COMP9208
Artificial Intelligence and Society

Lecture 2: 15 August 2025

Prof. Mikhail Prokopenko
Dr. Sheryl Chang

Canvas: https://canvas.sydney.edu.au/courses/66452

Email: mikhail.prokopenko@sydney.edu.au
Office: J12 (Computer Science), level 3W, room 324

Date	Week: Workshop	Assignment
08/08	1: History of AI	
15/08	2: History of AI + Perception and Actuation	
22/08	3: History of AI + Representation and Reasoning	
29/08	4: Machine Learning	
05/09	5: Natural Interaction	
12/09	6: Distributed AI (swarm intelligence)	
19/09	7: From Good Old-Fashioned AI (GOFAI) to Artificial General Intelligence (AGI)	# 1: Individual assignment (article review)
26/09	8: Goal-driven AI: search & rescue vs lethal autonomous weapon systems (LAWS)	
03/10	**BREAK**	
10/10	9: Review: language models, search trees, situated behaviours, neural networks	
17/10	10: Robotic swarms: RoboCop vs RoboCup	
24/10	11: Group Presentations	# 2: Group assignment
31/10	12: Adversarial Machine Learning	
07/11	13: Limits of cognition	
17/11	14: STUVAC	# 3: Individual assignment (data analysis report)

Brief review of last week

?

Definitions of Artificial Intelligence (AI)

- the ability of a digital computer or computer-controlled robot to perform tasks commonly associated with intelligent beings (Britannica)

- intelligence exhibited by machines or software. It is also the name of the scientific field which studies how to create computers and computer software that are capable of intelligent behaviour (Wikipedia)

- “the science and engineering of making intelligent machines” (John McCarthy)

- technology that enables computers and machines to simulate human intelligence and problem-solving capabilities (IBM)

- the ability of a computer system to perform tasks that would normally require human intelligence, such as learning, reasoning, and making decisions (DigitalNSW)

- the capacity of machines to mimic human cognitive functions such as learning, problem-solving, and pattern recognition, enabling them to perform tasks that normally require human intelligence (University of North Florida)

- the simulation of human intelligence processes by machines, especially computer systems (ChatGPT 3.5)

Artificial Intelligence (AI): socio-economic impacts

- "It is difficult to think of a major industry that AI will not transform. This includes healthcare, education, transportation, retail, communications, and agriculture. There are surprisingly clear paths for AI to make a big difference in all of these industries." (Andrew Ng)
- "It's very clear that AI is going to impact every industry. I think that every nation needs to make sure that AI is a part of their national strategy. Every country will be impacted." (Jensen Huang)
- “We don’t quite realize the impact that this is going to have in businesses and at work because it is just starting. But what we’ve seen so far is that over a very short amount of time, these AI systems have entered the workforce as collaborators.” (Mira Murati)
- “There are many valid concerns about AI, from its impact on jobs to its uses in autonomous weapons systems and even to the potential risk of superintelligence.“ BUT: "AI is a tool. The choice about how it gets deployed is ours." (Oren Etzioni)
- “If the government regulates against the use of drones or stem cells or artificial intelligence, all that means is that the work and the research leave the borders of that country and go someplace else.” (Peter Diamandis)
- “I do worry that organizations and even governments who own AI and data will have a competitive advantage and power, and those who don't will be left behind." (Rana el Kaliouby)

History of AI: main achievements and setbacks

Birth of artificial intelligence (1941 – 1956)

- neurology: the brain was found to be an electrical network of neurons that fired in all-or-nothing pulses.

- cybernetics (Norbert Wiener): methods to attain control and stability in electrical networks

- information theory (Claude Shannon): quantified how to store, process and communicate digital signals (i.e., all-or-nothing signals)

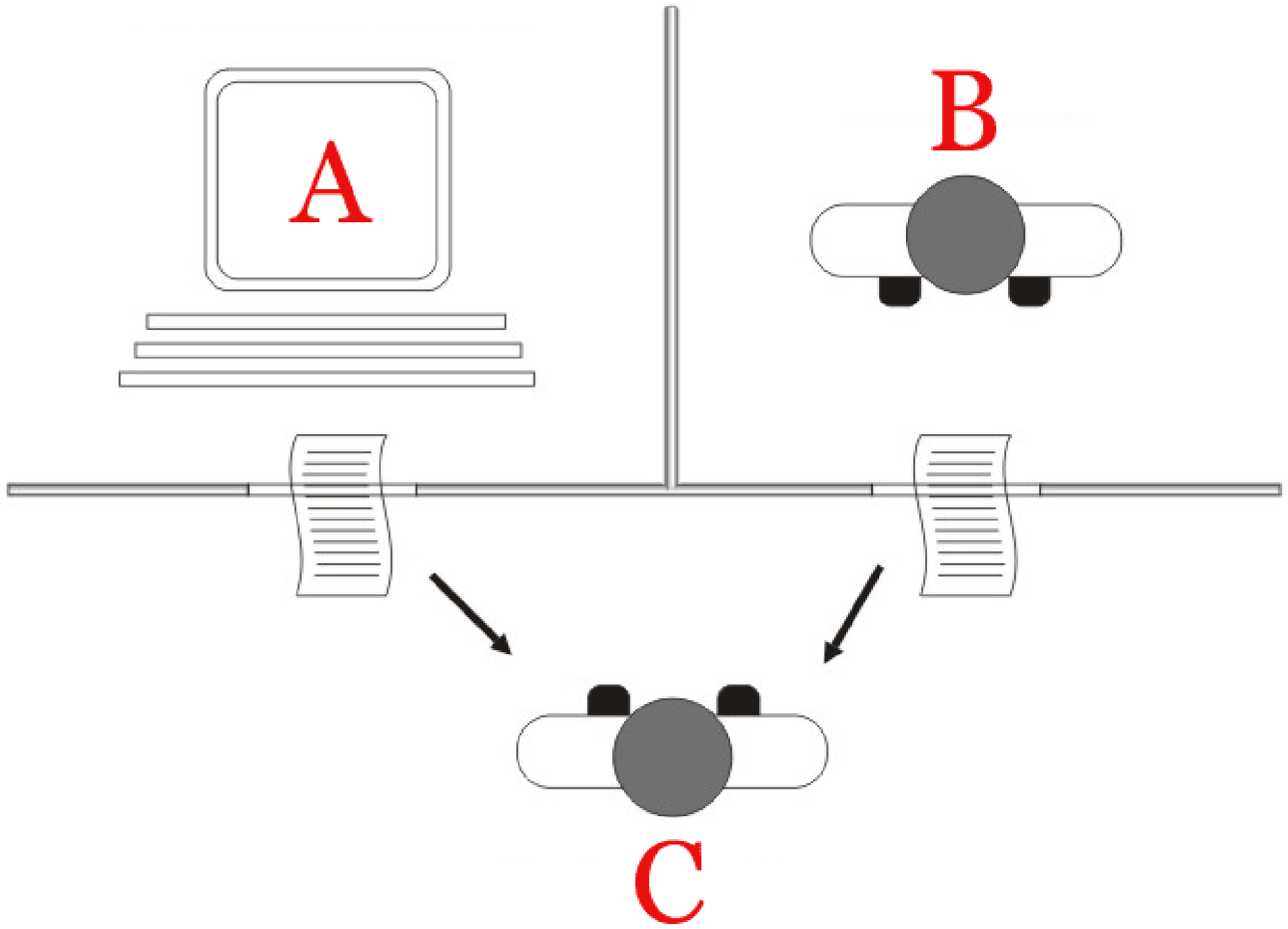
- theory of computation (Alan Turing): any form of computation can be described digitally

- …it may be possible to construct an "electronic brain“

- 1956: "Artificial intelligence research" founded as an academic discipline

Turing test

If a machine could carry on a conversation (over a teleprinter) that was indistinguishable from a conversation with a human being, then it was reasonable to say that the machine was "thinking".

