

An Introduction to Artificial Intelligence

MASS Rotterdam – Conference on *The Legal Future of Autonomous Vessels*
Erasmus School of Law – 19 June 2025

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?



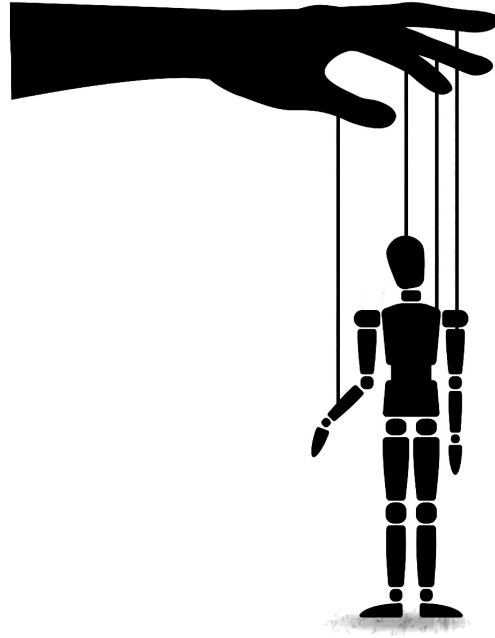
What is ~~Artificial~~ Intelligence?

- 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

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**

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?**

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?**

it depends on what we mean by “intelligence”...

Categories of AIs

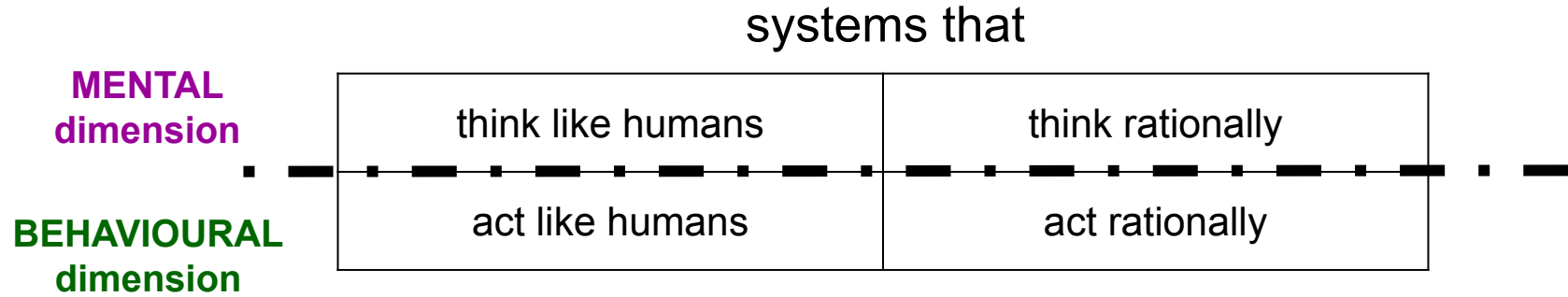
systems that

think like humans	think rationally
act like humans	act rationally

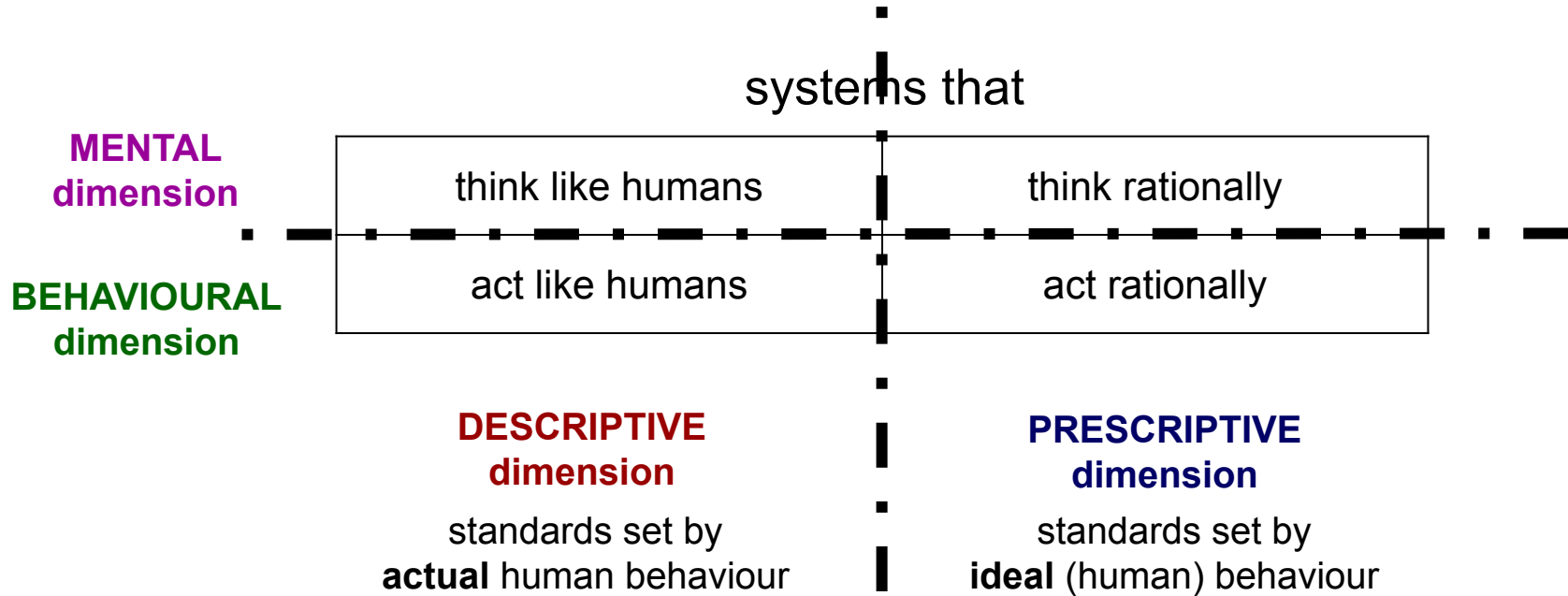
Russell and Norvig, "Artificial Intelligence: a Modern Approach", chapter 1.

<https://people.eecs.berkeley.edu/~russell/aima1e/chapter01.pdf>

Categories of AIs



Categories of AIs

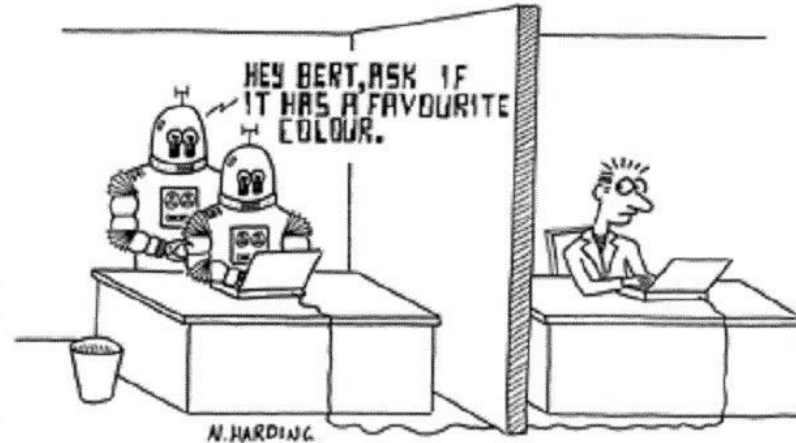


systems that

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 are required to achieve
artificial intelligence

EXPLAINABILITY argument

If they explain our
internal working



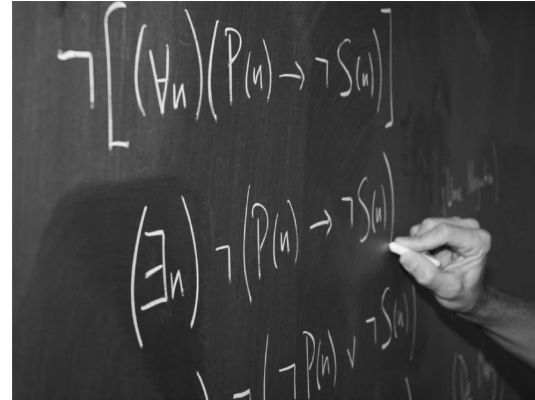
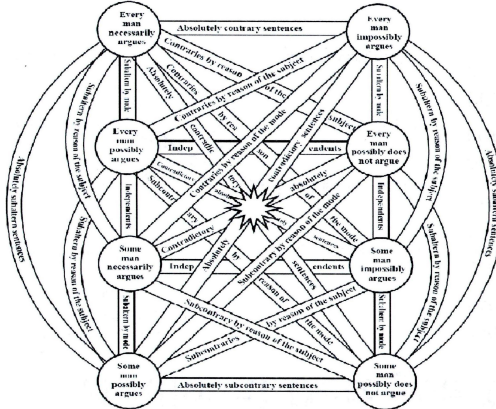
they can help to interpret AI
functioning

systems that

think like humans	think rationally
act like humans	act rationally

The “Laws of Thought” approach

AI producing logically valid inferences

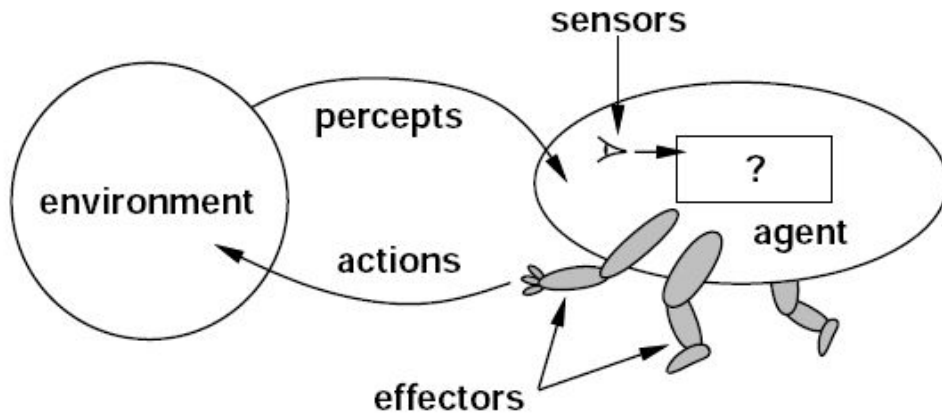


systems that

think like humans	think rationally
act like humans	act rationally

The “Rational Agent” approach

AI decision-making following standards of rationality



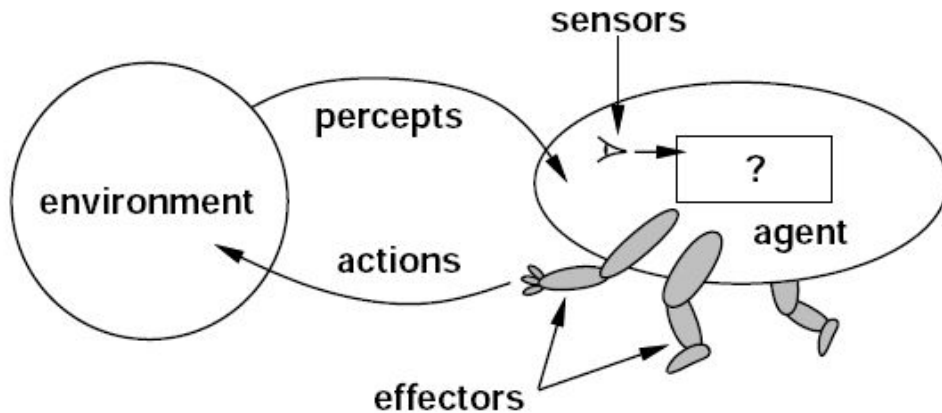
- the agent selects the best choice of **action**
- (to produce the means) to achieve its **goals**
- given its **beliefs**

systems that

think like humans	think rationally
act like humans	act rationally

The “Rational Agent” approach

AI decision-making following standards of rationality



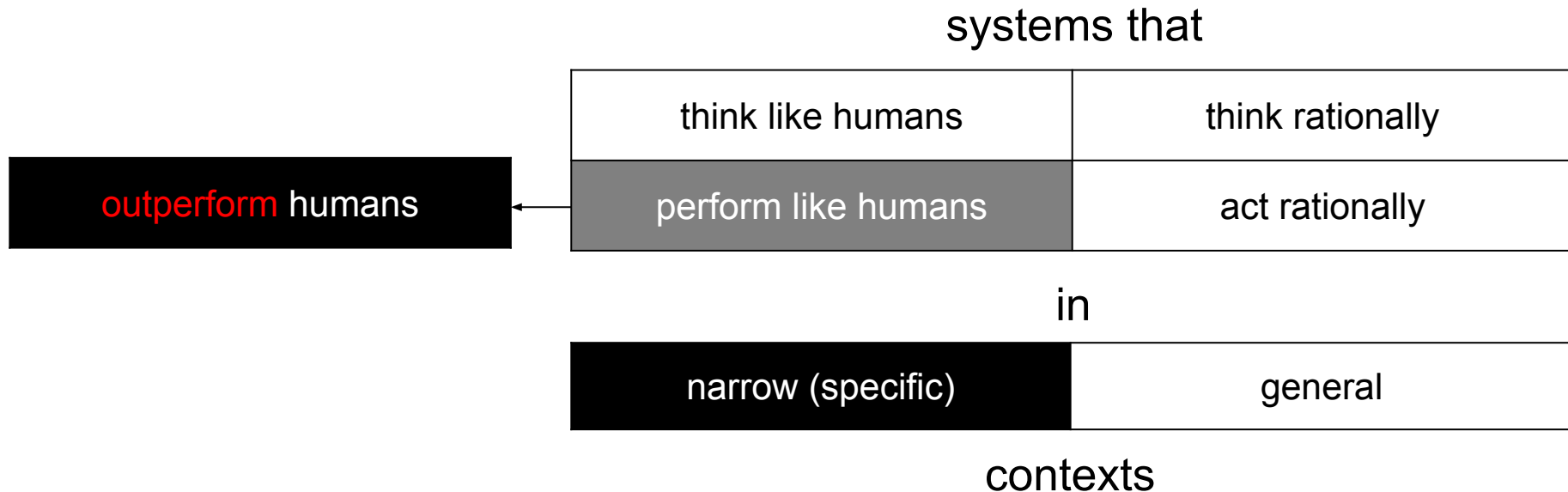
- the agent selects the best choice of **action**
- (to produce the means) to achieve its **goals**
- given its **beliefs**

➡ **“autonomous” entity**

Superhuman performances?

- In specific tasks, performance can be easily measured (quantified).

