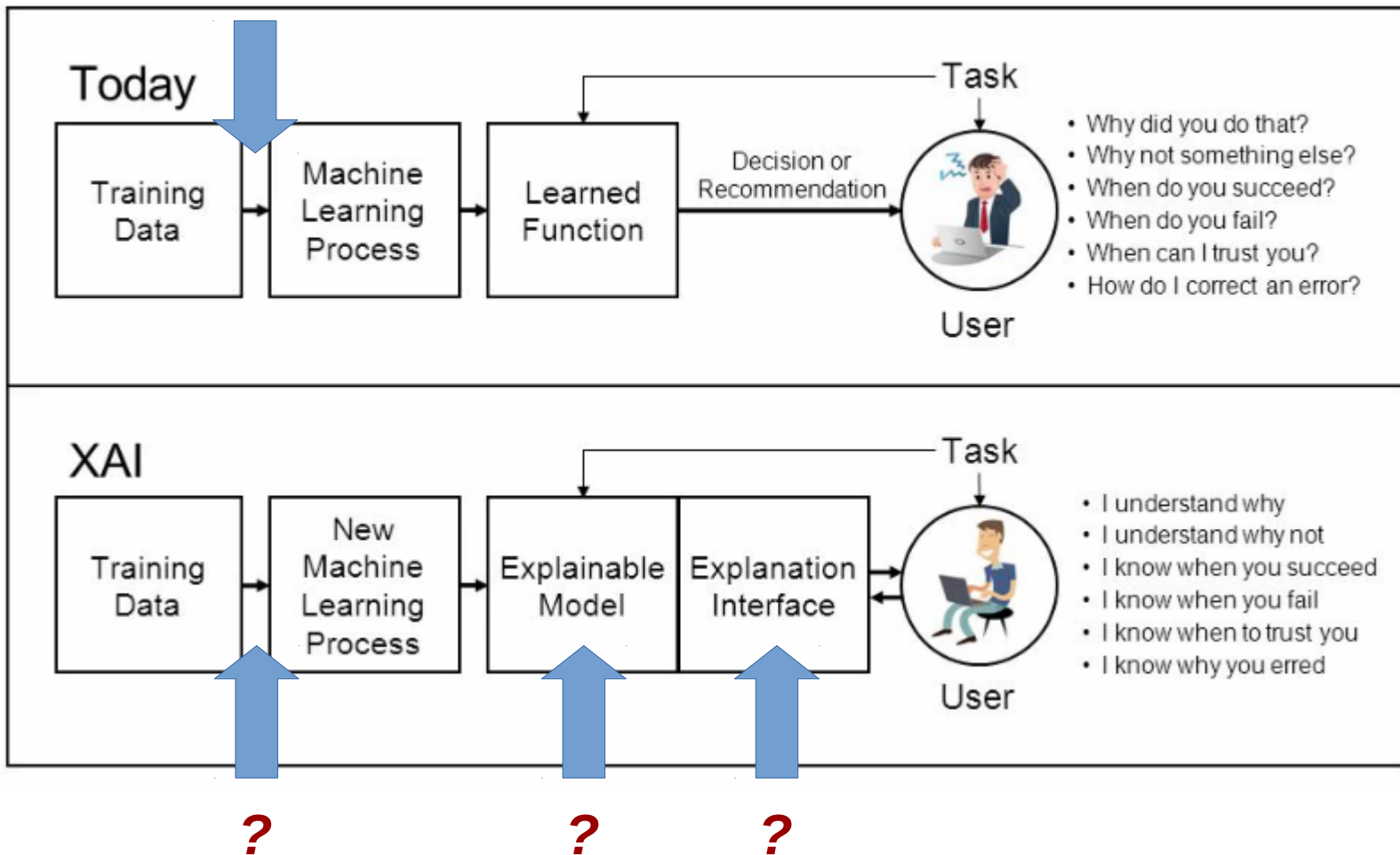
The call for *Explanable AI* (XAI)