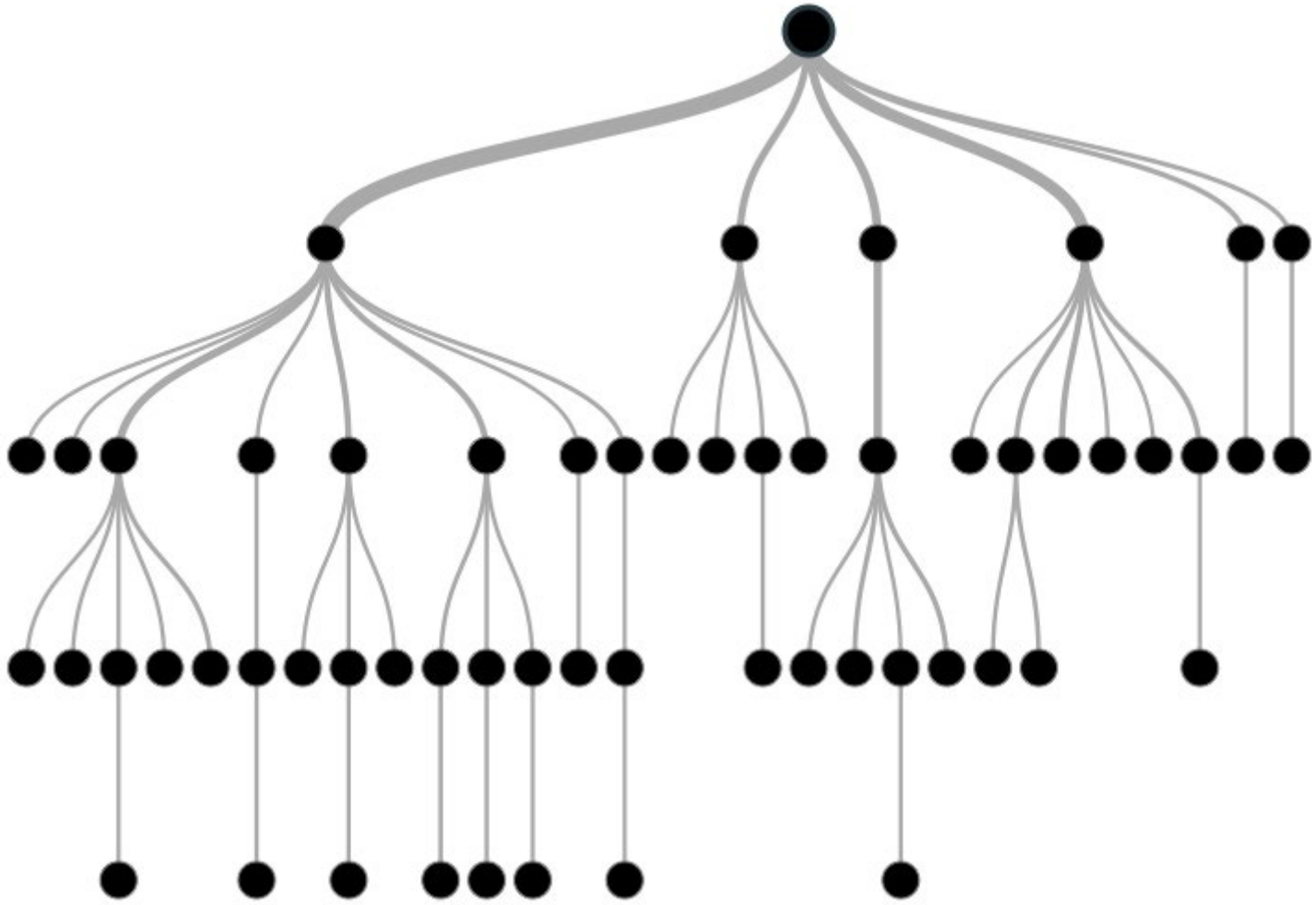


https://en.wikipedia.org/wiki/Turing_test

History of AI: main achievements and setbacks

Early successes (1956 – 1974)

- “Reasoning as search”:
  - the problem-solving process is framed as a search through a space of potential solutions, where each potential solution is represented as a state in the “search-space”,
  - the goal is to find a path to a solution by iteratively exploring and evaluating different states in the search space

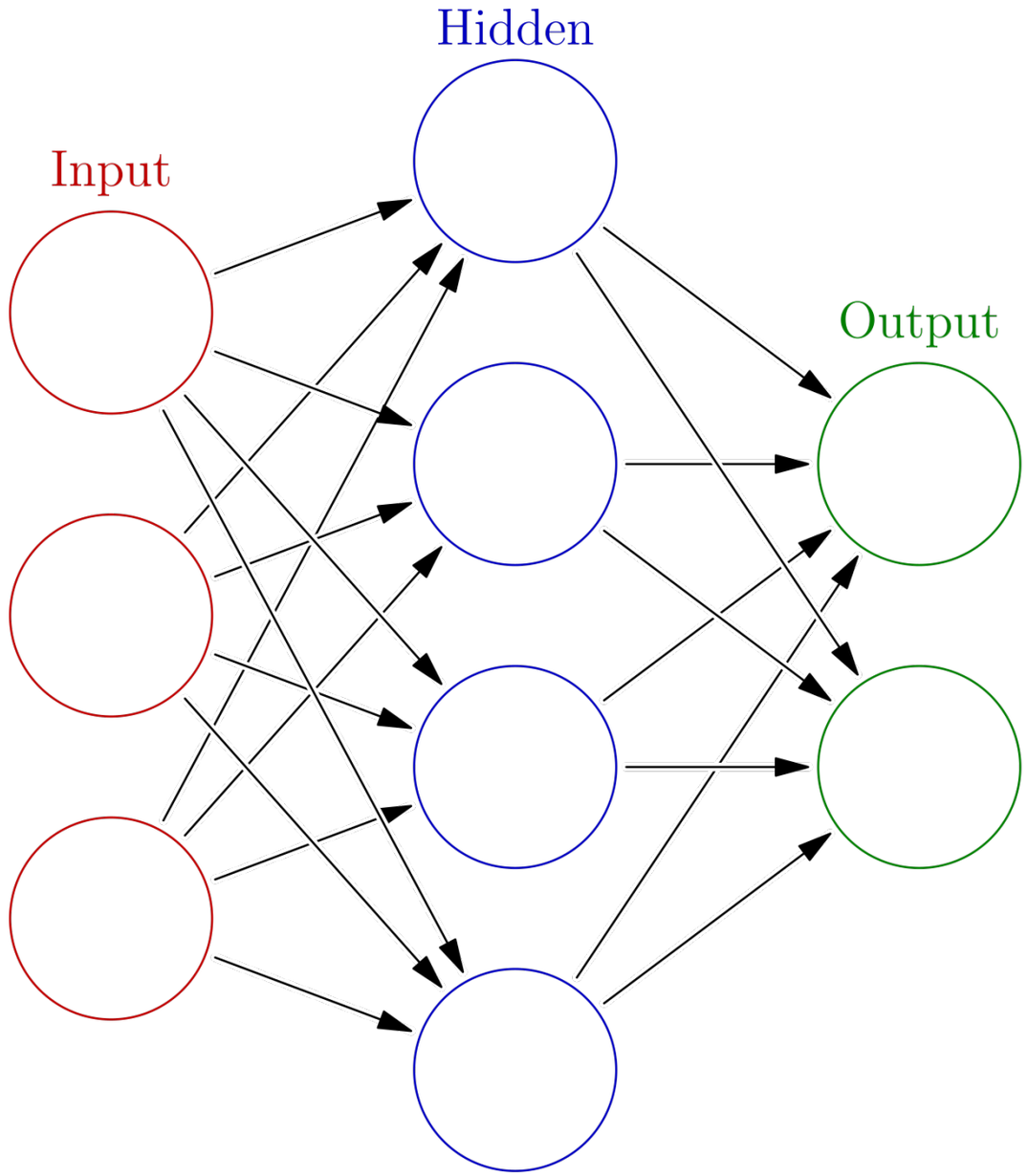


https://www.justinmath.com/the-first-wave-of-ai-reasoning-as-search/

History of AI: main achievements and setbacks

Early successes (1956 – 1974)

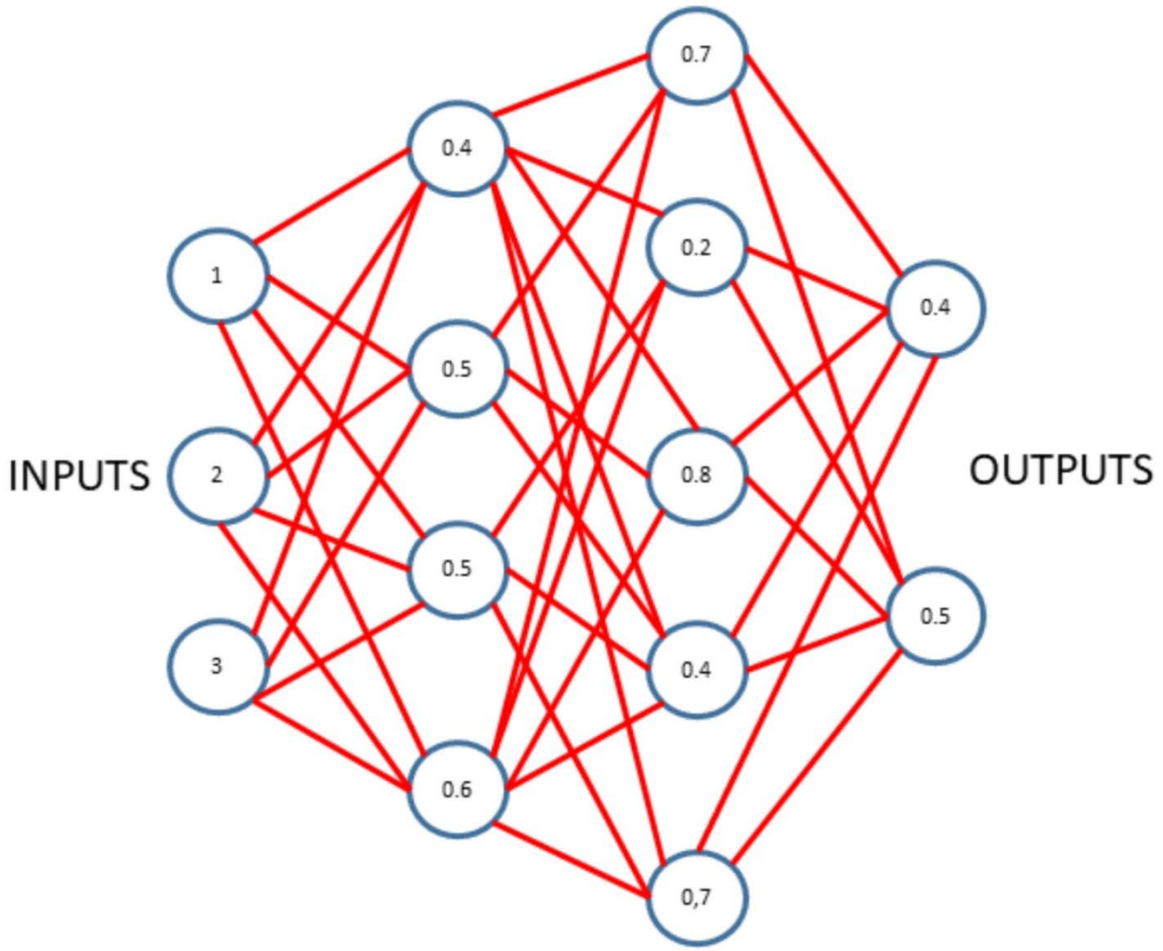
- Artificial neural networks:
  - interconnected nodes (artificial neurons), linked by edges,
  - each artificial neuron receives signals from neighbouring neurons, combines them using a non-linear function of the sum of its inputs, and sends the result (a number) to other neighbouring neurons.
  - the strength of the signal at each connection is determined by a weight, which is adjusted during the learning process, given some training data