 *systems can adapt to perform better than humans.*

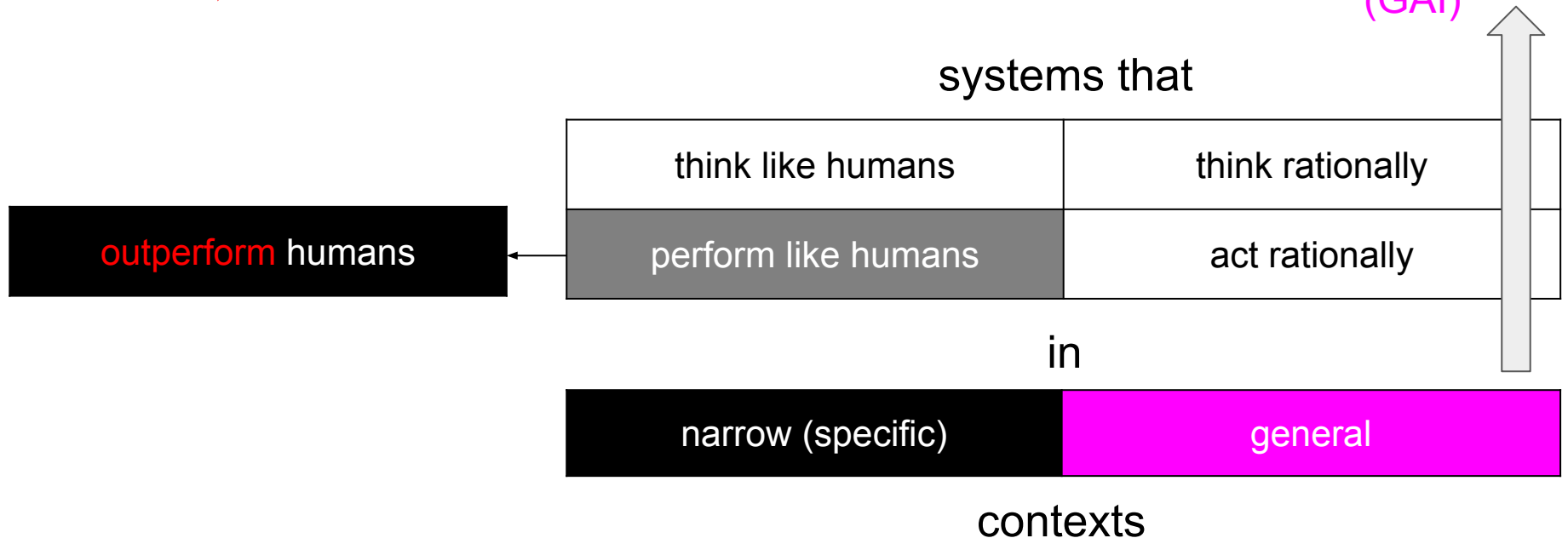


Superhuman performances?

- In specific tasks, performance can be easily measured (quantified).

 *systems can adapt to perform better than humans.*

General
Artificial
Intelligence
(GAI)

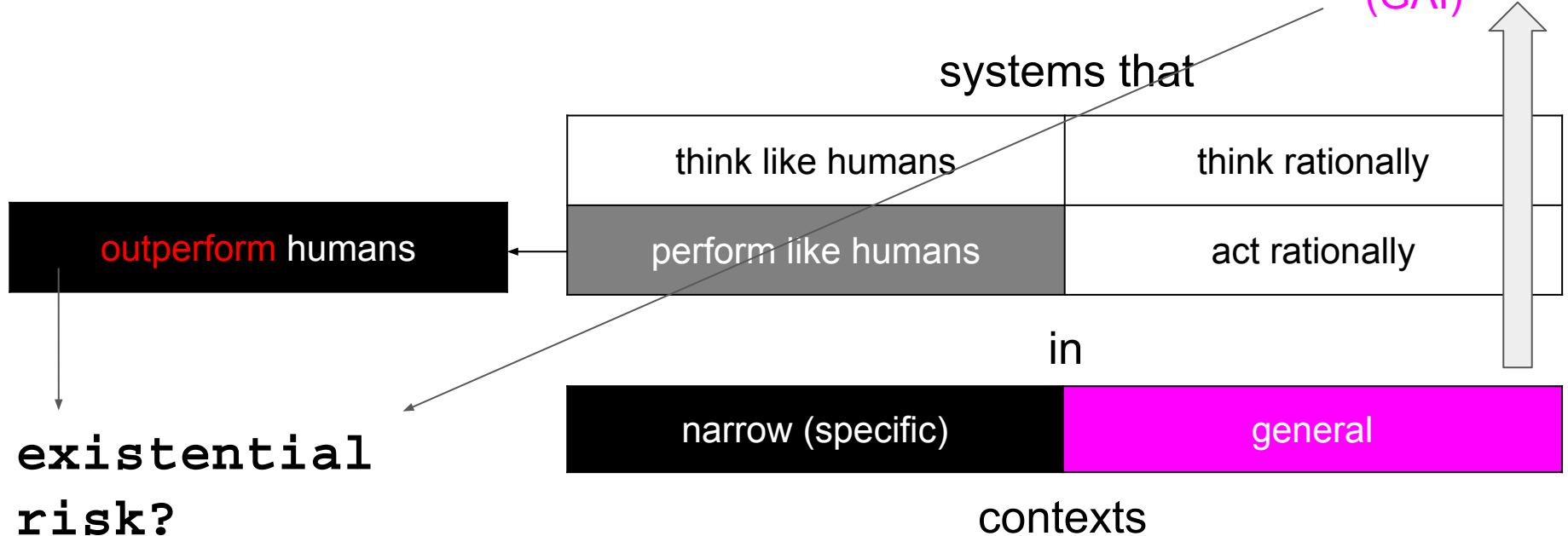


Superhuman performances?

- In specific tasks, performance can be easily measured (quantified).

→ *systems can adapt to perform better than humans.*

General
Artificial
Intelligence
(GAI)



Of the many AI waves

- This variety of topics has been developed through a cycles of *springs* (and *winters*) centered around different topics.

Of the many 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/20s)
 - generative AI (20s)



NOW

Key message

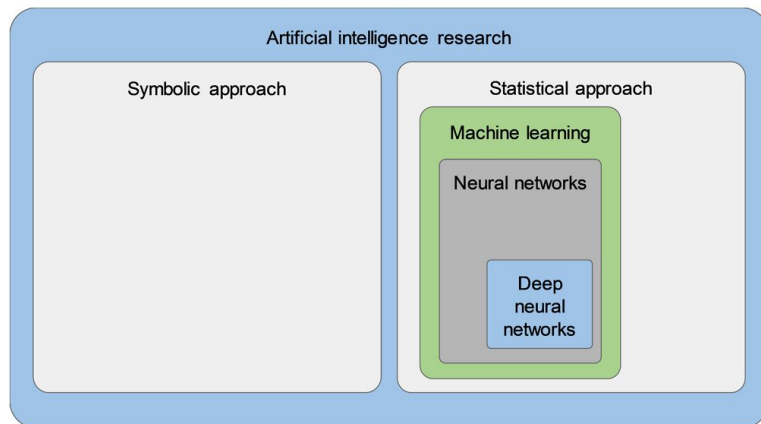
- Despite the contemporary (ab)use of the term, AI covers much more than the subfields that have brought the most recent advances:

AI $\neq$ **ML** or **DL**
machine learning deep learning

Key message

- Despite the contemporary (ab)use of the term, AI covers much more than the subfields that have brought the most recent advances:

AI $\neq$ **ML** or **DL**
machine learning deep learning



Key message

- Despite the contemporary (ab)use of the term, AI covers much more than the subfields that have brought the most recent advances:

AI $\neq$ **ML** or **DL**
machine learning deep learning

- This misappropriation is not new (even the term ML was coined by researchers to distinguish themselves from logic-based AI).

logician – symbolic

AI AS ENGINEERING OF THE “MIND”

empiricist – statistical

logician – symbolic



reasoning and decision-making

AI AS ENGINEERING OF THE “MIND”

induction of functions from data

empiricist – statistical

logician – symbolic



EXPLICITATION

reasoning and decision-making

AI AS ENGINEERING OF THE “MIND”