*statistical
alignment*



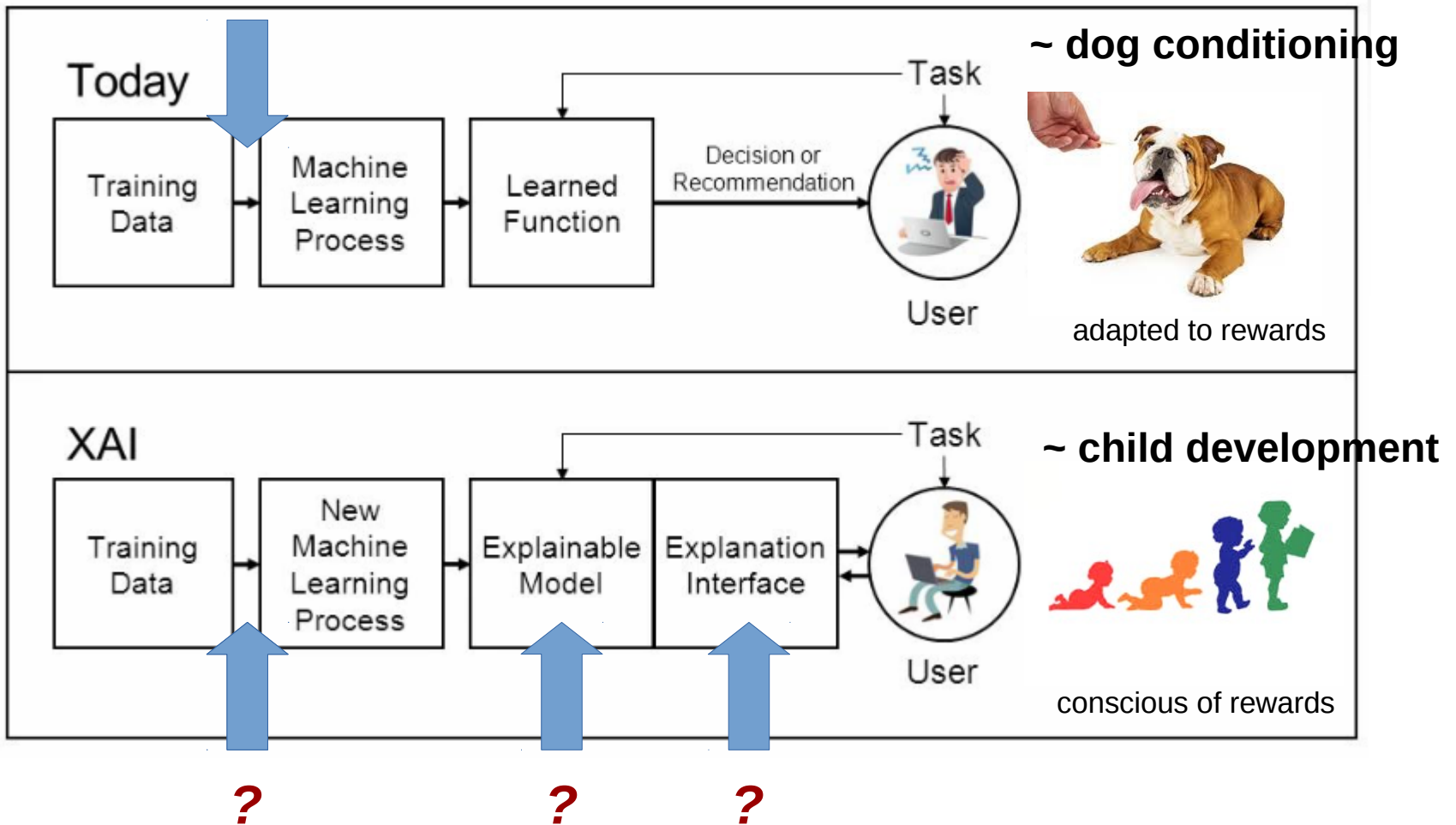
The call for *Explanable AI* (XAI)