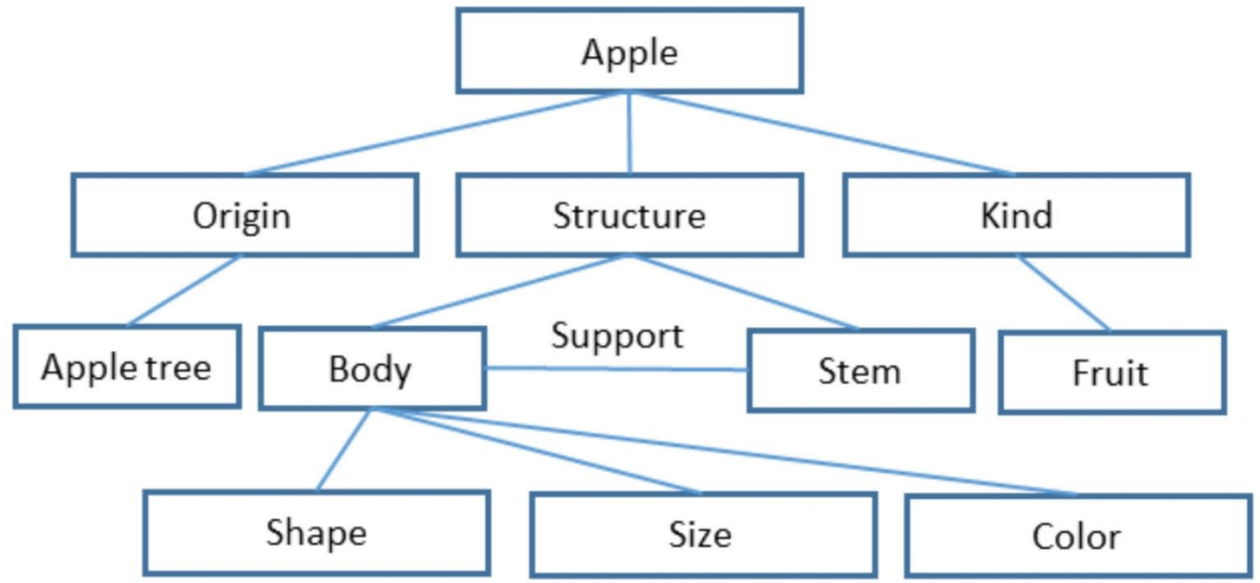
https://en.wikipedia.org/wiki/Neural_network_(machine_learning)

History of AI: main achievements and setbacks

Connectionist vs Symbolic AI approaches



(a)



(b1)

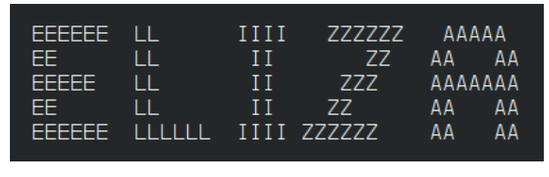
sibilings(X, Y) :- parent(Z, X) and parent)Z,Y)

(b2)

https://link.springer.com/chapter/10.1007/978-3-030-35664-4_13

History of AI: main achievements and setbacks

Early successes (1956 – 1974)



- Natural Language:
  - ELIZA was a computer program developed during 1964–1967 by Joseph Weizenbaum, MIT
  - ELIZA simulated conversation as a psychotherapist, by using a pattern matching and substitution methodology that gave users an illusion of understanding the conversation
  - ELIZA examined the text for keywords, applied values to identified keywords, and transformed the input into an output by using "decomposition rule" and "reassembly rule"

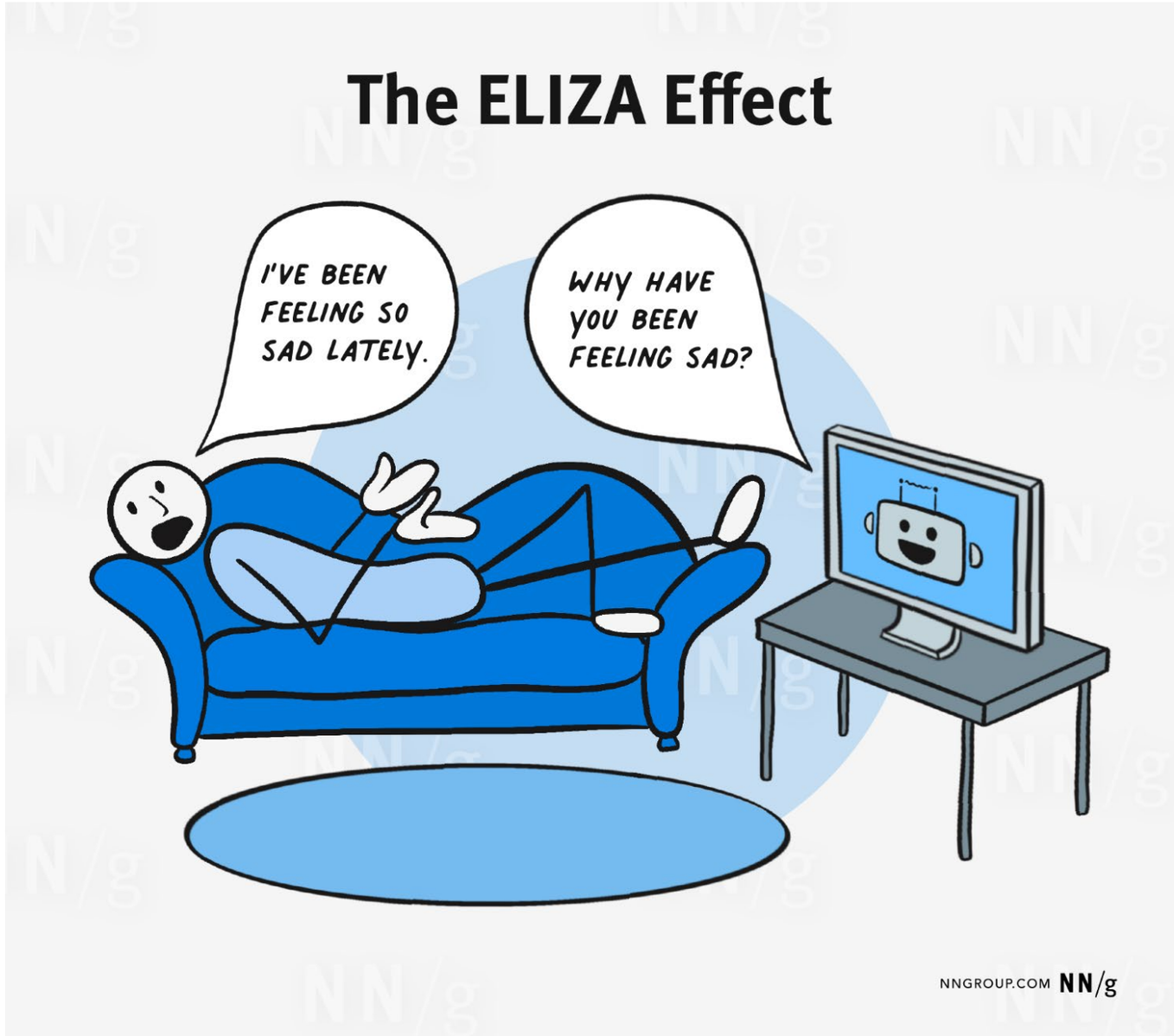
```
ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your boyfriend made you come here ?
YOU:   He says I'm depressed much of the time.
ELIZA: I am sorry to hear that you are depressed.
YOU:   It's true. I am unhappy.
ELIZA: Can you explain what made you unhappy ?
YOU:
```