induction of functions from data

INTERNALIZATION

empiricist – statistical

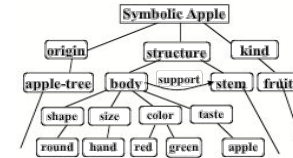
logacist – symbolic

EXPLICITATION

verbal input + explicit knowledge $\Rightarrow$ conclusion

RED
FRUIT

ROUND
SWEET

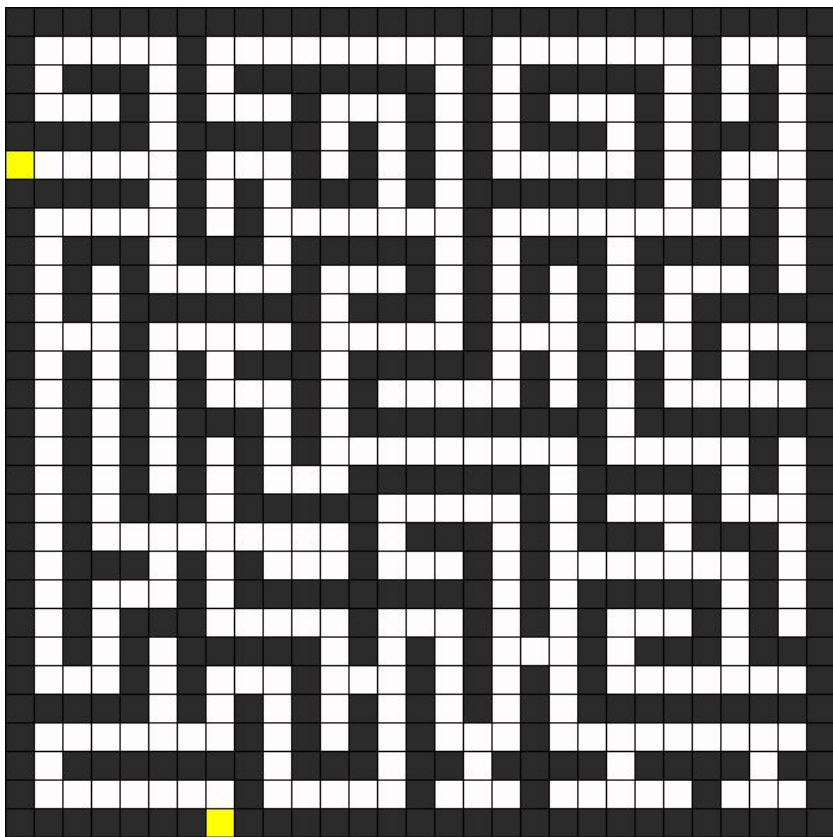


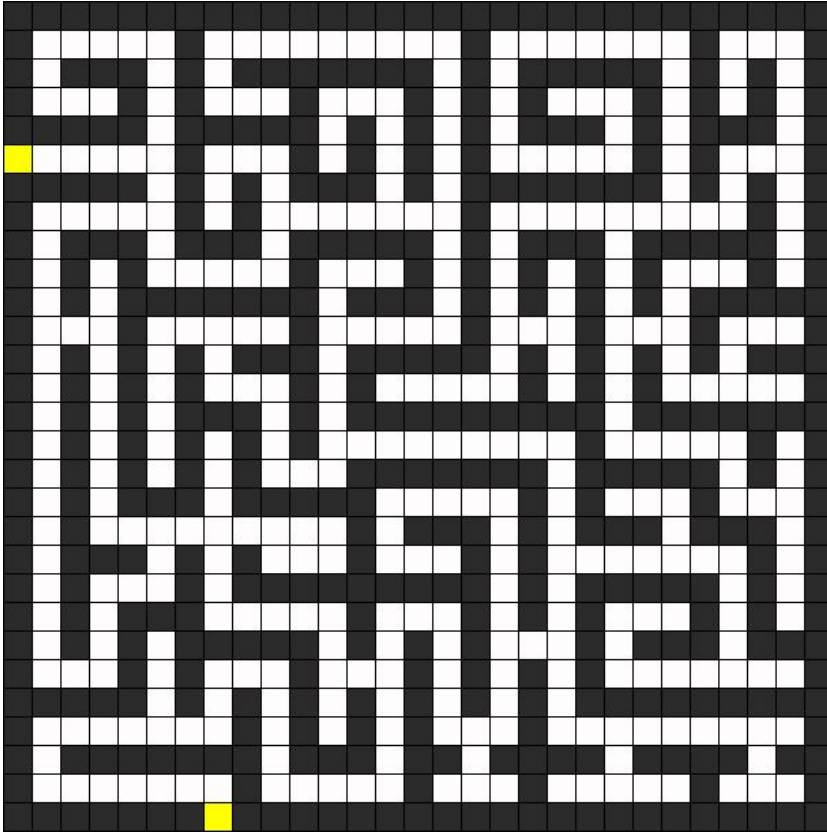
APPLE

INTERNALIZATION

empiricist – statistical

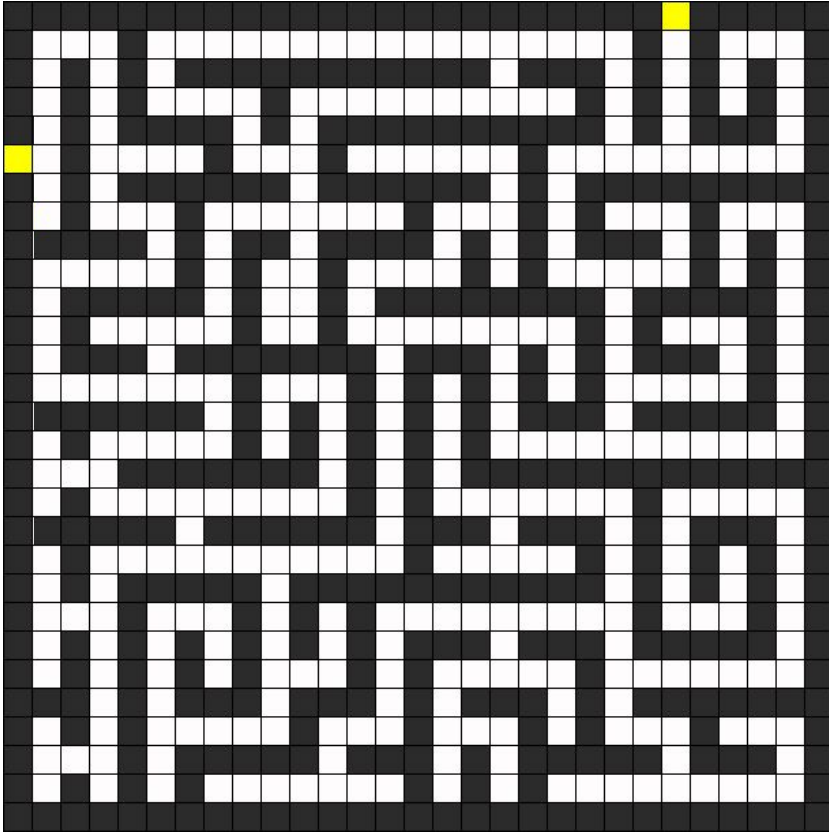
Working principles of Symbolic AI





Imperative approach

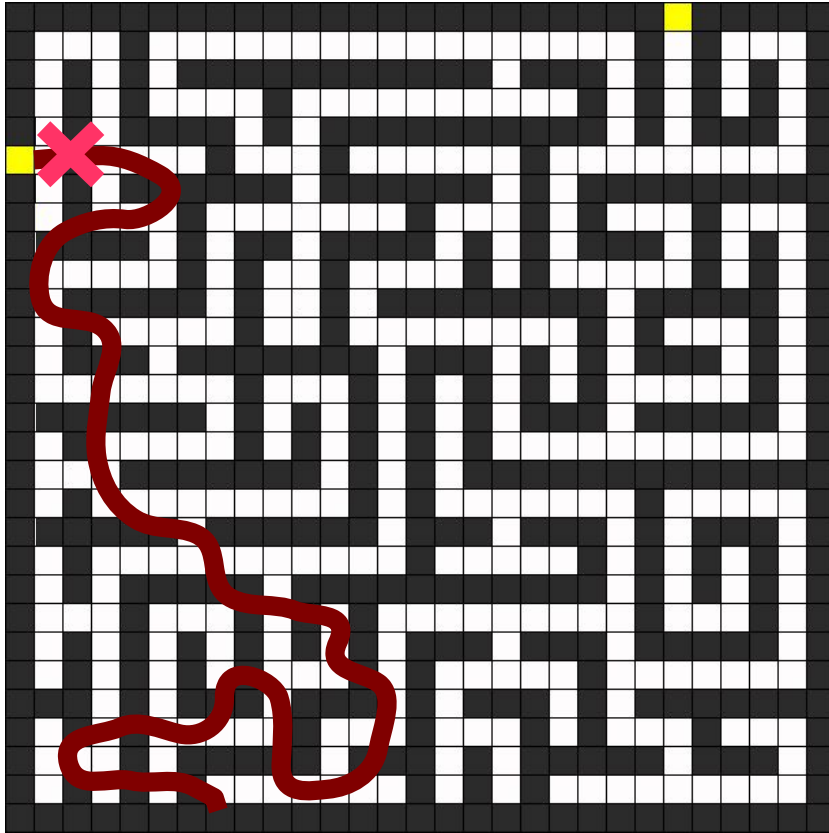
you command the directions



Imperative approach

you command the directions

*But what if the labyrinth
changes?*

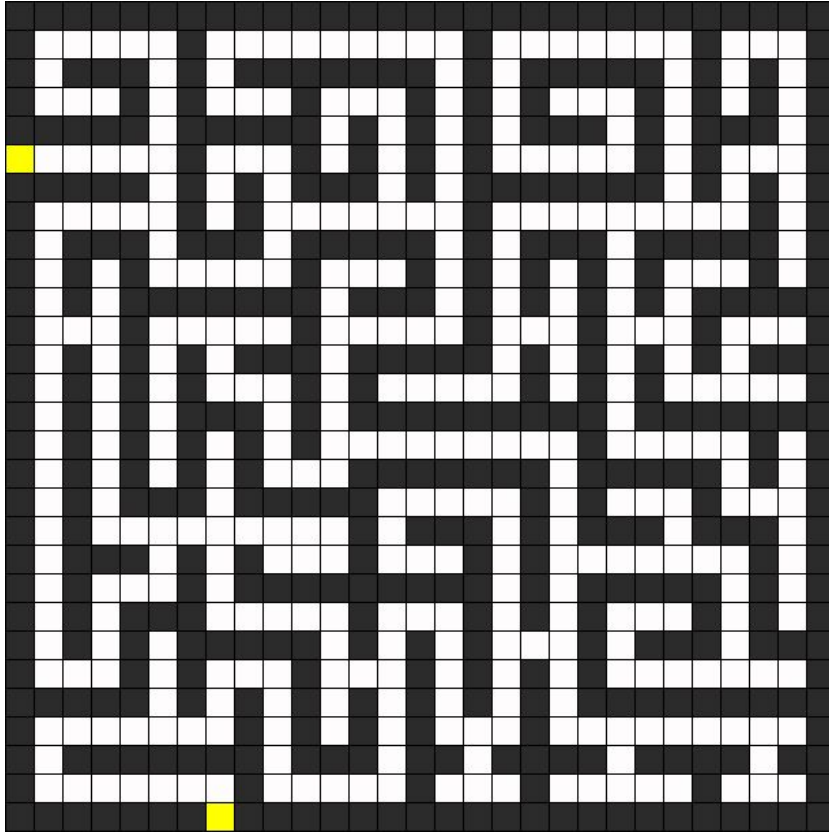


Imperative approach

you command the directions

*But what if the labyrinth
changes?*

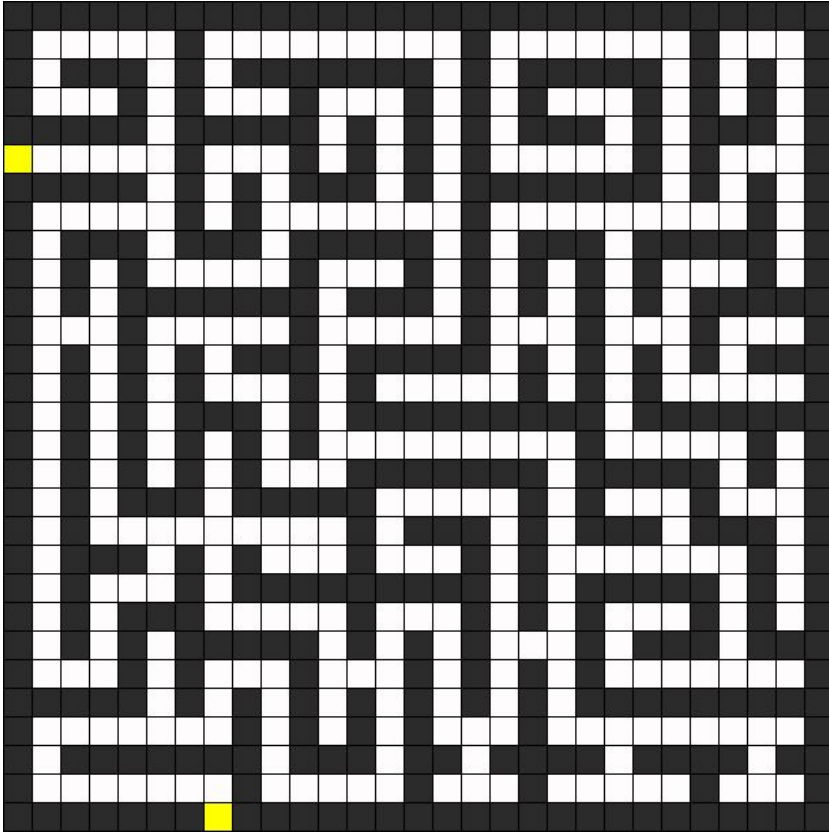
**We need to provide new
instructions...**



Declarative approach

you give just the labyrinth.

The computer finds the way.

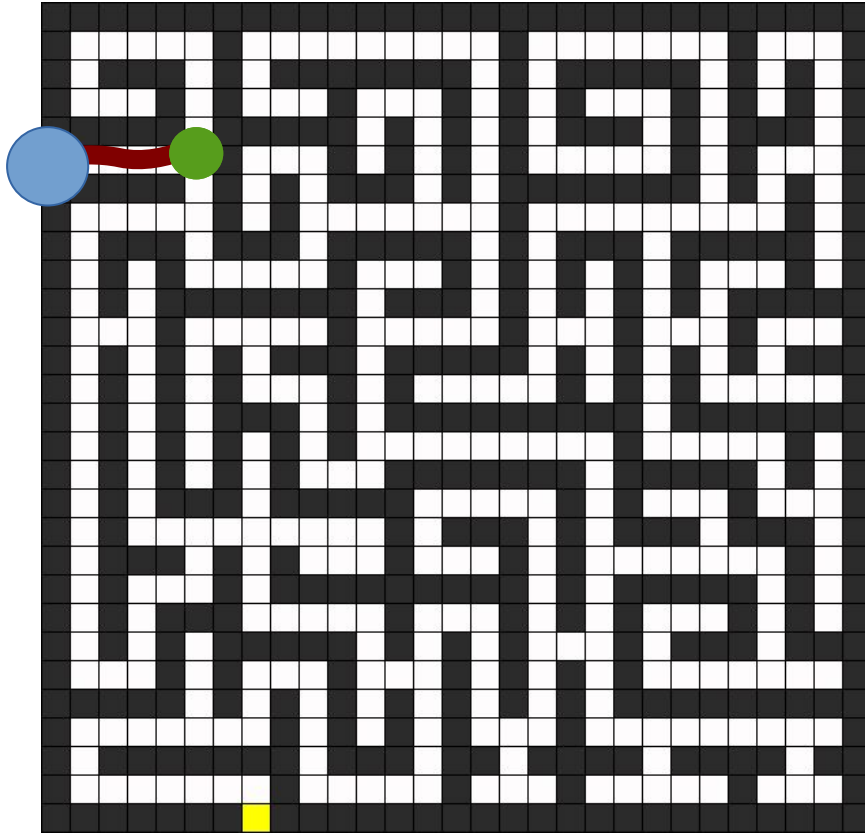


Declarative approach

you give just the labyrinth.

The computer finds the way.

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

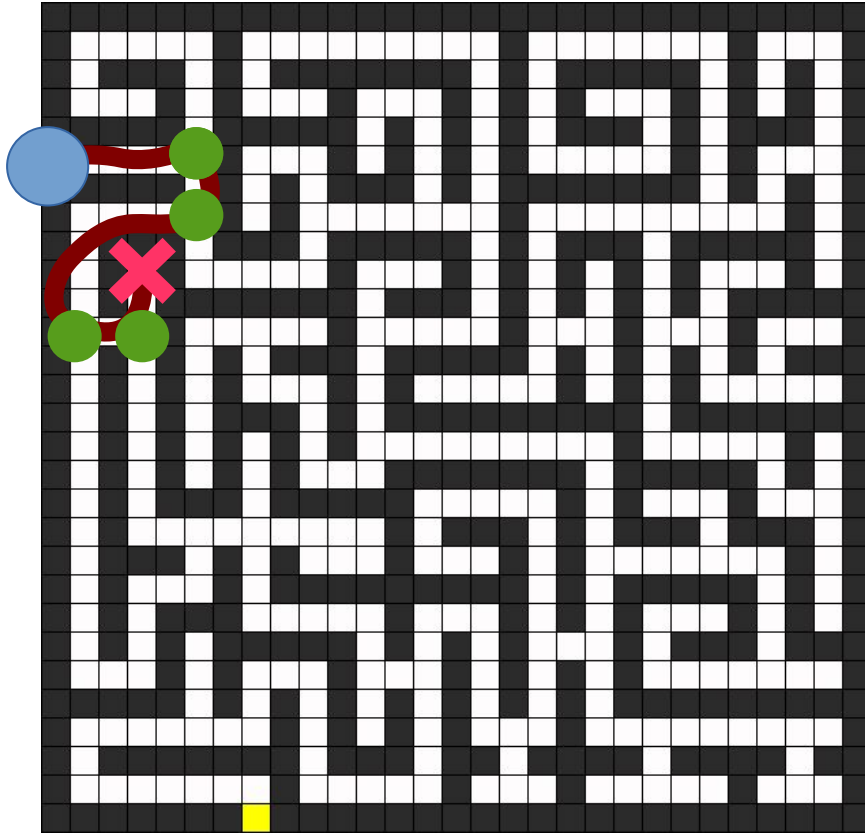


Declarative approach

you give just the labyrinth.

The computer finds the way.

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

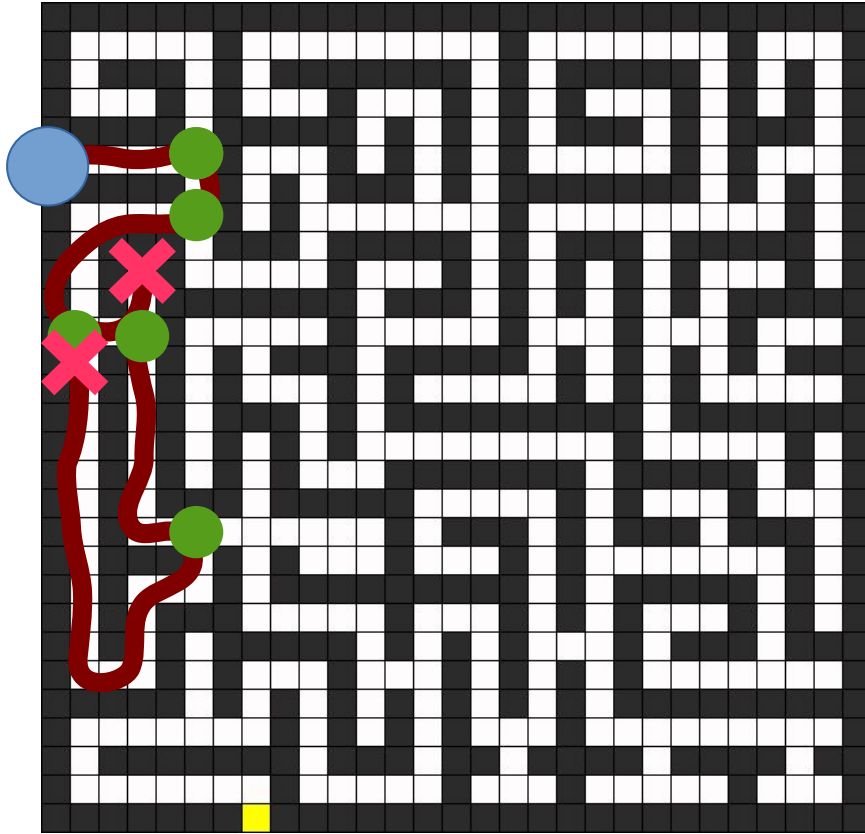


Declarative approach

you give just the labyrinth.

The computer finds the way.