*statistical
alignment*



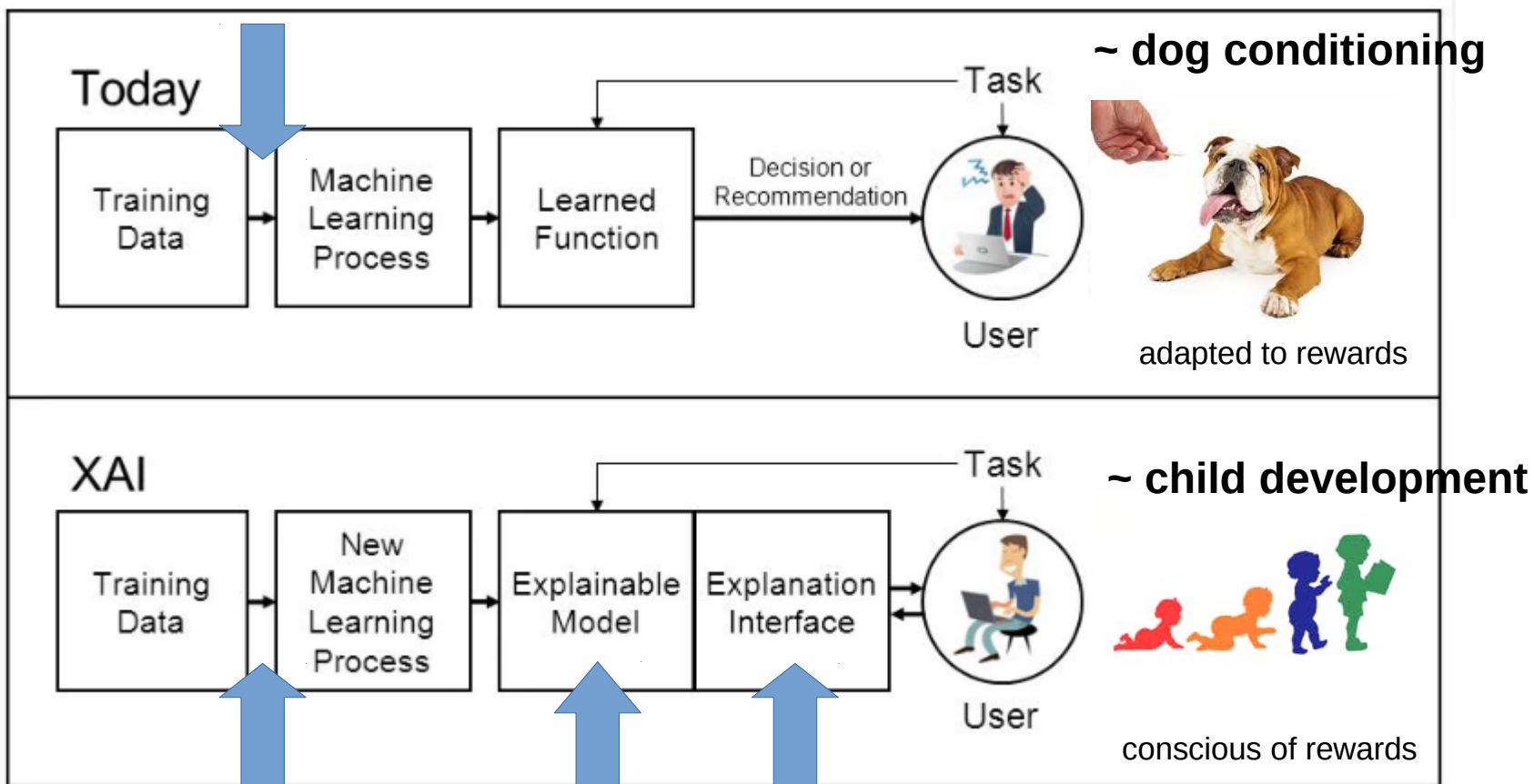
The call for *Explanaible AI* (XAI)

*statistical
alignment*



The call for *Explainable AI* (XAI)

**statistical
alignment**



grounding
experiential
(direct)