https://en.wikipedia.org/wiki/ELIZA

History of AI: main achievements and setbacks

Ethical concerns about ELIZA

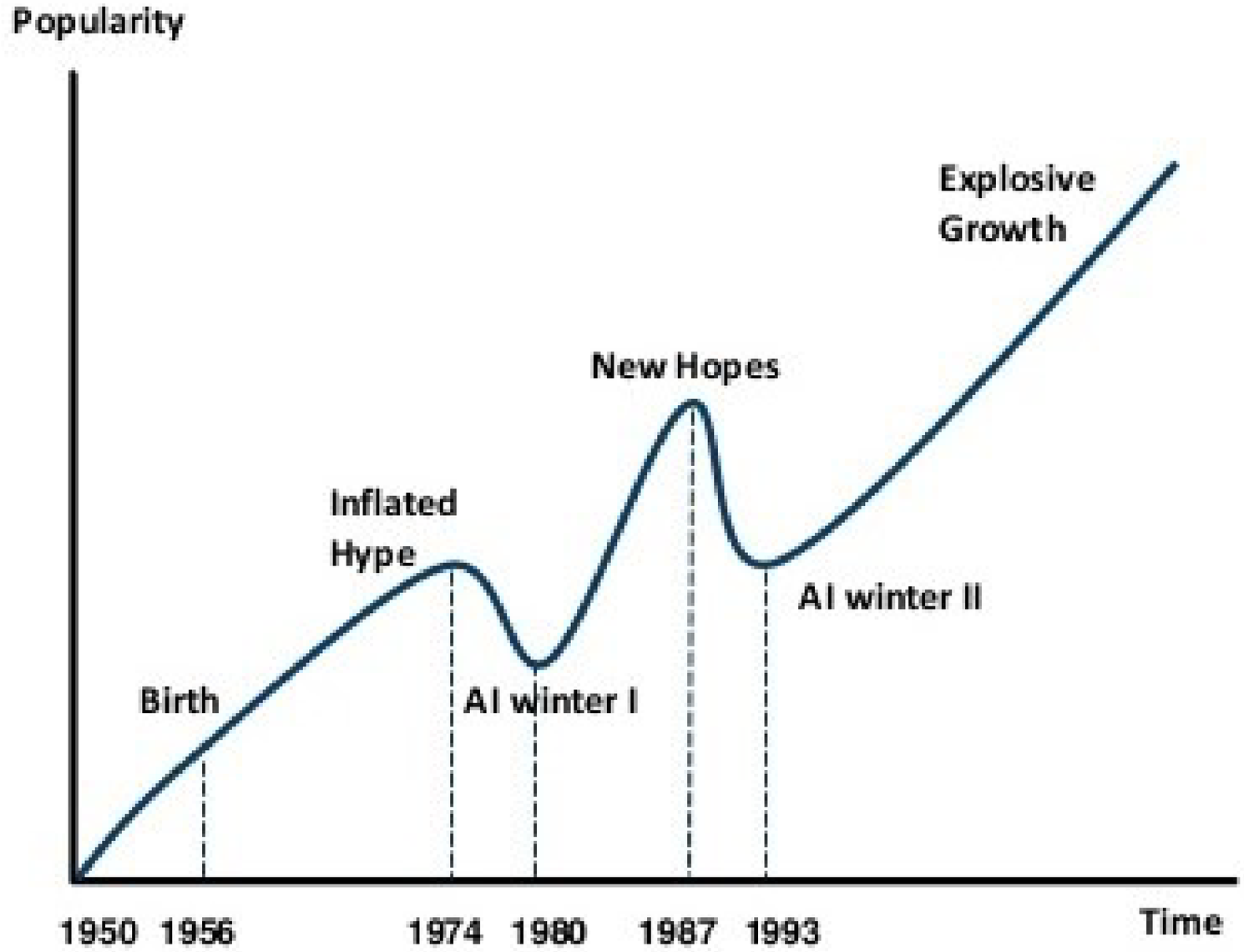
- "extremely short exposures to a relatively simple computer program could induce powerful delusional thinking in quite normal people“ (Joseph Weizenbaum)
- ELIZA effect: the tendency to project human traits — experience, understanding or empathy — into computer programs with a textual interface
- users assumed that ELIZA's questions implied interest and emotional involvement
- social engineering rather than explicit AI programming could pass a Turing test
- when human behavior is programmed into chatbots, gender stereotypes may appear



https://en.wikipedia.org/wiki/ELIZA_effect
https://www.nngroup.com/articles/eliza-effect-ai/

History of AI: main achievements and setbacks

AI HAS A LONG HISTORY OF BEING "THE NEXT BIG THING"...



Timeline of AI Development

- **1950s-1960s**: First AI boom - the age of reasoning, prototype AI developed
- **1970s**: AI winter I
- **1980s-1990s**: Second AI boom: the age of Knowledge representation (appearance of expert systems capable of reproducing human decision-making)
- **1990s**: AI winter II
- **1997**: Deep Blue beats Gary Kasparov
- **2006**: University of Toronto develops Deep Learning
- **2011**: IBM's Watson won Jeopardy
- **2016**: Go software based on Deep Learning beats world's champions

https://www.actuaries.digital/2018/09/05/history-of-ai-winters/

Questions

?

Homework

- check the information sources provided on different slides (at least some)
- many names were mentioned today (AI researchers, scientists, entrepreneurs, writers) – please try to find more information about these people
- several movies about AI were mentioned as well – please try to watch at least one before the next lecture, and prepare to give a few brief comments about the movie:
  - what was most striking about the AI technology in the movie?
  - what societal impact did the AI make?
  - what were the main AI limitations, biases and risks?

?

Quick quiz

- Alan Turing:
- Norbert Wiener:
- Claude Shannon:
- John McCarthy:
- Herbert A. Simon:
- Allen Newell:
- Marvin Minsky:
- Joseph Weizenbaum:

Quick quiz

- **Alan Turing**: mathematician, computer scientist and logician, considered to be the father of theoretical computer science and artificial intelligence
- **Norbert Wiener**: computer scientist, mathematician and philosopher, the founder of cybernetics, one of the first to theorise that all intelligent behavior is the result of feedback mechanisms
- **Claude Shannon**: mathematician, electrical engineer, computer scientist and cryptographer known as the "father of information theory", built the first machine learning device, Theseus.
- **John McCarthy**: computer scientist, one of the founders of Artificial Intelligence
- **Herbert A. Simon**: political scientist focussed on decision-making; received the Nobel Memorial Prize in Economic Sciences in 1978 and the ACM's Turing Award in computer science in 1975
- **Allen Newell**: researcher in computer science and cognitive psychology; received the ACM's Turing Award, along with Herbert A. Simon, in 1975 for their contributions to artificial intelligence and the psychology of human cognition
- **Marvin Minsky**: cognitive and computer scientist, co-founded the MIT's AI laboratory, proposed frames as a framework for knowledge representation
- **Joseph Weizenbaum**: computer scientist, created an early natural language processing computer program ELIZA, one of the first chat-bots

Homework: some answers

- Geoffrey Hinton:
- Oren Etzioni:
- Fei-Fei Li:
- Sam Altman:
- Mira Murati:
- Rene Haas:
- Andrew Ng:
- Jensen Huang:
- Rana el Kaliouby:
- Kristalina Georgieva:

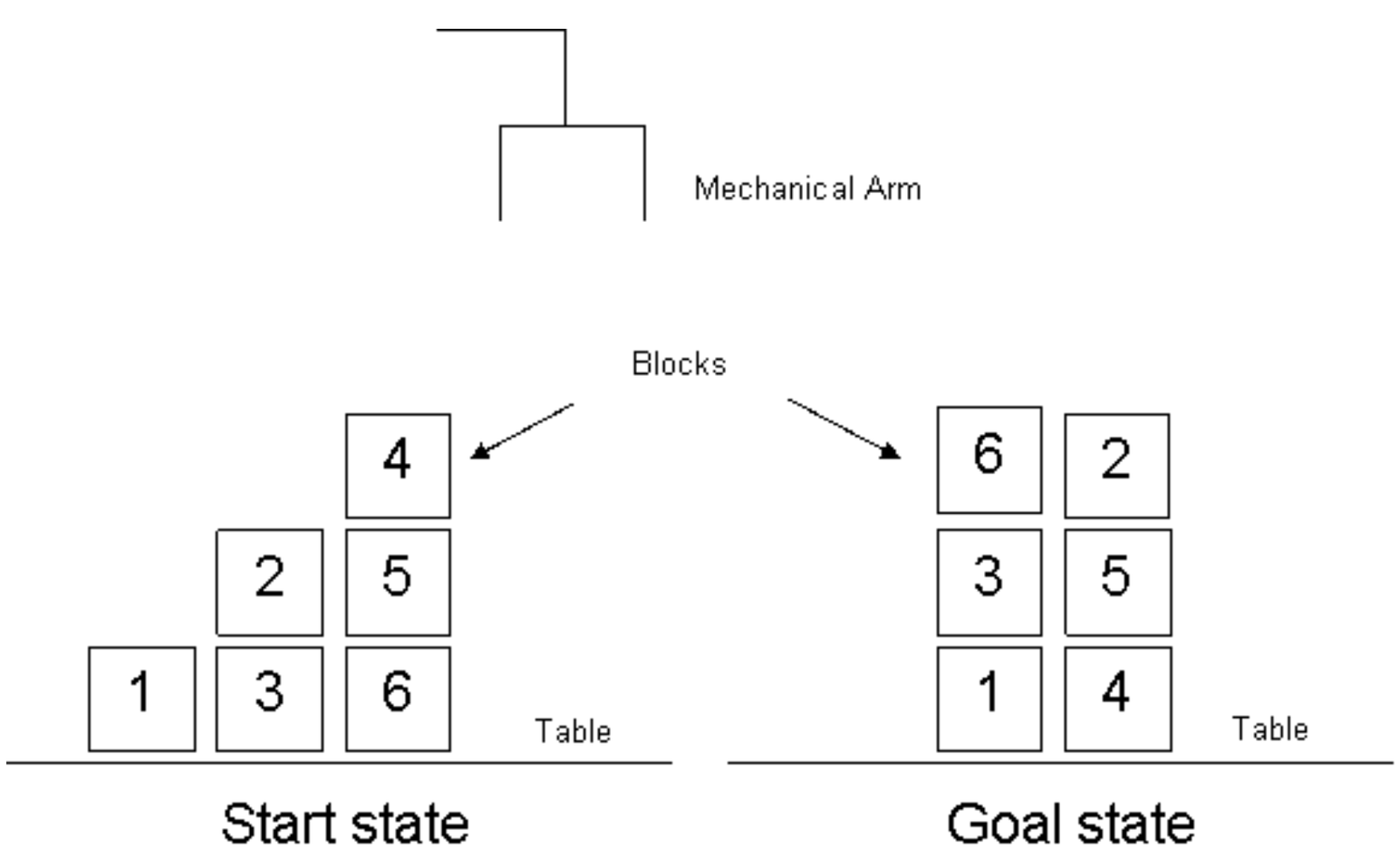
Homework: some answers

- **Geoffrey Hinton**: computer scientist and cognitive psychologist, most noted for his work on artificial neural networks and deep learning
- **Oren Etzioni**: computer scientist, entrepreneur, founding CEO of the Allen Institute for Artificial Intelligence
- **Fei-Fei Li**: computer scientist, known for establishing ImageNet, the dataset that enabled rapid advances in computer vision in the 2010s
- **Sam Altman**: entrepreneur and investor, the CEO of OpenAI since 2019
- **Mira Murati:** engineer, researcher and tech executive, the CTO of OpenAI since 2018 (left in 2024)
- **Rene Haas**: engineer and tech executive, the CEO of Arm Holdings since 2022
- **Andrew Ng**: computer scientist, technology entrepreneur in machine learning and AI, a co-founder and head of Google Brain, the former Chief Scientist at Baidu
- **Jensen Huang**: businessman, electrical engineer, and the co-founder, president and CEO of Nvidia
- **Rana el Kaliouby**: computer scientist, co-founder and the CEO of Affectiva
- **Kristalina Georgieva**: International Monetary Fund Managing Director

History of AI: main achievements and setbacks

Early successes (1956 – 1974)

- Micro-worlds:
  - artificial simple situations (toy problems)
  - a "blocks world" with coloured blocks of various shapes and sizes placed on a flat surface
  - a mechanical robot arm has to pick and place the blocks
  - automated planning and scheduling



https://www.rose-hulman.edu/class/cs/archive/other-old/archive/fall06/materials/search_project.htm

History of AI: main achievements and setbacks

Early optimism (1956 – 1974)

- 1958: "within ten years a digital computer will be the world's chess champion" and "within ten years a digital computer will discover and prove an important new mathematical theorem.“ (Herbert A. Simon and Allen Newell)

- 1965: "machines will be capable, within twenty years, of doing any work a man can do.“ (Herbert A. Simon)

- 1967: "Within a generation ... the problem of creating 'artificial intelligence' will substantially be solved.“ (Marvin Minsky)

- 1970: "In from three to eight years we will have a machine with the general intelligence of an average human being.” (Marvin Minsky)

Questions

?

History of AI: main achievements and setbacks

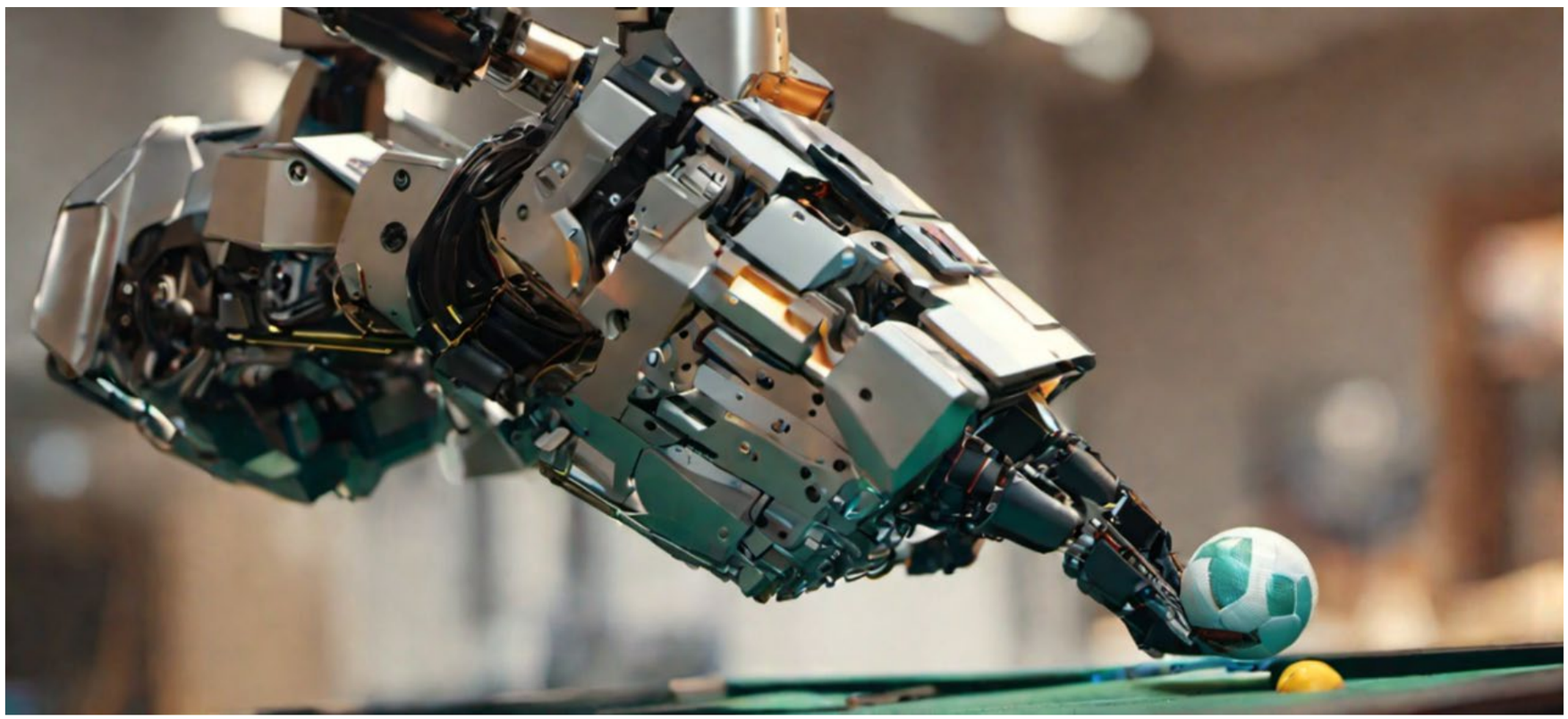
First AI winter (1974 – 1980)

- Limited computer power: there was not enough memory or processing speed to complete practical tasks
- Intractability and the combinatorial explosion: many AI solutions of the "toy problems" did not scale up into useful systems
- Commonsense knowledge and reasoning: many AI applications required huge amounts of information about the world
- Moravec's paradox: "it is comparatively easy to make computers exhibit adult level performance on intelligence tests or playing checkers, and difficult or impossible to give them the skills of a one-year-old when it comes to perception and mobility"
- The frame and qualification problems: how to describe default commonsense assumptions in a virtual environment

https://www.linkedin.com/pulse/unraveling-moravecs-paradox-enigma-robots-struggling-everyday-sharma

History of AI: main achievements and setbacks

Moravec's paradox: abstract reasoning needs very little computation, but perception and sensorimotor skills require large computational resources



the robot needs to

- recognize the ball using complex image processing
- calculate the exact position where the ball is placed currently
- predict where the ball will be next
- coordinate its mechanical arm to reach for the ball in the predicted position
- grasp the ball using hand-eye coordination and accurate motor control

https://www.linkedin.com/pulse/unraveling-moravecs-paradox-enigma-robots-struggling-everyday-sharma

History of AI: main achievements and setbacks

How to explain Moravec's paradox

- for AI, the hard problems are easy and the easy problems are hard – why?
- human skills are shaped by evolution based on natural selection
- the evolutionary processes had more time to develop more basic skills like sensory perception; motor skills including object catching or grasping; recognising a face, a voice, or an intention; visualising an object; as well as some social skills like judging motivation of other people or paying attention