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

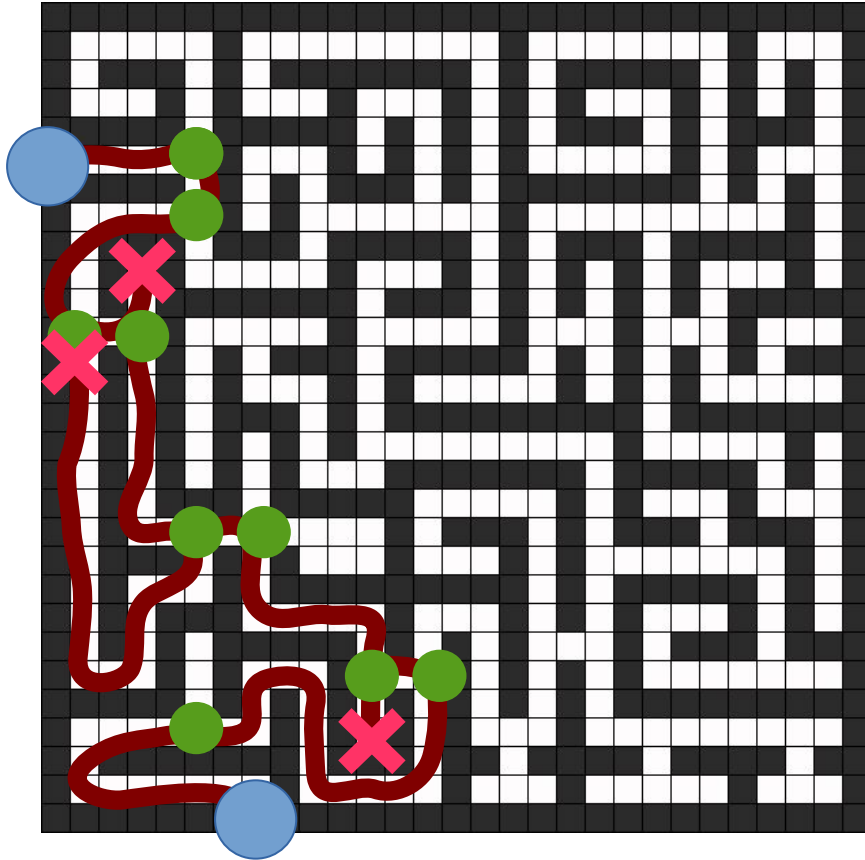


Declarative approach

you give just the labyrinth.

The computer finds the way.

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



Declarative approach

you give just the labyrinth.

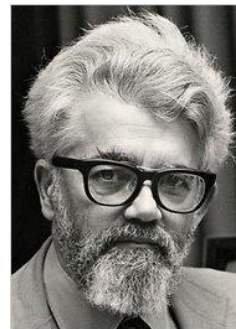
The computer finds the way.

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

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.



Intelligence as **search**: Garry Kasparov vs IBM Deep Blue (1997)

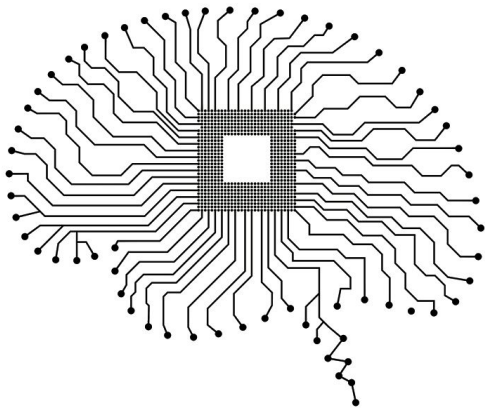


200 million positions per second
alpha-beta min-max search
static evaluation heuristics ***based on matches played by masters***

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 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

- First AI developers never believed the mind was a monolithic system, so they tinkered with heuristics, *ad-hoc* methods, and opportunistically with logic.

(the first chatbot)