conceptualizing

communicating

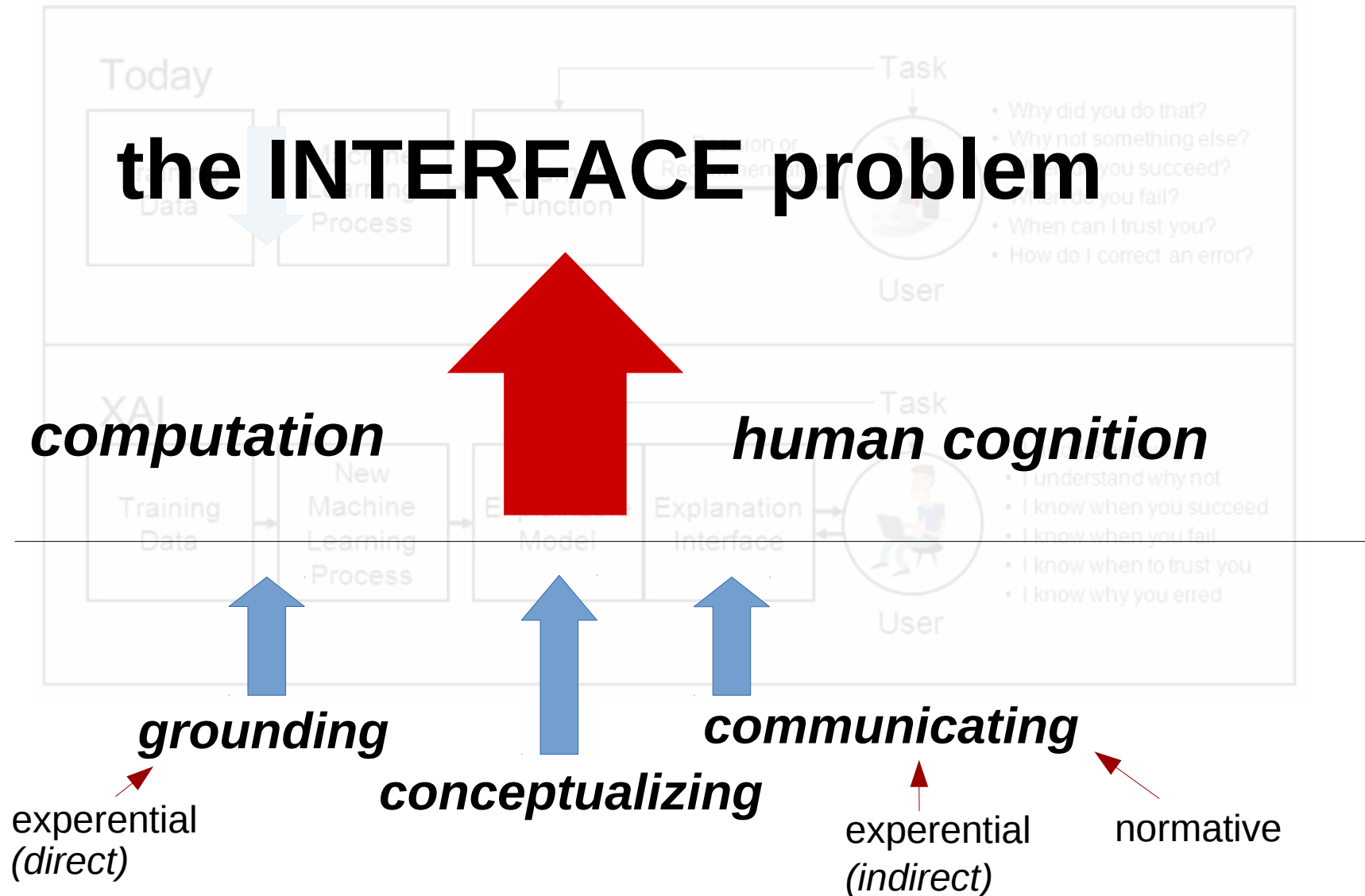
experiential
(indirect)

normative

The call for *Explainable AI* (XAI)

*statistical
alignment*

the INTERFACE problem

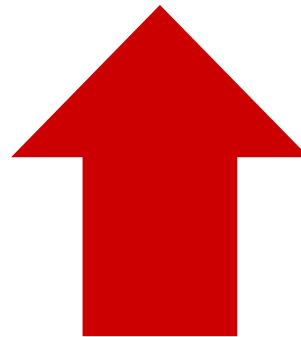


Possible approaches

- *bottom-up*: use statistical ML to recreate functions *mimicking* to some extent human cognition
- *top-down*: conceive algorithms *reproducing by design* functions observable in human cognition

the INTERFACE problem

computation



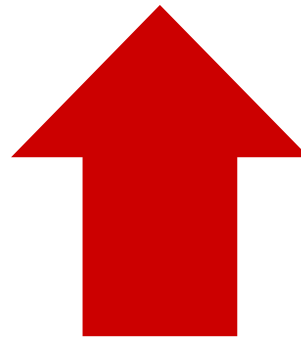
human cognition

Possible approaches

- *bottom-up*: use statistical ML to recreate functions *mimicking* to some extent human cognition
 - *top-down*: conceive algorithms *reproducing by design* functions observable in human cognition
- only here we have control on what we want to reproduce*