  → hence, it's harder to reverse-engineer
- the evolutionary processes had less time to develop more complex skills like abstract thought, mathematical and analytical reasoning, engineering, logic and scientific reasoning in general

  → hence, it's easier to reverse-engineer

https://en.wikipedia.org/wiki/Moravec's_paradox

History of AI: main achievements and setbacks

Fast-forward to RoboCup-2016 (Leipzig, Germany): Standard Platform League

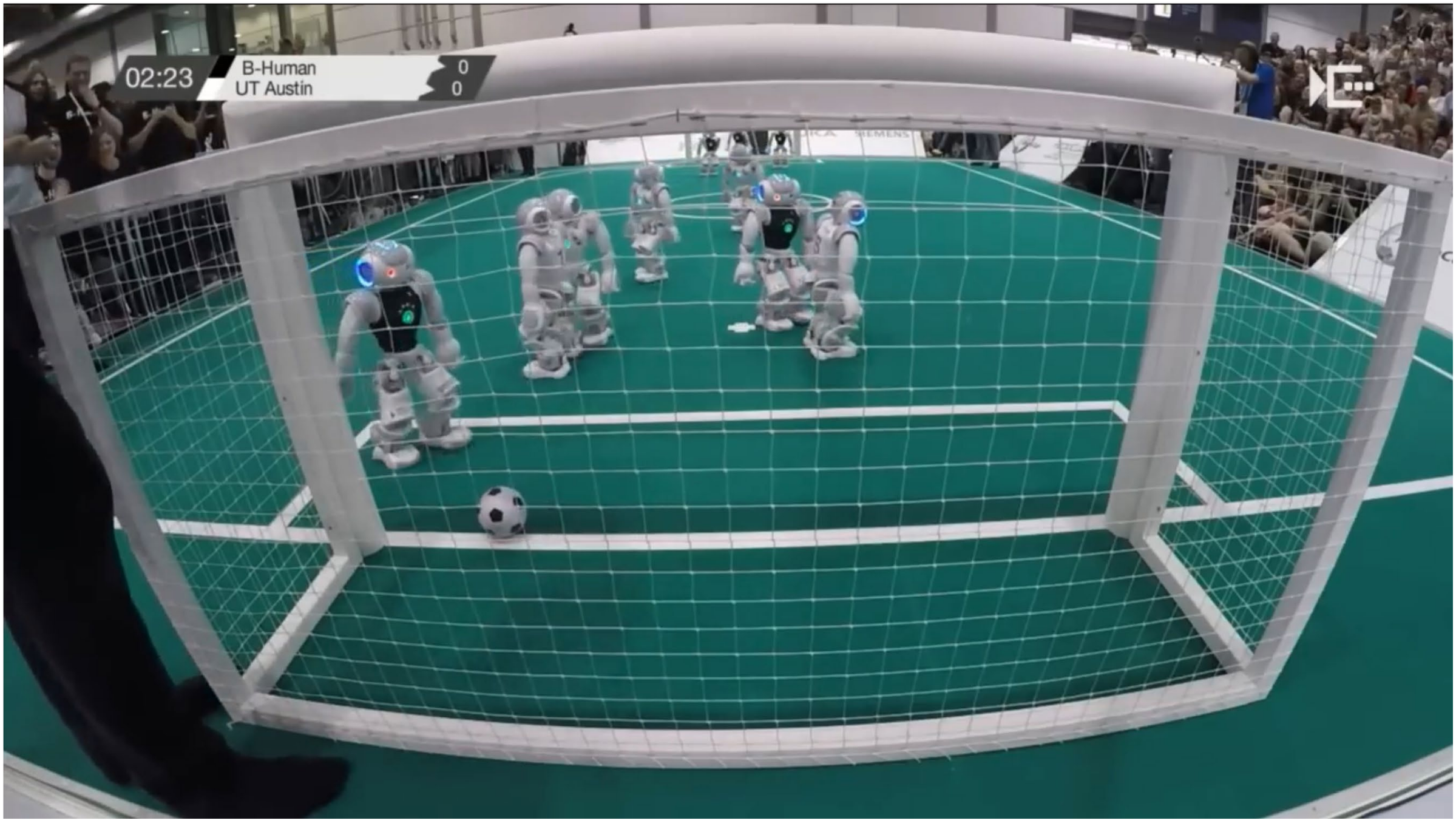
https://www.youtube.com/watch?v=XgRw42oHN-Y

Discussion

?

History of AI: main achievements and setbacks

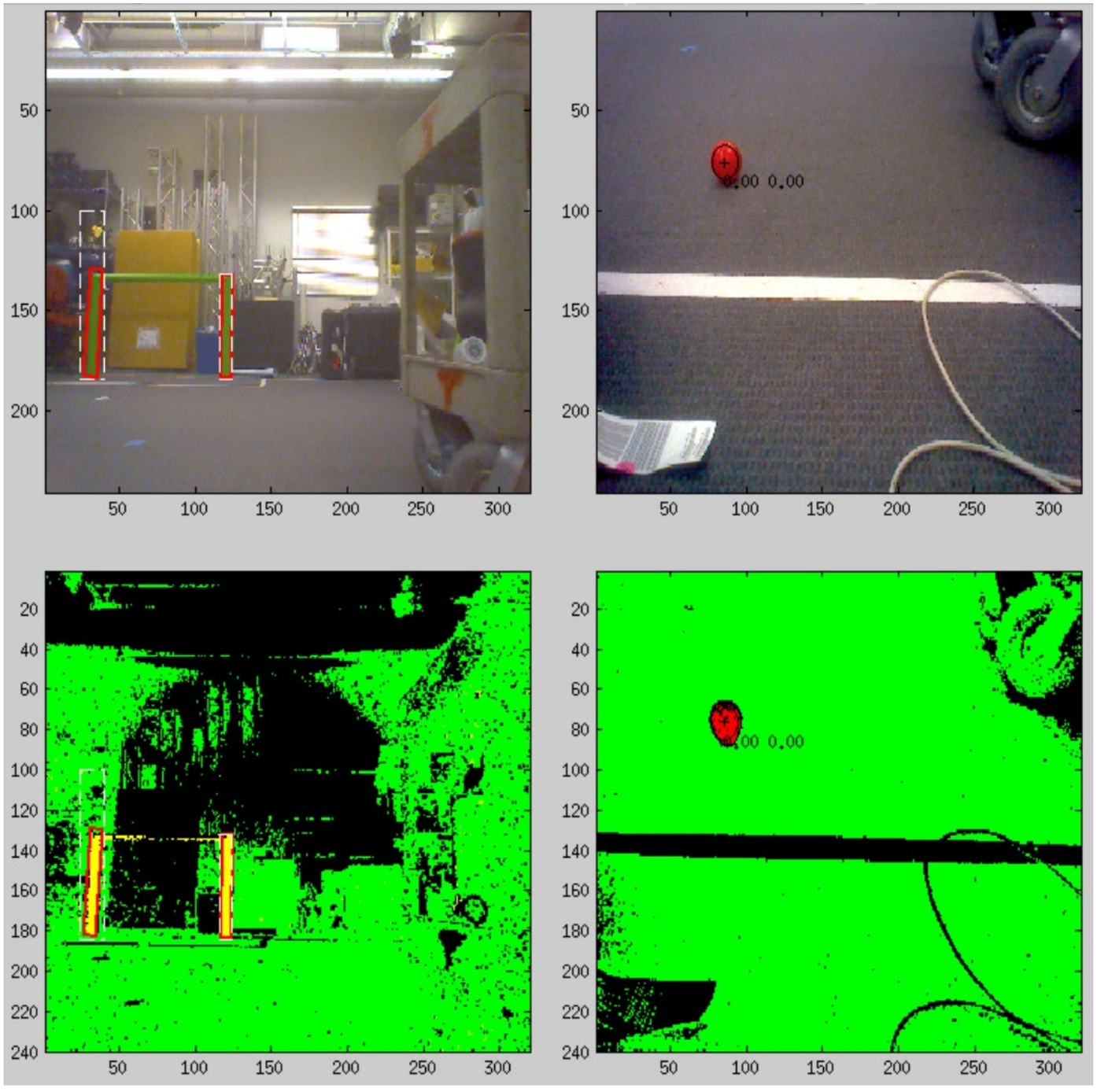
Standard Platform League: image segmentation