ELIZA

Weizenbaum ~1965

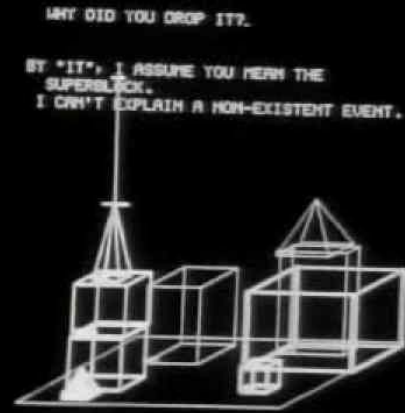
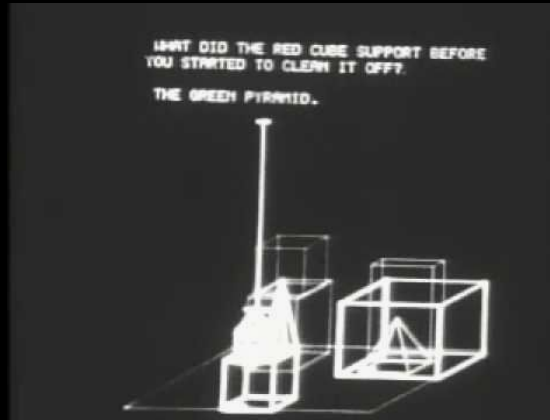
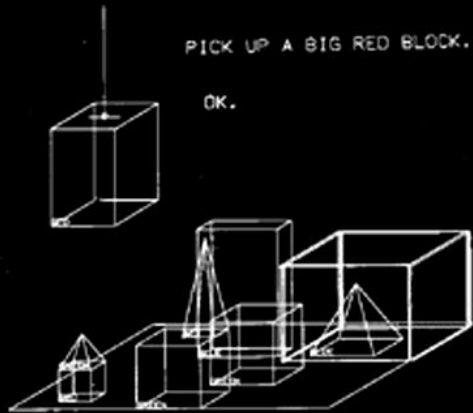
```
ELIZA: Is something troubling you ?
YOU: I am bored
ELIZA: Do you know anyone else who is bored ?
YOU: Maybe
ELIZA: You aren't sure ?
YOU: I don't really know
ELIZA: Why don't you really know ?
YOU: Well, I haven't really asked
ELIZA: Do you say you haven't really asked for some special reason ?
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 LLLLLLLL 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



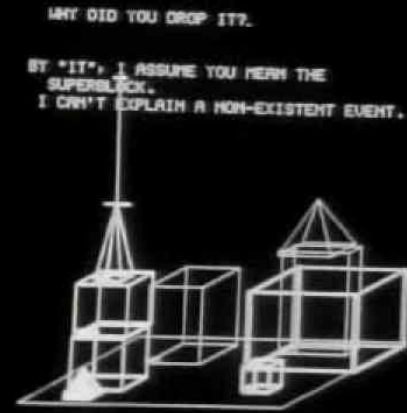
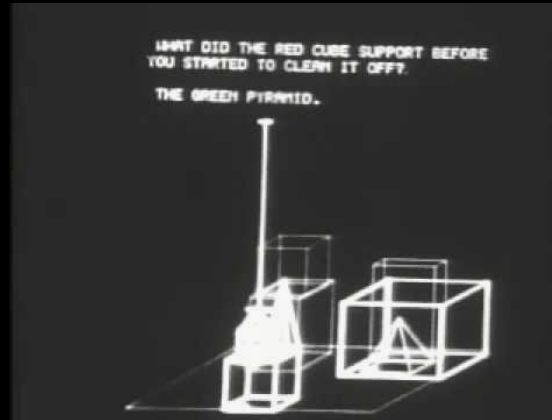
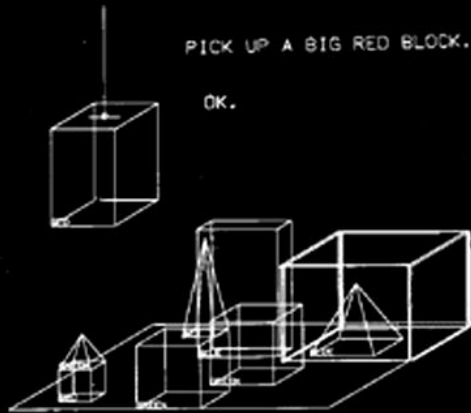
Deeper linguistic understanding
but limited to simple *blocks* worlds

<https://hci.stanford.edu/~winograd/shrdlu/>

SHRDLU

Winograd ~1969

...but these successes were impossible to generalize.



Deeper linguistic understanding
but limited to simple *blocks* worlds

<https://hci.stanford.edu/~winograd/shrdlu/>

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.



NOT APPROVED

Facing overwhelming difficulties to go beyond 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.



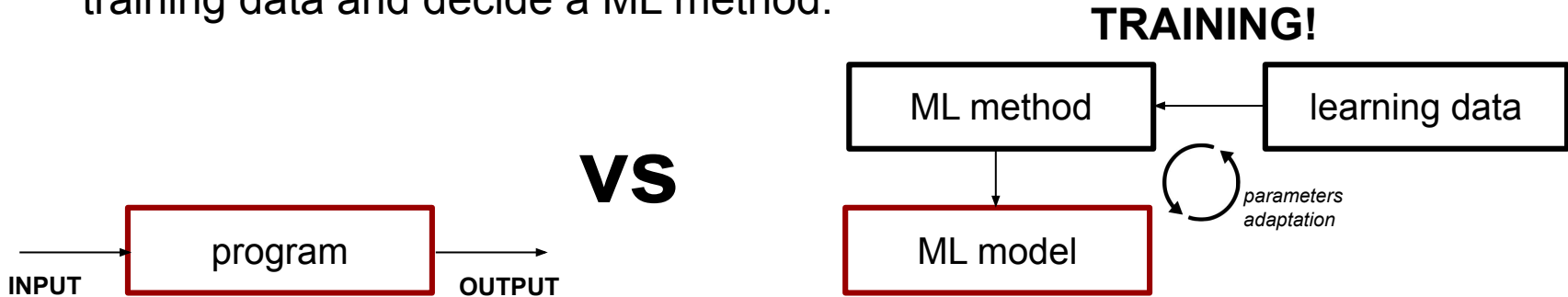
Machine learning

- *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.



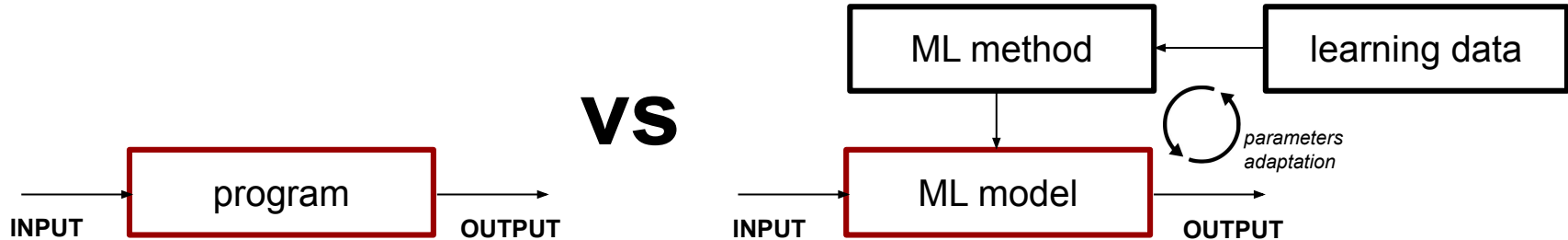
Machine learning

- *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.



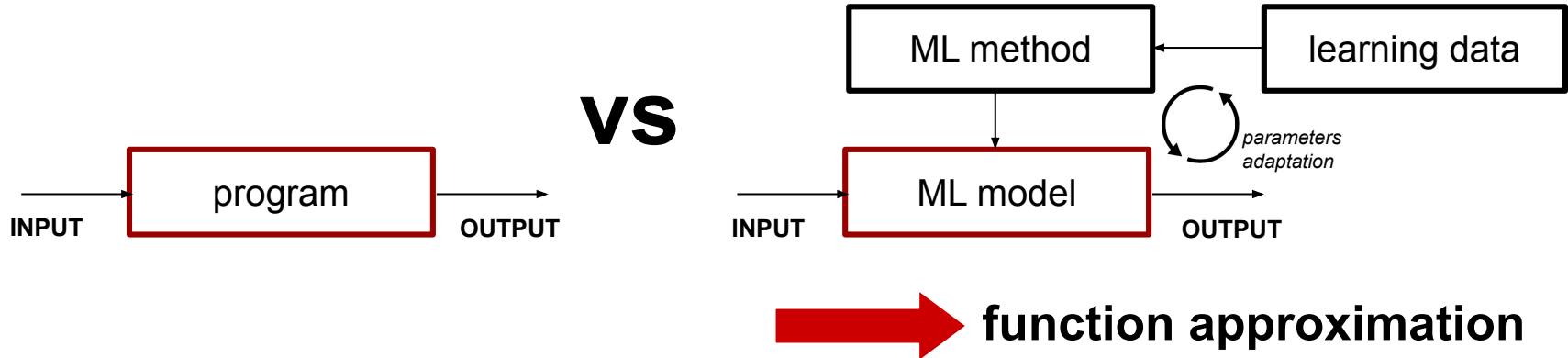
Machine learning

- *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.

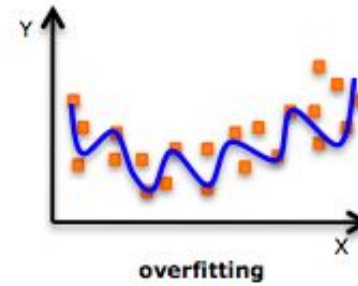
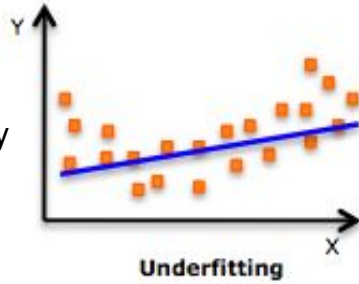


Machine learning

- *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.



max error
min model complexity



min error
max model complexity

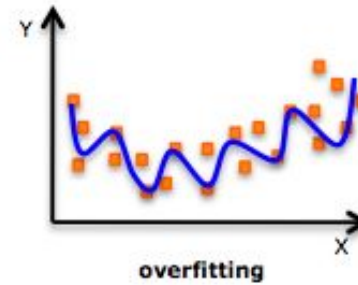
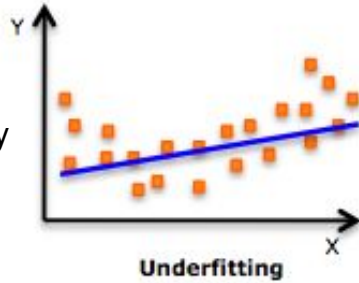
How to adapt to training data is often not straightforward!

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



➡ function approximation

max error
min model complexity



min error
max model complexity

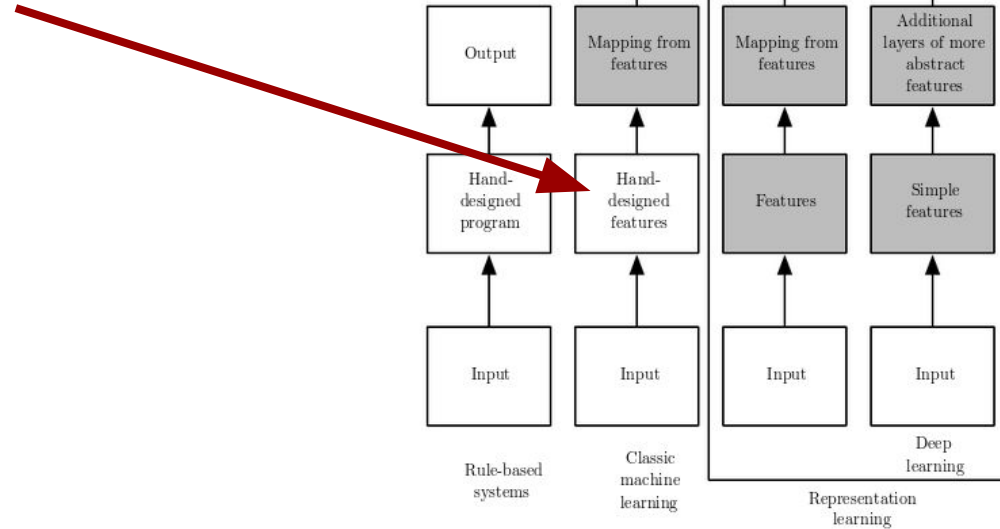
How to adapt to training data is often not straightforward!

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

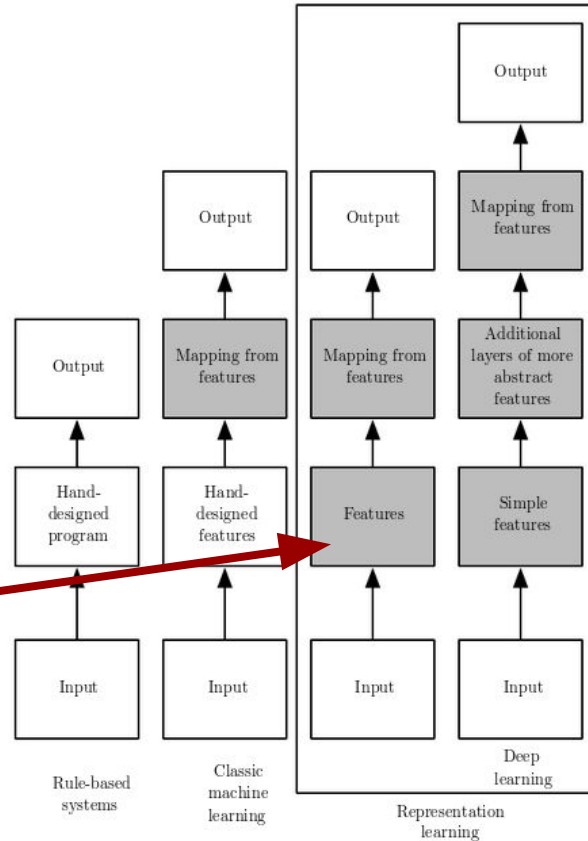


➔ function approximation

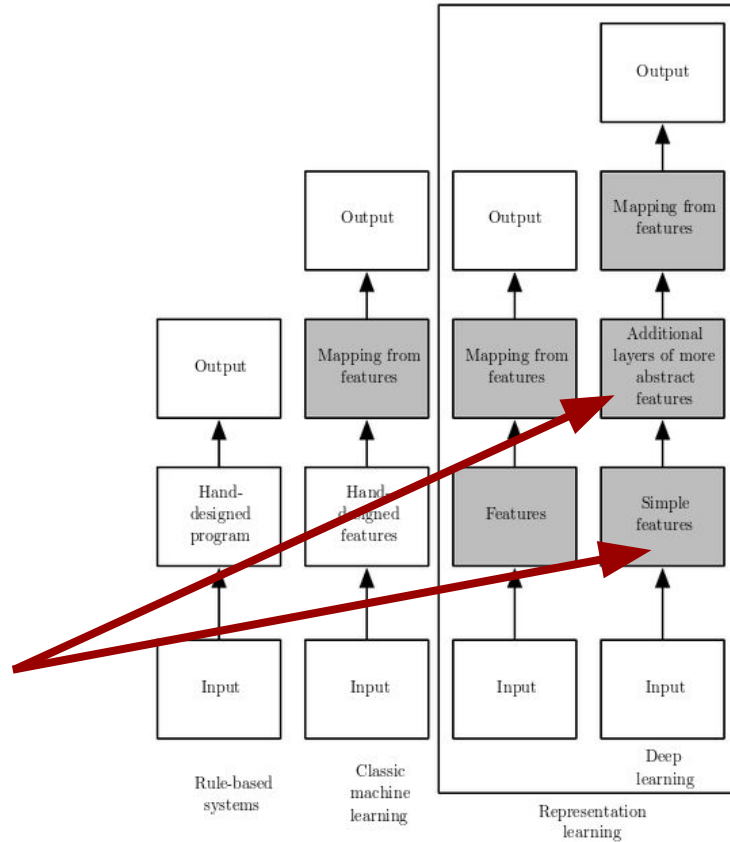
To reduce data requirements,
in classic ML features deemed to be
relevant are manually selected by the
developer from the available input.

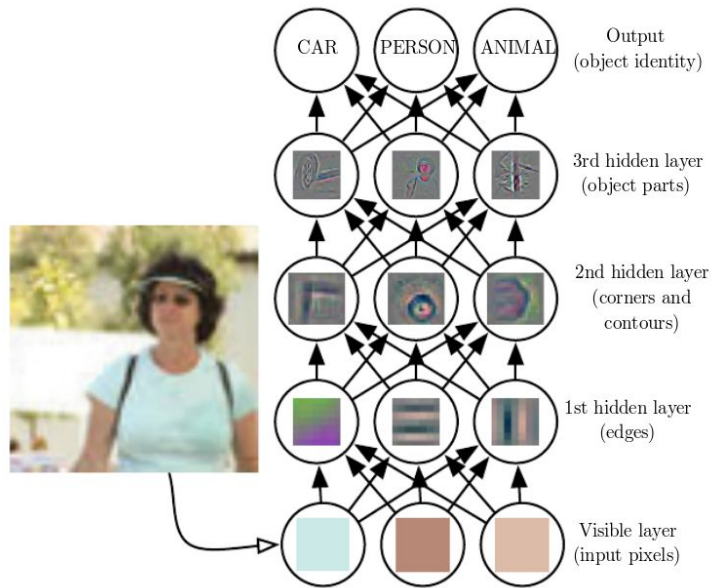


When this is not possible,
features have to be extracted as
well, through some
representation learning.

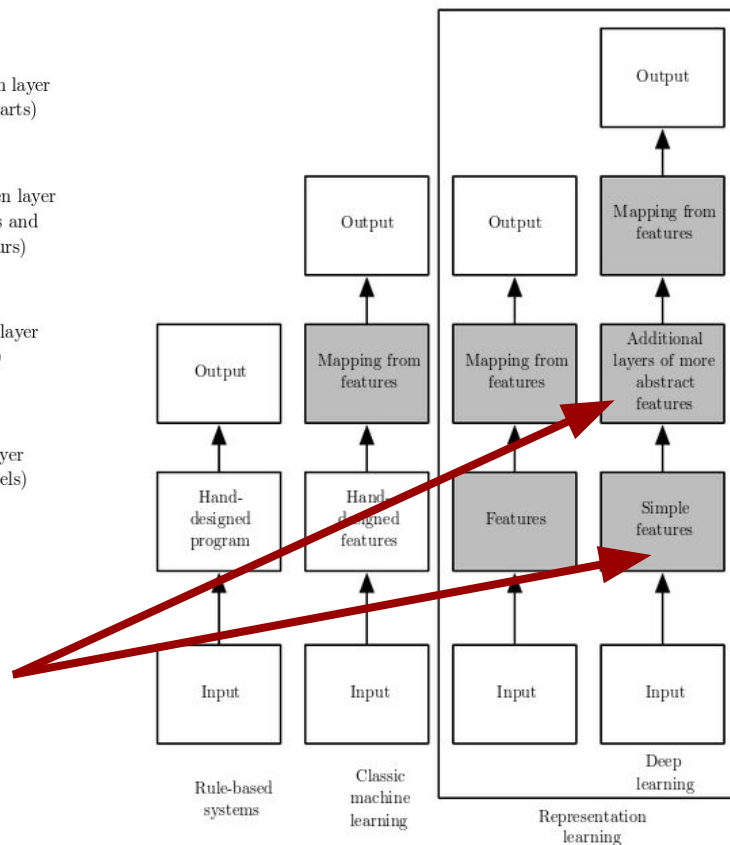


Deep learning relies on a **hierarchy** of representation learning, producing different levels of abstractions



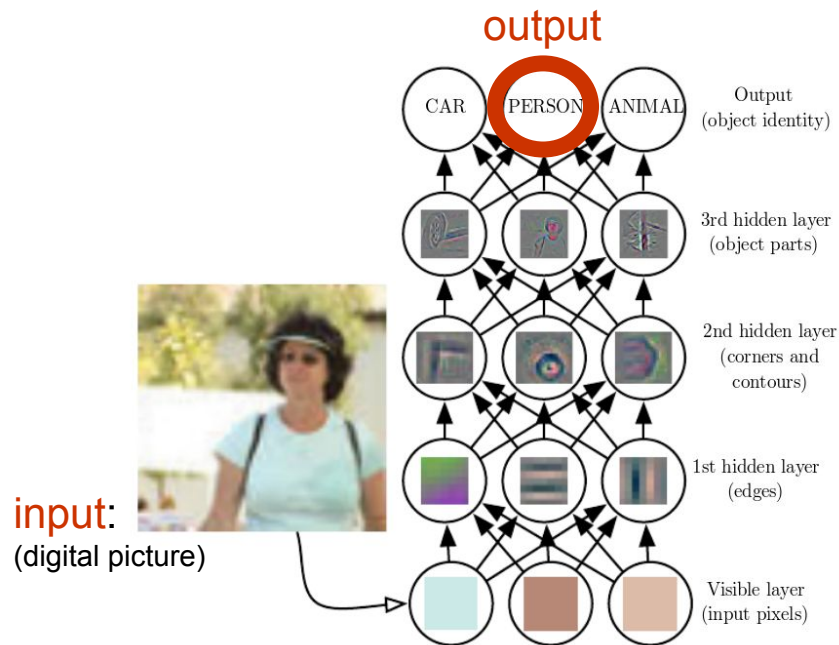


Deep learning relies on a hierarchy of representation learning, producing different levels of abstractions



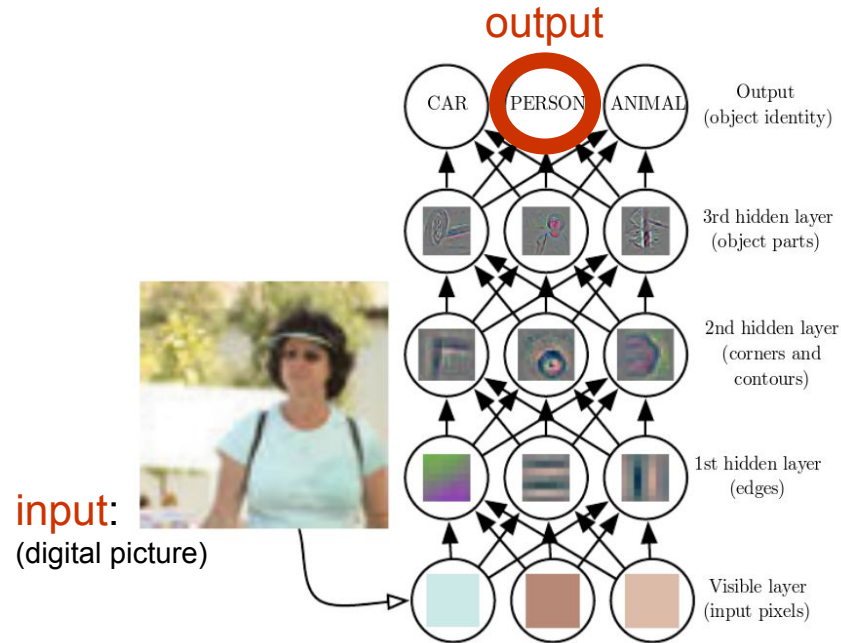
Classifiers

Classifiers associate symbols to (typically) non-symbolic inputs



Classifiers

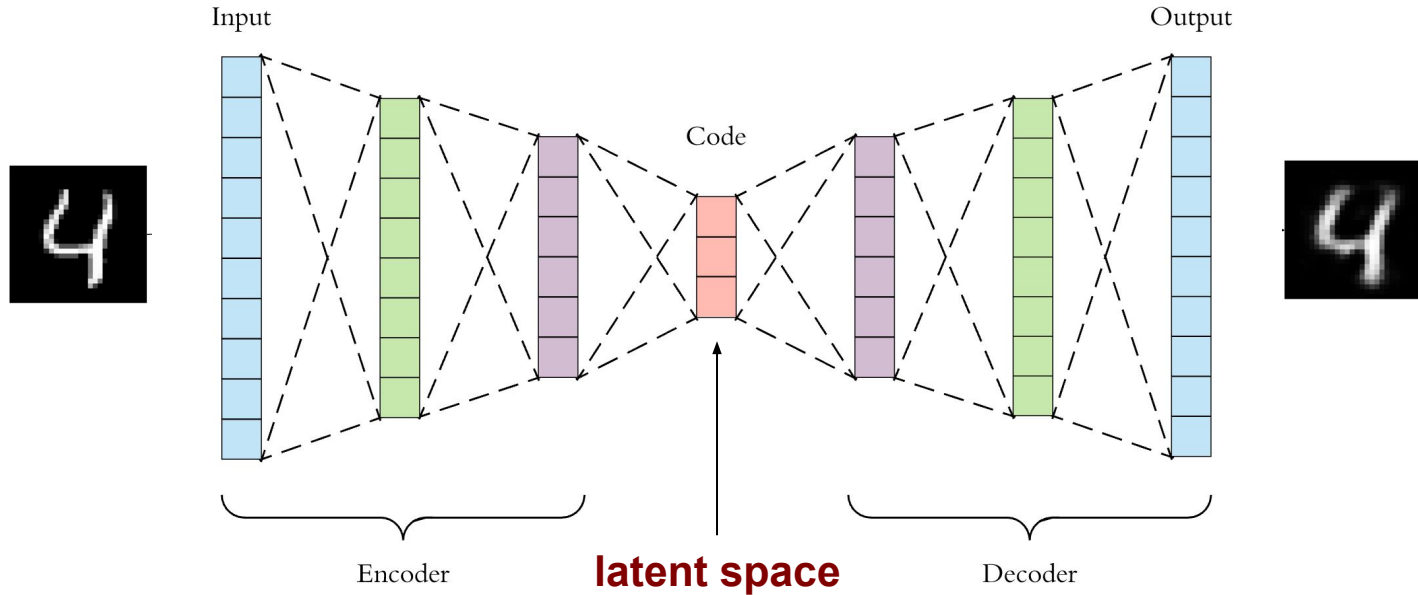
Classifiers associate symbols to (typically) non-symbolic inputs



SUPERVISED method: you need the labels for the training set.

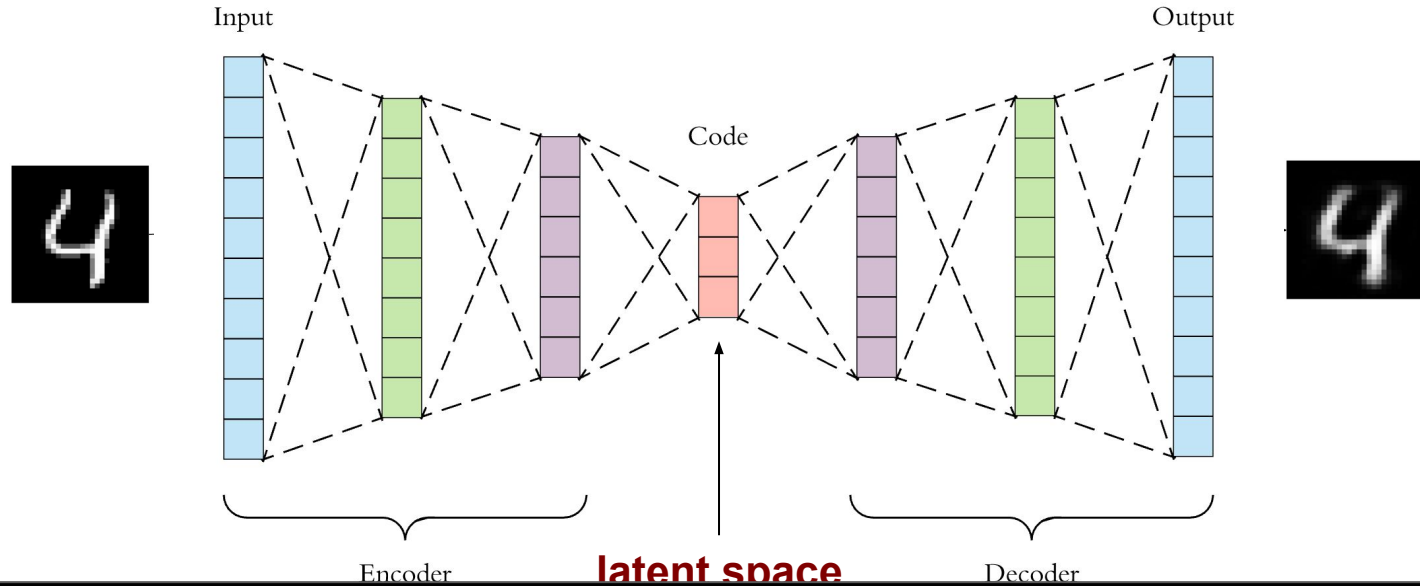
Autoencoders

Autoencoders function as “compressors”: what are the features that make the output as much similar as possible to the input



Autoencoders

Autoencoders function as “compressors”: what are the features that make the output as much similar as possible to the input



UNSUPERVISED method: you just need the data (no labels).

Why it works? Plausibly it targets convexity regions in the world

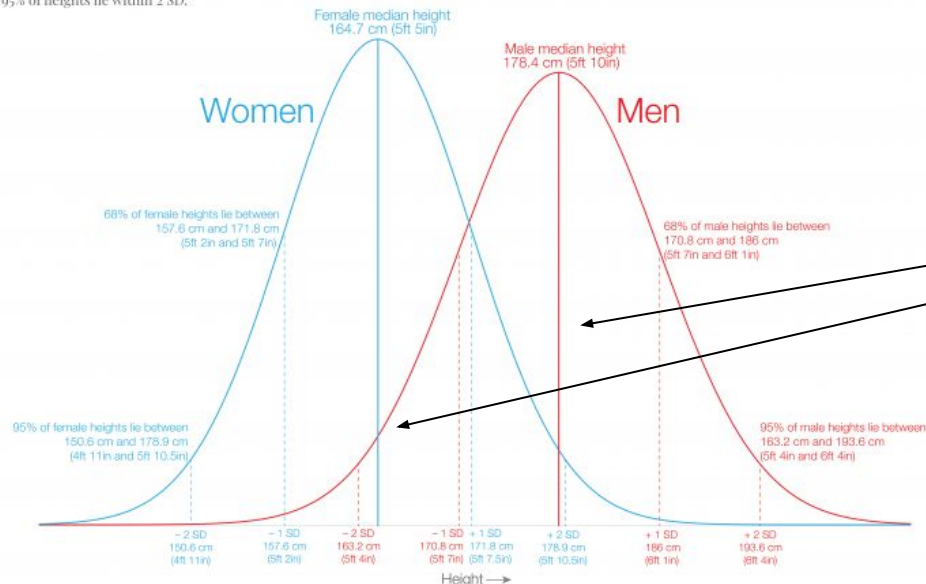
The distribution of male and female heights

The distribution of adult heights for men and women based on large cohort studies across 20 countries in North America, Europe, East Asia and Australia. Shown is the sample-weighted distribution across all cohorts born between 1980 and 1994 (so reaching the age of 18 between 2008 and 2012).

Since human heights within a population typically form a normal distribution:

- 68% of heights lie within 1 standard deviation (SD) of the median height;
- 95% of heights lie within 2 SD.

Our World
in Data



maintain the
prototypes,
not the
instances

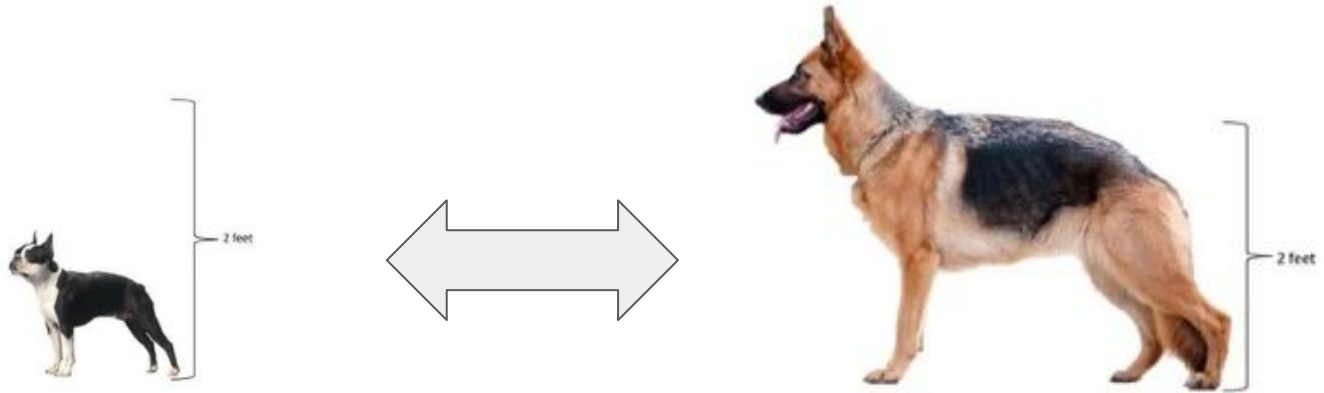
Note: this distribution of heights is not globally representative since it does not include all world regions due to data availability.

Data source: Jelenkovic et al. (2016). Genetic and environmental influences on height from infancy to early adulthood: An individual-based pooled analysis of 45 twin cohorts.