the INTERFACE problem

computation



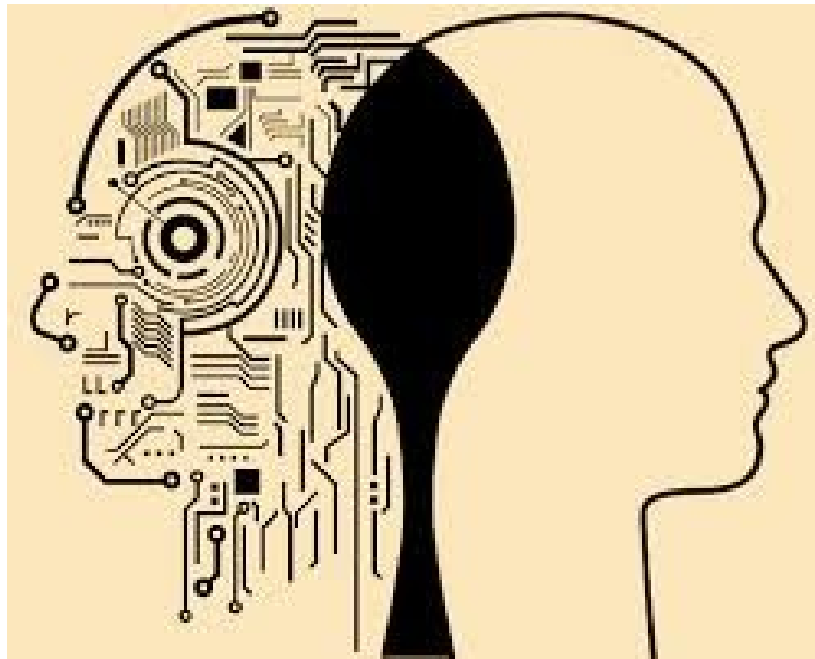
human cognition

Will cognitive architectures be
the third AI wave?

Conclusions

No AGI in view

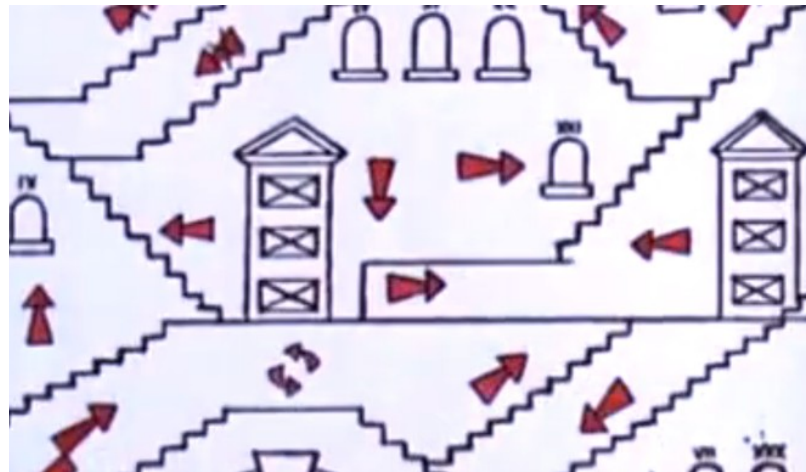
- I believe (with many others) that crucial pieces are still missing to embed *general intelligence* into a single artificial device.



- These pieces might be simple or not, it's the ML method that is not satisfactory to design them.

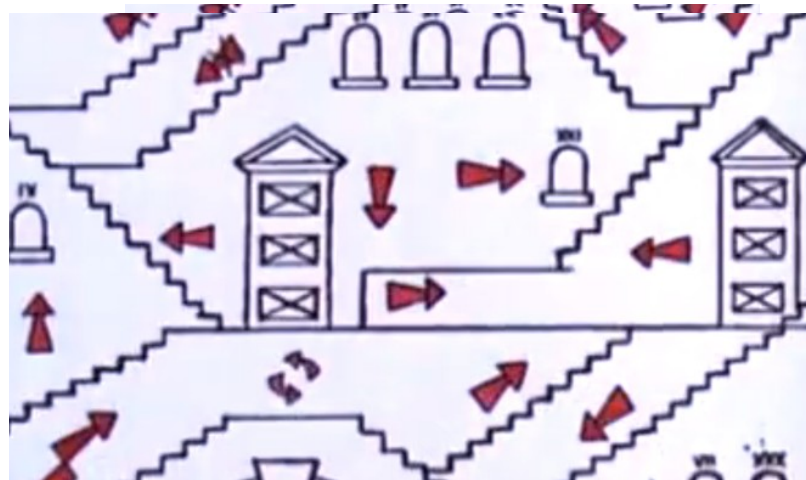
Rise of *artificially dumber* systems

- However, already today, the introduction of ubiquitous *cyber-physical connections* in all human activities raises serious concerns at societal and at cognitive level.
 - high risks to be entangled in **artificially dumber** systems.