https://www.youtube.com/watch?v=XgRw42oHN-Y

History of AI: main achievements and setbacks

Standard Platform League: image segmentation

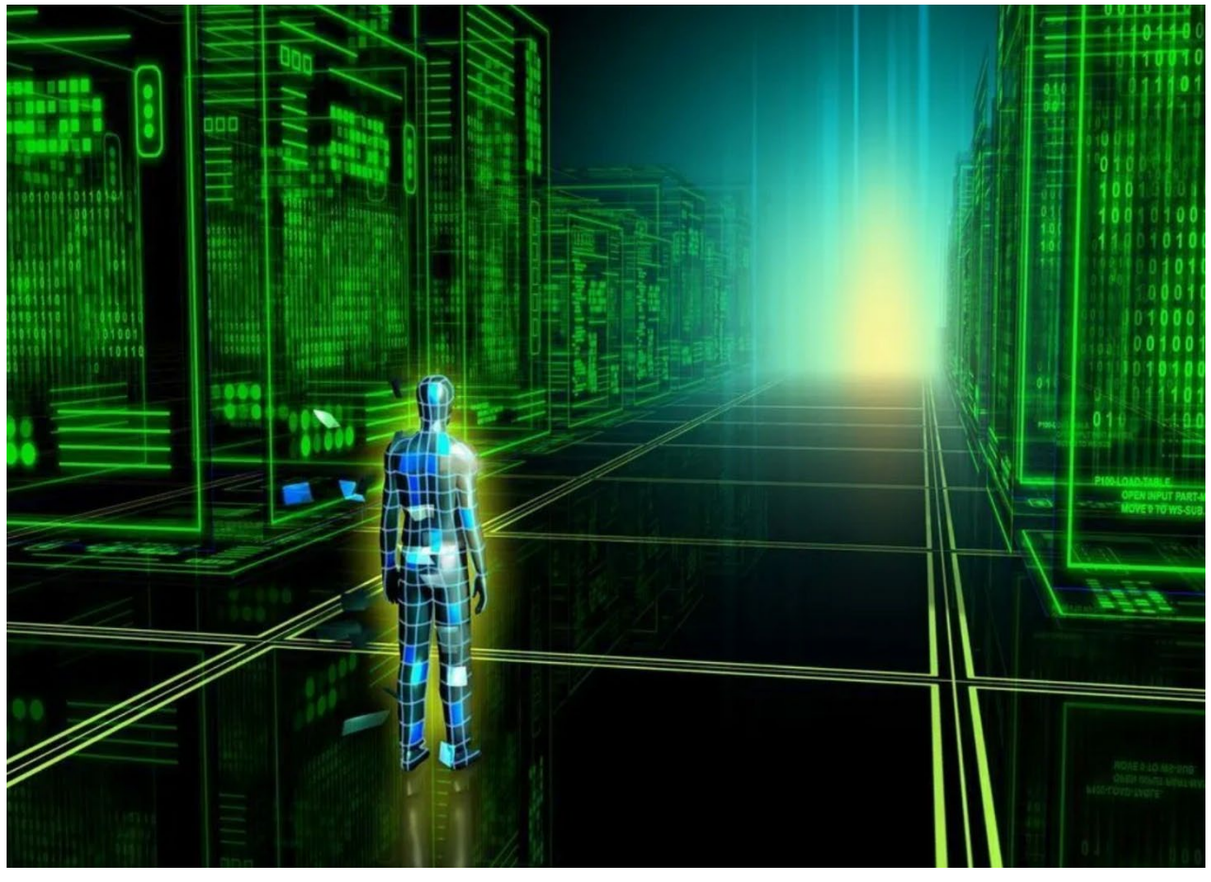


https://fling.seas.upenn.edu/~robocup/files/2014Report.pdf

History of AI: main achievements and setbacks

How to overcome Moravec's paradox

- In the 1980s, a new direction in AI and robotics – "Nouvelle AI“ – was proposed (Rodney Brooks): intelligent machines with "No cognition. Just sensing and action.”
- AI can be trained using massive datasets such as videos and 3D environments to enhance their perceptual tasks
- Simulators can replicate real-world scenarios, helping AI to learn and adapt more efficiently



https://www.linkedin.com/pulse/unraveling-moravecs-paradox-enigma-robots-struggling-everyday-sharma
https://medium.com/simudyne/simulation-the-bedrock-of-ai-12153eaf7971

History of AI: main achievements and setbacks

Fast-forward to RoboCup-2023 (Bordeaux, France): Small Size League

https://www.youtube.com/watch?v=ZLCrtapFmfQ

History of AI: main achievements and setbacks

RoboCup Small Size League: Augmented reality

https://www.youtube.com/watch?v=nAm6R5h48Pc

History of AI: main achievements and setbacks

RoboCup Small Size League: Gazebo simulator

https://www.youtube.com/watch?v=KY_sauKgqxI

History of AI: main achievements and setbacks

RoboCup-2016 (Leipzig, Germany): Simulation 2D League final

https://www.youtube.com/watch?v=h8edvmkYp0A

Break

?

History of AI: main achievements and setbacks

Back to: How to overcome Moravec's paradox

- In the 1980s, a new direction in AI and robotics – "Nouvelle AI“ – was proposed (Rodney Brooks): intelligent machines with "No cognition. Just sensing and action.”
- AI can be trained using massive datasets such as videos and 3D environments to enhance their perceptual tasks
- Simulators can replicate real-world scenarios, helping AI to learn and adapt more efficiently

https://www.linkedin.com/pulse/unraveling-moravecs-paradox-enigma-robots-struggling-everyday-sharma
https://medium.com/simudyne/simulation-the-bedrock-of-ai-12153eaf7971

History of AI: main achievements and setbacks

Socio-economic impact of Moravec's paradox

> “As the new generation of intelligent devices appears, it will be the stock analysts and petrochemical engineers and parole board members who are in danger of being replaced by machines. The gardeners, receptionists, and cooks are secure in their jobs for decades to come.”
>
> (Steven Pinker, The Language Instinct, 1994)

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

History of AI: main achievements and setbacks

Socio-economic impact of Moravec's paradox

> “As the new generation of intelligent devices appears, it will be the stock analysts and petrochemical engineers and parole board members who are in danger of being replaced by machines. The gardeners, receptionists, and cooks are secure in their jobs for decades to come.”
>
> (Steven Pinker, The Language Instinct, 1994)

Socio-economic impact of the first AI winter (1974 – 1980)

The agencies which funded AI research became frustrated with the lack of progress and eventually cut off almost all funding for strategic AI research

- …the misuse of artificial intelligence has the potential to devalue human life
- Deciding is a computational activity, something that can ultimately be programmed
- Choice, however, is the product of judgment, not calculation. Comprehensive human judgment is able to include non-mathematical factors such as emotions

(Joseph Weizenbaum, Computer Power and Human Reason, 1976)

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

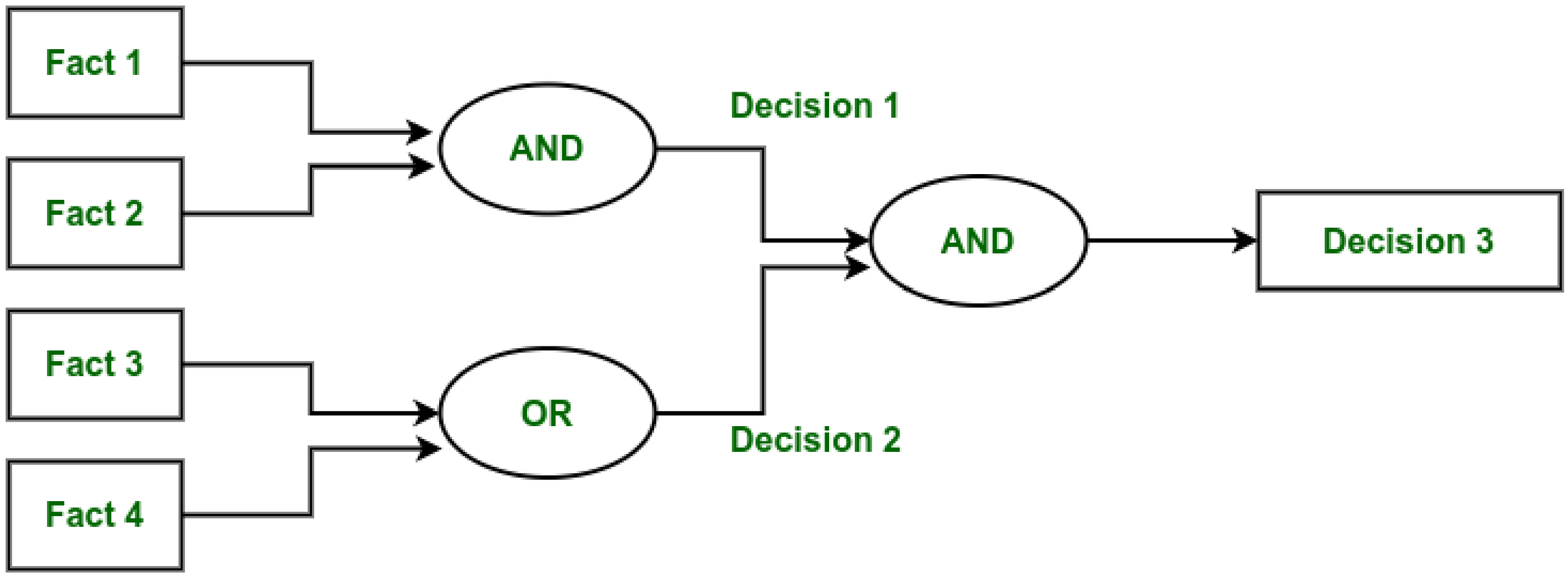
Discussion

?