This is a visualization from OurWorldInData.org, where you find data and research on how the world is changing.

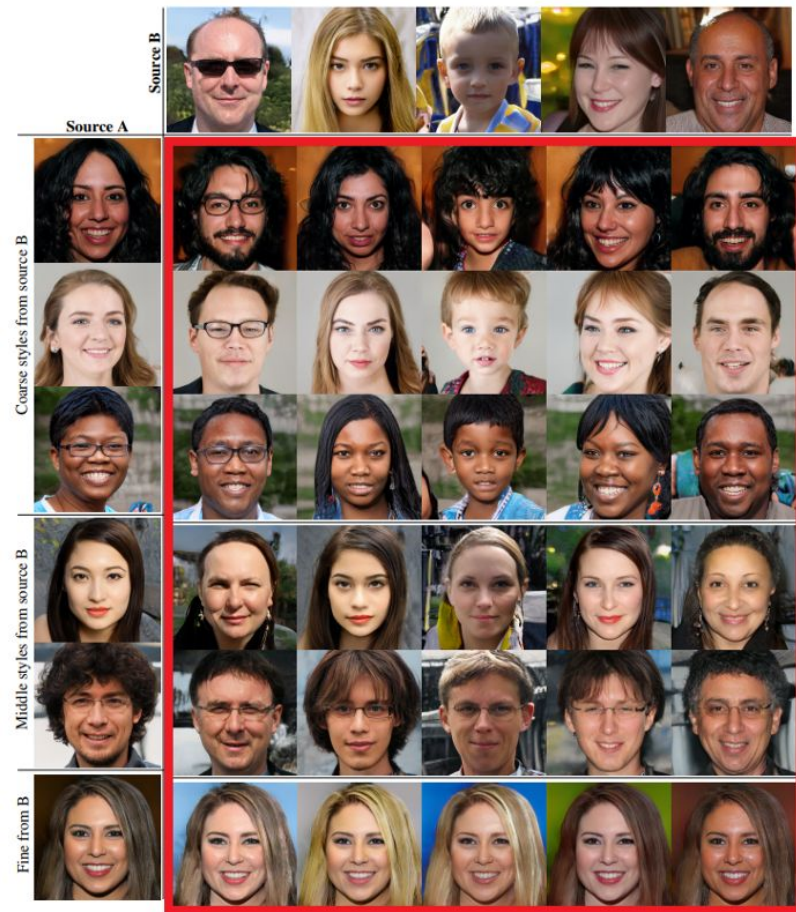
Licensed under CC-BY by the author Cameron Appel.

Humans exhibit as well some form of latent spaces!

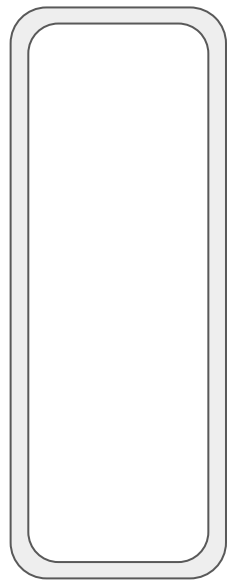


- Given for instance two arbitrary dogs, we can imagine all kind of dogs which are in-between the two! (i.e. the two dog concepts behave as two points in a mental space; we can move from one towards the other, finding new points, generating new dogs).

what you can do
with a latent
space?



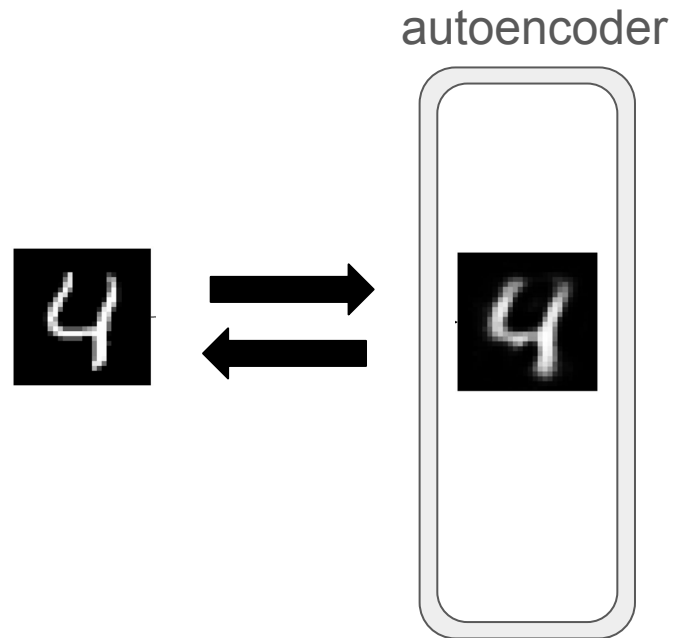
AI models as trained “mirrors”



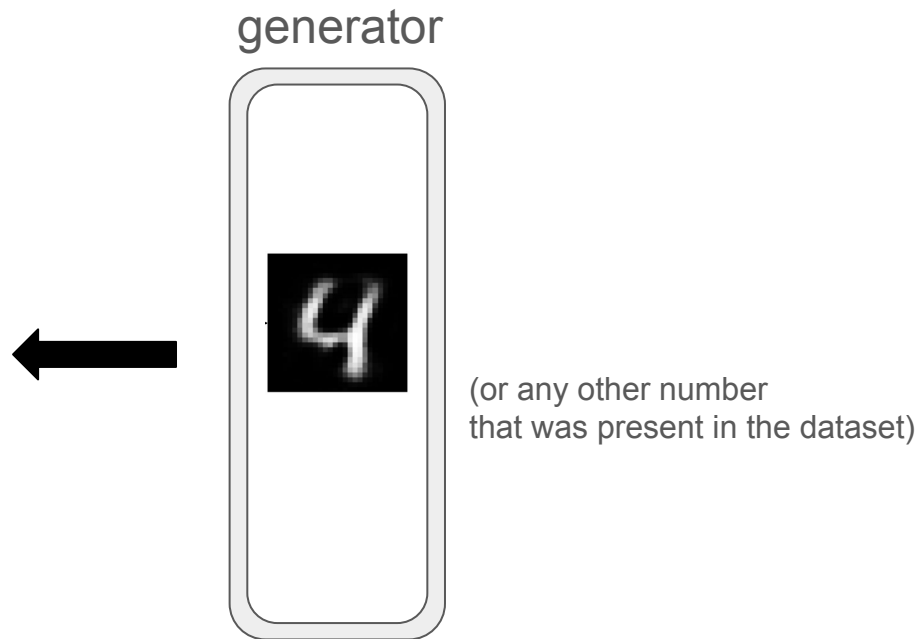
AI models as trained “mirrors”



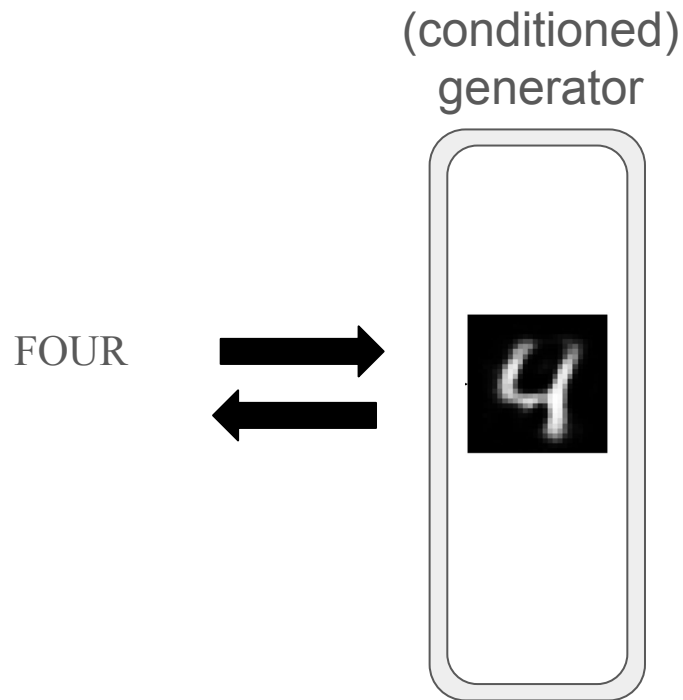
AI models as trained “mirrors”



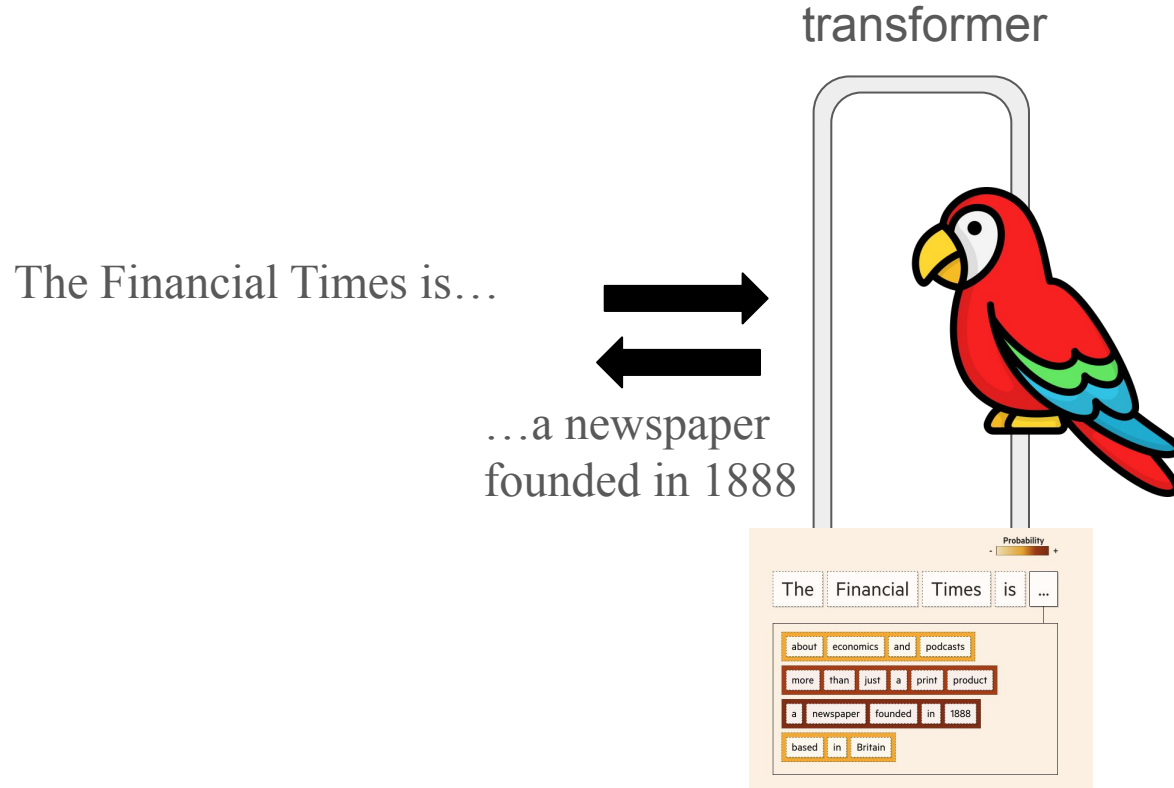
AI models as trained “mirrors”



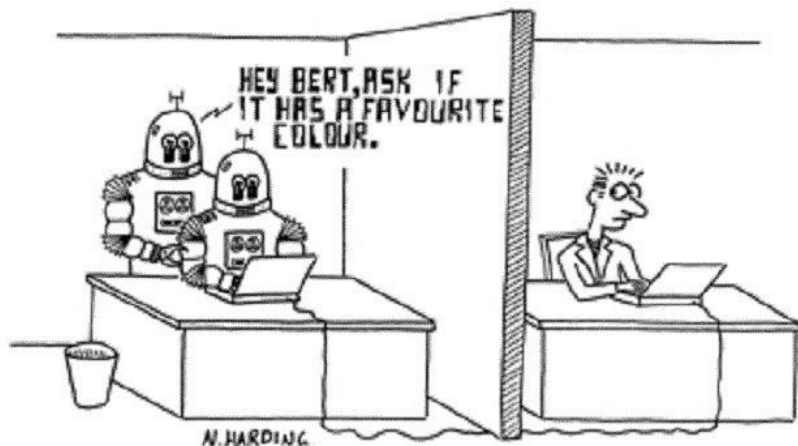
AI models as trained “mirrors”



AI models as trained “mirrors”



AI models as trained “mirrors”



Does GPT-4 Pass the Turing Test?

Cameron Jones and Benjamin Bergen
UC San Diego,
9500 Gilman Dr, San Diego, CA
cameron@ucsd.edu

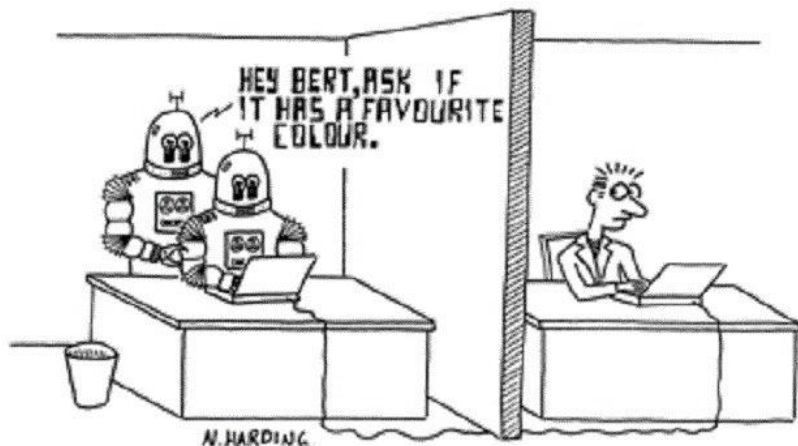
Abstract

We evaluated GPT-4 in a public online Turing Test. The best-performing GPT-4 prompt passed in 41% of games, outperforming baselines set by ELIZA (27%) and GPT-3.5 (14%), but falling short of chance and the baseline set by human participants (63%). Participants' decisions were based mainly on linguistic style (35%) and socio-emotional traits (27%), supporting the idea that intelligence is not sufficient to pass the Turing Test. Participants' demographics, including education and familiarity with LLMs, did not predict detection rate, suggesting that even those who understand systems deeply and interact with them frequently may be susceptible to deception. Despite known limitations as a test of intelligence, we argue that the Turing Test continues to be relevant as an assessment of naturalistic communication and deception. AI models with the ability to masquerade as humans could have widespread societal consequences, and we analyse the effectiveness of different strategies and criteria for judging human likeness.

Keywords: Turing Test, Large Language Models, GPT-4, interactive evaluation



AI models as trained “mirrors”



human 63%
GPT4 41%
ELIZA 27%
GPT3.5 14%

Does GPT-4 Pass the Turing Test?

Cameron Jones and Benjamin Bergen
UC San Diego,
9500 Gilman Dr, San Diego, CA
cameron@ucsd.edu

Abstract

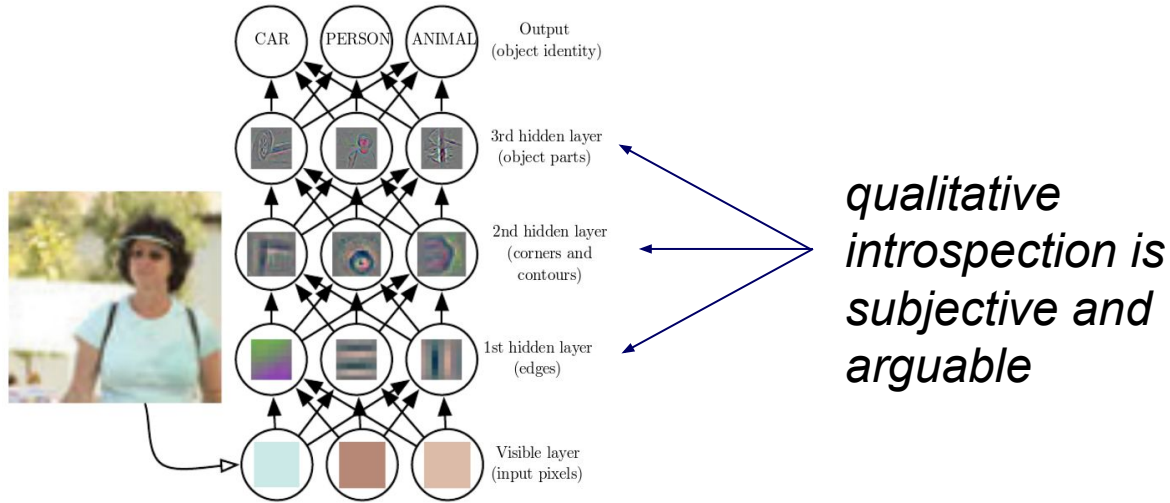
We evaluated GPT-4 in a public online Turing Test. The best-performing GPT-4 prompt passed in 41% of games, outperforming baselines set by ELIZA (27%) and GPT-3.5 (14%), but falling short of chance and the baseline set by human participants (63%). Participants' decisions were based mainly on linguistic style (35%) and socio-emotional traits (27%), supporting the idea that intelligence is not sufficient to pass the Turing Test. Participants' demographics, including education and familiarity with LLMs, did not predict detection rate, suggesting that even those who understand systems deeply and interact with them frequently may be susceptible to deception. Despite known limitations as a test of intelligence, we argue that the Turing Test continues to be relevant as an assessment of naturalistic communication and deception. AI models with the ability to masquerade as humans could have widespread societal consequences, and we analyse the effectiveness of different strategies and criteria for judging humanlikeness.

Keywords: Turing Test, Large Language Models, GPT-4, interactive evaluation



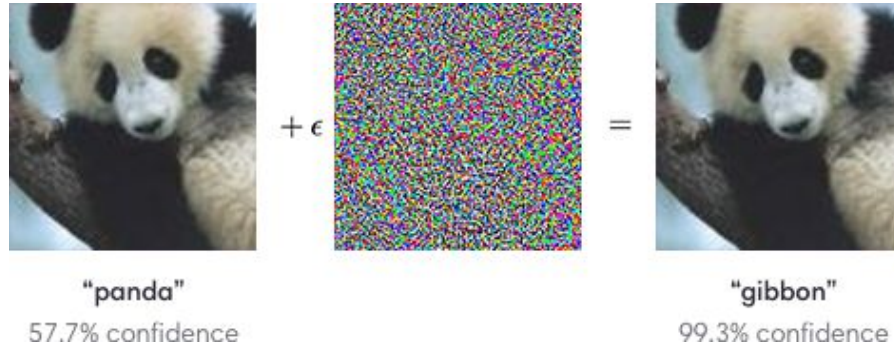
The dark sides of contemporary AI

Problem 1: Opacity



- The developer does not have direct control on which features will be considered to be **relevant** to the task.

Problem 2: 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.

Problem 3: Deception

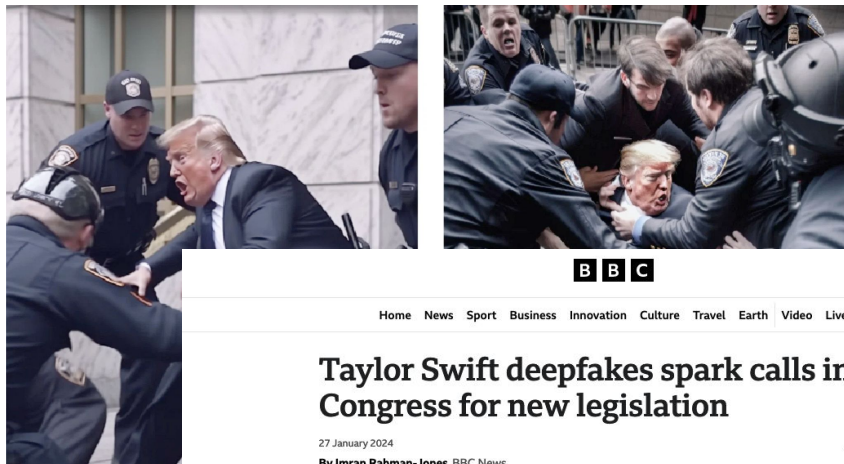
- Knowing what is relevant to *our vision*, someone can play dirty tricks.



Face to face (2016): <https://www.youtube.com/watch?v=ohmajJTcpNk>

Voice to lips (2017): <https://www.youtube.com/watch?v=9Yq67CjDqvw>

Problem 3: Deception



Taylor Swift deepfakes spark calls in Congress for new legislation

27 January 2024

By Imran Rahman-Jones, BBC News

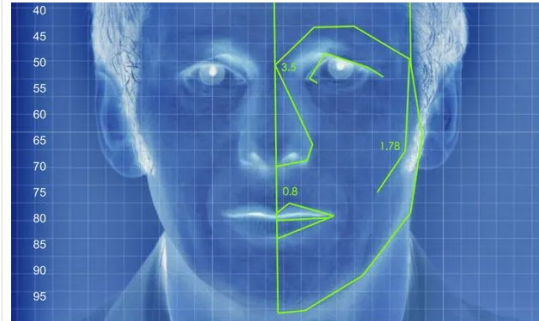
Share



Problem 4: Societal sustainability

New AI can guess whether you're gay or straight from a photograph

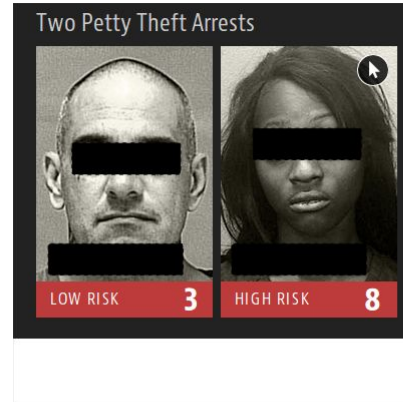