Rise of *artificially dumber* systems

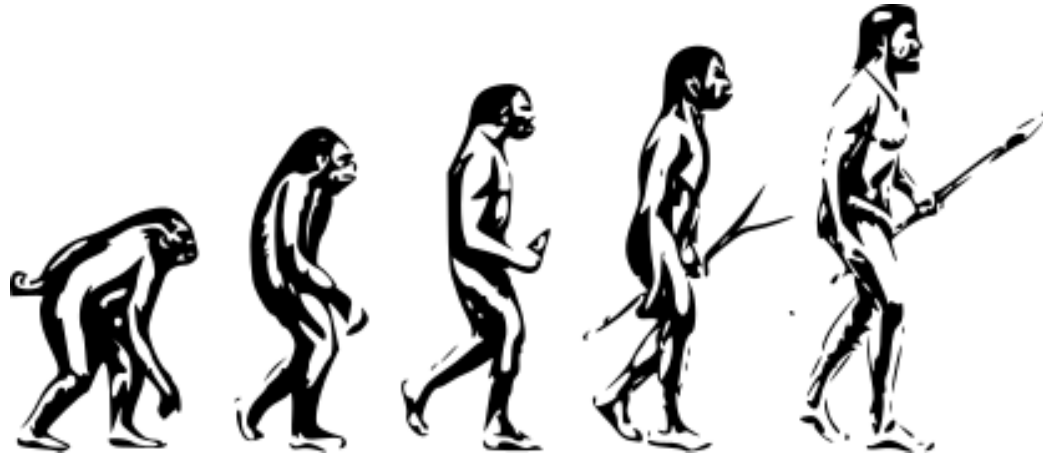
- However, already today, the introduction of ubiquitous *cyber-physical connections* in all human activities raises serious concerns at societal and at cognitive level.
 - high risks to be entangled in **artificially dumber** systems.



- The potential impact is too critical to be belittled for the belief in technologically-driven '*magnificent and progressive fate*'.

AI as an extension to humans

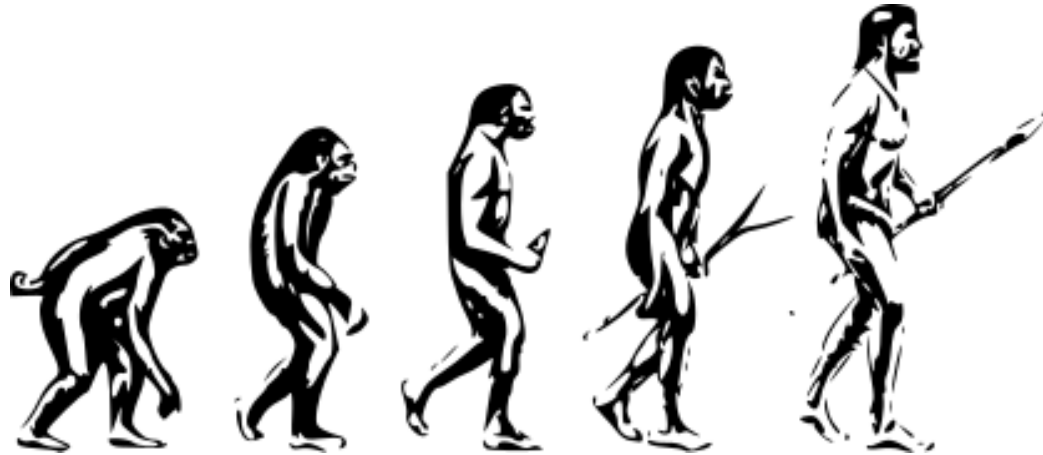
- Humans, as species, evolved *being shaped by their tools*.



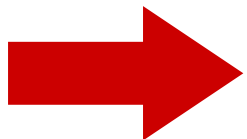
- We should look at our tools not as *means*, but as **forces that determine** not only **our societies**, but also **our very existence**.

AI as an extension to humans

- Humans, as species, evolved *being shaped by their tools*.



- We should look at our tools not as *means*, but as **forces that determine** not only **our societies**, but also **our very existence**.



If we want to decide upon our existence, then we have also to decide upon our tools.