An algorithm deduced the sexuality of people on a dating site with up to 91% accuracy, raising tricky ethical questions



▲ An illustrated depiction of facial analysis technology similar to that used in the experiment. Illustration: Alamy
Artificial intelligence can accurately guess whether people are gay or straight based on photos of their faces, according to new research that suggests machines can have significantly better “gaydar” than humans.

The **study** from Stanford University - which found that a computer algorithm could correctly distinguish between gay and straight men 81% of the time,

<https://www.theguardian.com/technology/2017/sep/07/new-artificial-intelligence-can-tell-whether-youre-gay-or-straight-from-a-photograph> (2017)



COMPAS: software used in the US predicting future crimes and criminals argued to be biased against African Americans (2016)

Angwin J. et al. ProPublica, May 23 (2016). *Machine Bias: risk assessments in criminal sentencing*

SyRI (System Risk Indication) used in the Netherlands to create risk alerts for welfare frauds by processing and linking personal data of citizens argued to be discriminatory and unlawful (2018)

<https://pilpnjcm.nl/en/proceedings-risk-pro-filing-dutch-citizens-syri>

Problem 4: Societal sustainability



JURI SAYS:

I'm predicting judgments of the European Court of Human Rights with an accuracy of **87.2% over the *last month*.**

JURI reads published documents from previous years and decisions of the cases judged by the European Court of Human Rights and predicts decisions the Court will make. Every month it learns from its mistakes.



<https://jurisays.com/>

PLOS ONE

OPEN ACCESS PEER-REVIEWED

RESEARCH ARTICLE

A general approach for predicting the behavior of the Supreme Court of the United States

Daniel Martin Katz , Michael J. Bommarito II, Josh Blackman

Published: April 12, 2017 • <https://doi.org/10.1371/journal.pone.0174698>

Article	Authors	Metrics	Comments	Media Coverage
---------	---------	---------	----------	----------------

Abstract

Introduction.
Research principles and prior work
Data and feature engineering
Model construction
Model testing and results
Conclusion and future research
Acknowledgments
Author Contributions

Abstract

Building on developments in machine learning and prior work in the science of judicial prediction, we construct a model designed to predict the behavior of the Supreme Court of the United States in a generalized, out-of-sample context. To do so, we develop a time-evolving random forest classifier that leverages unique feature engineering to predict more than 240,000 justice votes and 28,000 cases outcomes over nearly two centuries (1810–2015). Using only data available prior to decision, our model outperforms null (baseline) models at both the justice and case level under both parametric and non-parametric tests. Over nearly two centuries, we achieve 70.2% accuracy at the case outcome level and 71.9% at the justice vote level. More recently, over the past century, we outperform an *in-sample* optimized null model by nearly 5%. Our performance is consistent with, and improves on the general level of prediction demonstrated by prior work; however, our model is distinctive because it can be applied out-of-sample to the entire past and future of the Court, not a single term. Our results represent an important advance for the science of quantitative legal prediction and portend a range of other potential applications.

<https://journals.plos.org/plosone/article?id=10.1371/journal.pone.0174698>

Problem 4: Societal sustainability

WIRED

Technology | Science | Culture | Gear | Business | Politics | More ▾

Privacy

Co-op is using facial recognition tech to scan and track shoppers

Branches of the Southern Co-op are using facial recognition to look for potential shoplifters. The roll-out raises concerns about the creep of surveillance tech in the private sector

<https://www.wired.co.uk/article/coop-facial-recognition> (2020)

NL#TIMES

TOP STORIES HEALTH CRIME POLITICS BUSINESS TECH



Jumbo - Credit: Jumbo / Jumbo

CRIME TECH INNOVATION JUMBO SHOPLIFTING AI » MORE TAGS

SUNDAY, 11 FEBRUARY 2024 - 08:15

SHARE THIS:



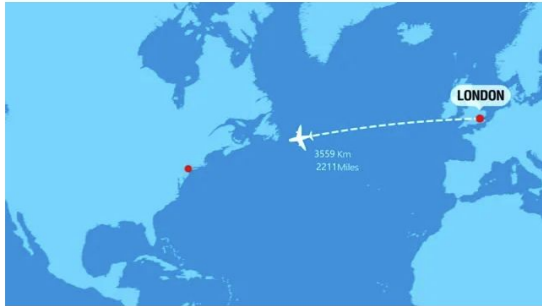
Jumbo takes extra measures against self-scan shoplifting

Supermarket chain Jumbo will take extra measures to stop shoplifting over the next few weeks. For example, there will be more clearly visible camera surveillance and more and smarter random checks at self-checkouts. There will also be extra communication to customers that they must pay for all groceries and that they will be

<https://nltimes.nl/2024/02/11/jumbo-takes-extra-measures-self-scan-shoplifting> (2024)

Problem 5: Environmental sustainability

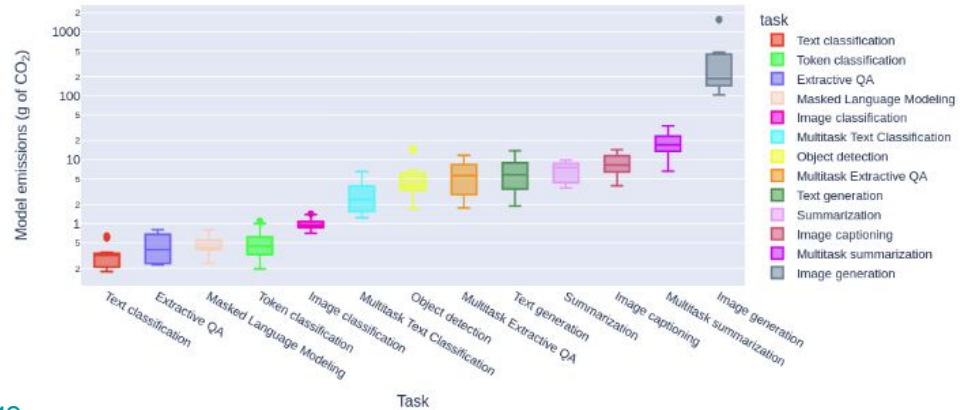
Their research revealed that BLOOM's training led to 25 metric tons of carbon dioxide emissions, doubling when considering the broader life cycle. While this may seem like a lot — **equivalent to around 60 flights between London and New York** — it's significantly less than other LLMs of the same size, primarily due to BLOOM being trained on a French supercomputer powered mostly by nuclear energy.



Comparatively, OpenAI's GPT-3 and Meta's OPT were estimated to emit over 500 and 75 metric tons of carbon dioxide, respectively, during training. These estimates, however, are based on limited data or external assessments, highlighting the need for a standardized method of measuring emissions.

training an LLM:
between 75 to 500
tons of CO₂

using an image generator:
0.2-0.5Kg CO₂ per 1000 queries



<https://medium.com/@thesab/unmasking-the-dirty-secret-of-large-language-models-their-carbon-footprint-9bac7ae2da5e>

<https://arxiv.org/pdf/2311.16863.pdf>

A continuous progress?

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

A continuous progress?

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



Google
DeepMind (2016)



Microsoft (2018)



Uber (2019)



Google AlphaStar (2019)



OpenAI GPT-3 (2020)



OpenAI DALL·E (2021)



Google
Bart/Gemini (2023-...)



Claude (2023-...)



Meta
Llama (2023-...)



OpenAI GPT-4 (2023)
DALL·E 3
Sora (2024)



DeepSeek (2025)

A continuous progress?

- By using a mixture of ML techniques, several human or super-human performances are achieved every year in specific tasks (*mostly by US corporate-driven research, but now also China...*).
- More and more sensitive applications are being researched and deployed in the wild. All big players are strongly involved in vertical integration (hardware, physical resources, talent acquisition).
Geopolitical matters have become prominent (and rightly so).

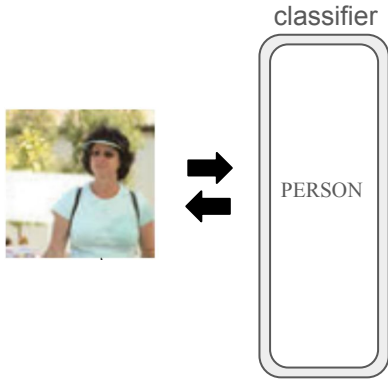
A continuous progress?

- By using a mixture of ML techniques, several human or super-human performances are achieved every year in specific tasks (*mostly by US corporate-driven research, but now also China...*).
- More and more sensitive applications are being researched and deployed in the wild. All big players are strongly involved in vertical integration (hardware, physical resources, talent acquisition).
Geopolitical matters have become prominent (and rightly so).
- Yet, problems of **generalization, explainability, transparency, responsibility, fairness...** are still there.

Conclusions

AI is much wider than ML

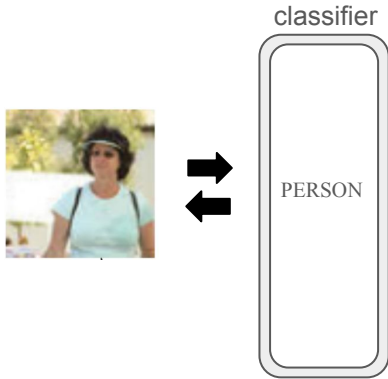
AUTOMATIC



ML *internalizes*
("mirrors") behaviour
associated to the task

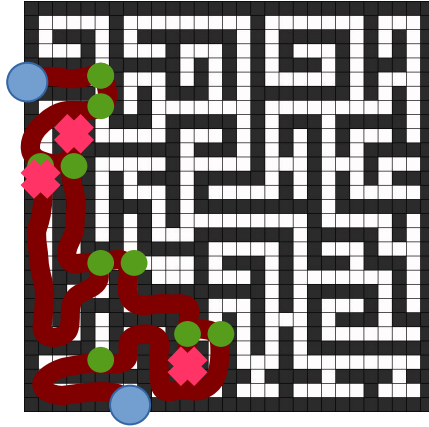
AI is much wider than ML

AUTOMATIC



ML *internalizes*
("mirrors") behaviour
associated to the task

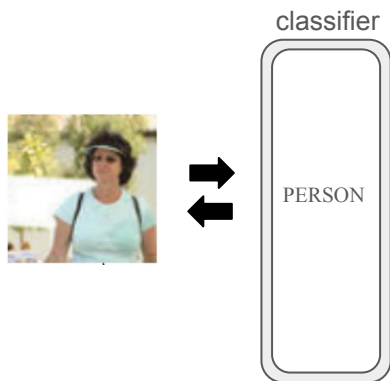
AUTOMATED



Symbolic methods
mechanize search
to find solutions

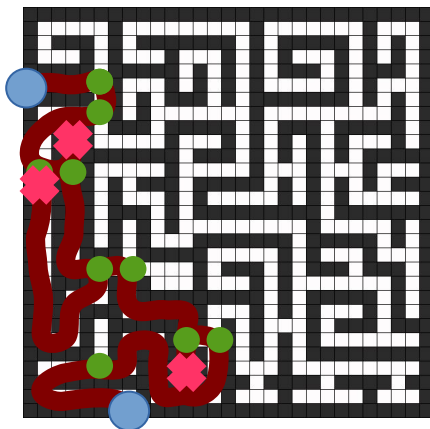
AI is much wider than ML

AUTOMATIC



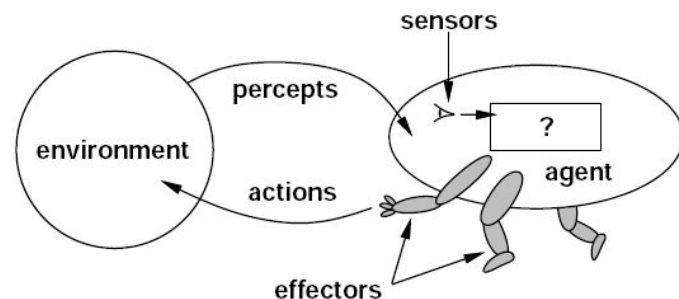
ML *internalizes*
("mirrors") behaviour
associated to the task

AUTOMATED



Symbolic methods
mechanize search
to find solutions

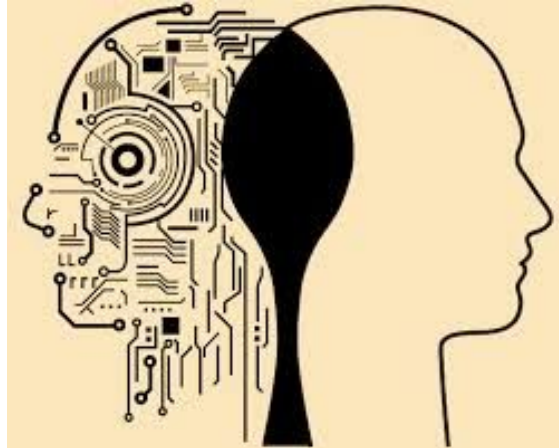
(partially) AUTONOMOUS



Agents creates its own
goals/policies to achieve
higher-order goals

No AGI in view

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



No AGI in view

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



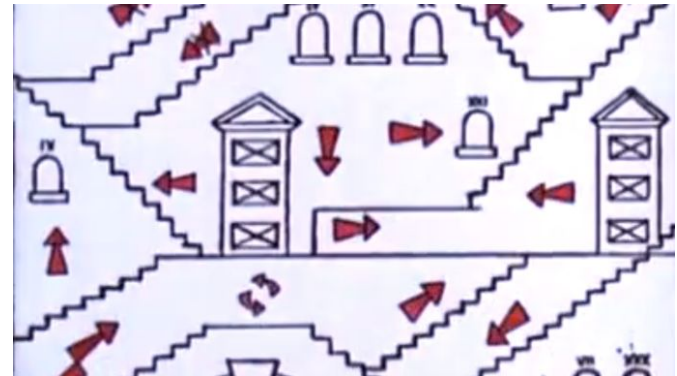
These pieces might be simple or not, it is the ML method that is not satisfactory to design them. Plausibly a mixture of methods (symbolic, cognitive modeling), offer a sounder basis.

“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.

- lock-in with geopolitical and functional dependencies
- socio-economic disruptions
- loss of control, understanding and ownership
- cognitive deskilling



“Outperforming” humans



Touching numerals from 1 to 9

Masking task

Limited-hold memory task (Human, 5 numerals)

<https://langint.pri.kyoto-u.ac.jp/ai/en/publication/SanaInoue/Inoue2007.html>

“Outperforming” humans



training + sufficient memory... can we call it intelligence?

If we are pursuing **rationality**
(*rational systems, rational
institutions, etc.*),
it is rather **implausible** that this
will be obtained only by
empirical means.

An Introduction to Artificial Intelligence

MASS Rotterdam – Conference on *The Legal Future of Autonomous Vessels*

Erasmus School of Law – 19 June 2025

Giovanni Sileno

g.sileno@uva.nl