History of AI: main achievements and setbacks

Boom (1980 – 1987)

- Expert systems:
  - emulates the decision-making ability of a human expert
  - knowledge is represented mainly as if–then rules
  - a knowledge base (facts and rules) and an inference engine (deduction of new facts)
  - can always explain and describe how the problem was solved
  - solves a problem in real-time
  - significantly reduces the cost of expertise for each user

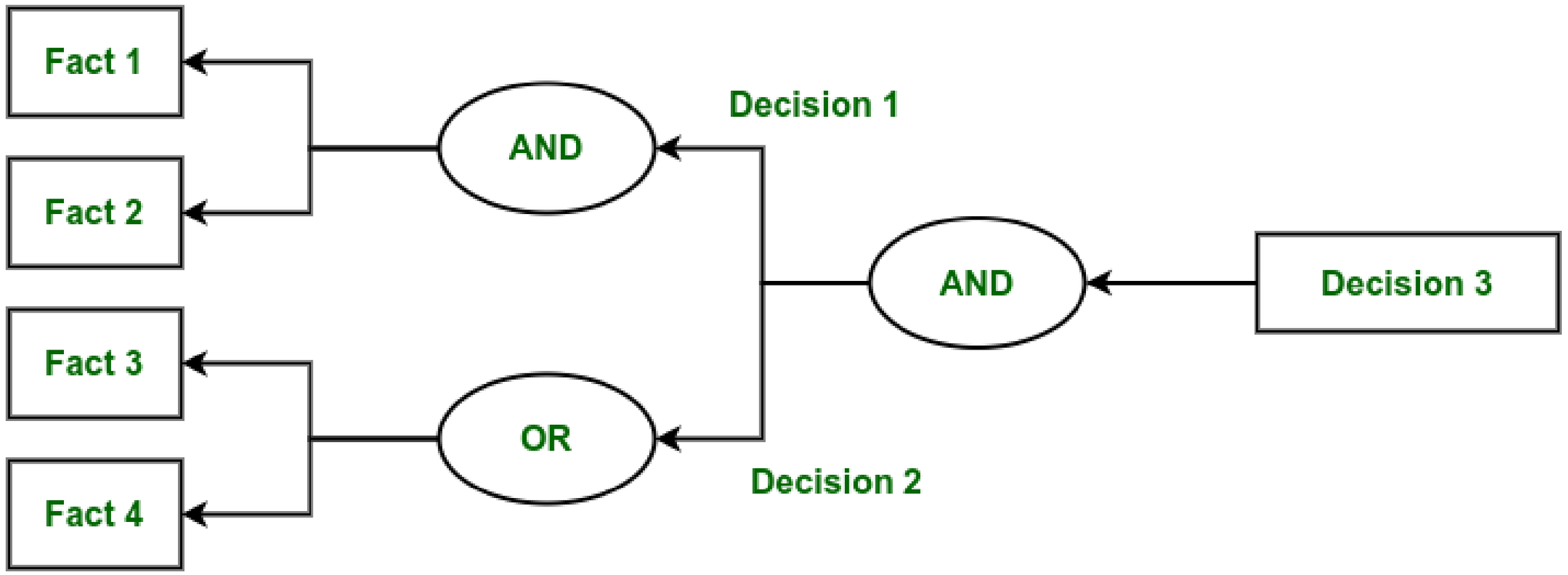


https://www.geeksforgeeks.org/expert-systems/

History of AI: main achievements and setbacks

Boom (1980 – 1987)

- Expert systems:
  - emulates the decision-making ability of a human expert
  - knowledge is represented mainly as if–then rules
  - a knowledge base (facts and rules) and an inference engine (deduction of new facts)
  - can always explain and describe how the problem was solved
  - solves a problem in real-time
  - significantly reduces the cost of expertise for each user

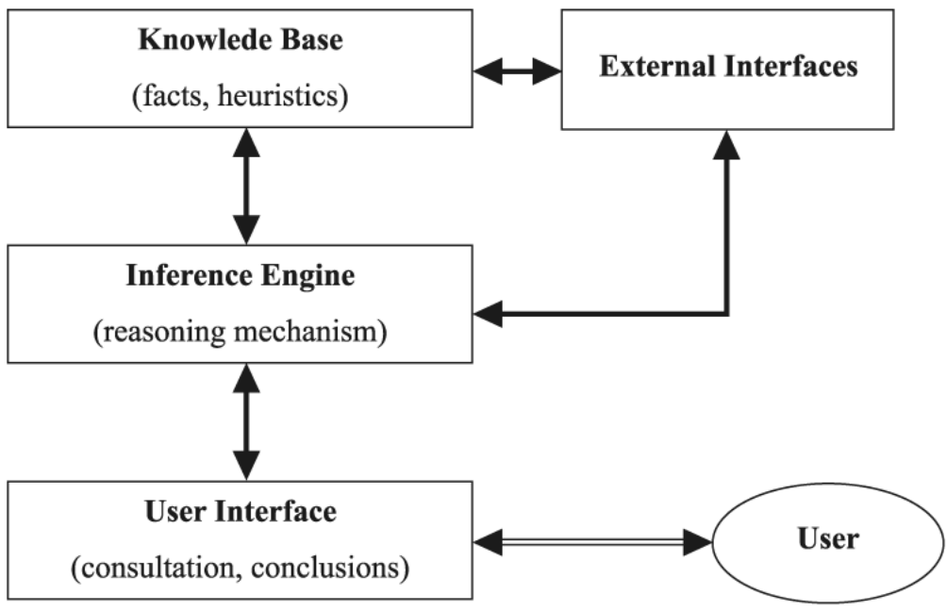


https://www.geeksforgeeks.org/expert-systems/

History of AI: main achievements and setbacks

Boom (1980 – 1987)

- Expert systems applications:
  - medical diagnosis and prognosis
  - design, testing, control and repair of complex engineering systems
  - planning and monitoring of software development projects
  - scientific experiment planning
  - forecasting environmental conditions and impacts (e.g., crop damage)
  - planning and scheduling of resources and various manufacturing tasks
  - identification of geological or chemical compound structures from measurement data



DENDRAL: a chemical-analysis expert system, developed in 1965 by the AI researcher Edward Feigenbaum and the geneticist Joshua Lederberg (Stanford University in California)

Input: spectrographic data obtained from a complicated compound of chemical elements

Output: a hypothesis on the compound's molecular structure

https://www.geeksforgeeks.org/expert-systems/
https://expertsystem101.weebly.com/dendral.html

History of AI: main achievements and setbacks

Boom (1980 – 1987)

- Expert systems limitations:
  - ❖ domain-specific
  - ❖ knowledge acquisition problem: require knowledge engineers to distil and input the data
  - ❖ no real understanding of concepts and their relationships, no commonsense
  - ❖ simple task can potentially become computationally expensive
  - ❖ hard to scale up, for example, increase the number of rules
  - ❖ hard to modify and maintain consistency of large knowledge bases
  - ❖ do not learn: “could not learn concepts that most children know by the time they are 3 years old” (Marvin Minsky)

https://en.wikipedia.org/wiki/Expert_system
https://www.javatpoint.com/advantages-and-disadvantages-of-expert-system

History of AI: main achievements and setbacks

Boom (1980 – 1987)

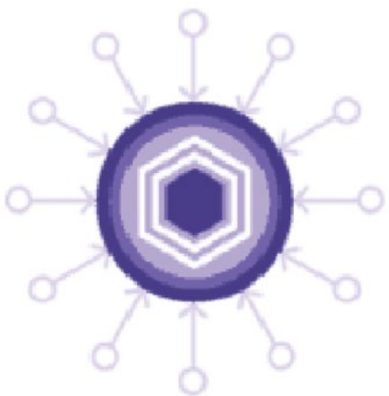
- Ethical concerns around expert systems:
  - ❖ bias and fairness: can reflect the biases of experts and knowledge engineers; e.g., if a medical expert system is trained primarily on data from certain demographics, it may not provide equally accurate diagnoses or treatments for all patients
  - ❖ transparency and accountability: difficult to challenge or correct errors, or check whether decisions are based on relevant factors
  - ❖ privacy concerns: may process sensitive data, such as medical records or financial information
  - ❖ dependence and autonomy: decisions can have profound consequences (e.g., healthcare or law), and over-reliance on automated reasoning could potentially harm patients or clients
  - ❖ job displacement and socio-economic impact: tasks traditionally performed by humans are automated, and large numbers of workers may be displaced without adequate support or retraining opportunities; this may further increase inequalities if certain groups are disproportionately affected by job losses
  - ❖ informed consent to automated decision-making: users must be able to give informed consent regarding the use of automated recommendations or decisions, particularly in sensitive areas such as healthcare or criminal justice

ChatGPT 3.5

History of AI: main achievements and setbacks

Boom (1980 – 1987)

- Cyc project (Doug Lenat, 1984):
  - a massive knowledge base (database) and ontology that would contain all the commonsense facts that the average person knows, down to appropriate level of detail
  - an adequately expressive representation language
  - an inference engine exponentially faster than the engines used in expert systems to be able to infer human-level conclusions, given their knowledge of the world
  - open architecure
  - by 1994, Cyc's ontology grew to about 100,000 terms, and by 2017, it grew to about 1,500,000 terms
  - by 1994, Cyc's knowledge base grew to about 1 million commonsense rules, and by 2017, it grew to about 24.5 million rules

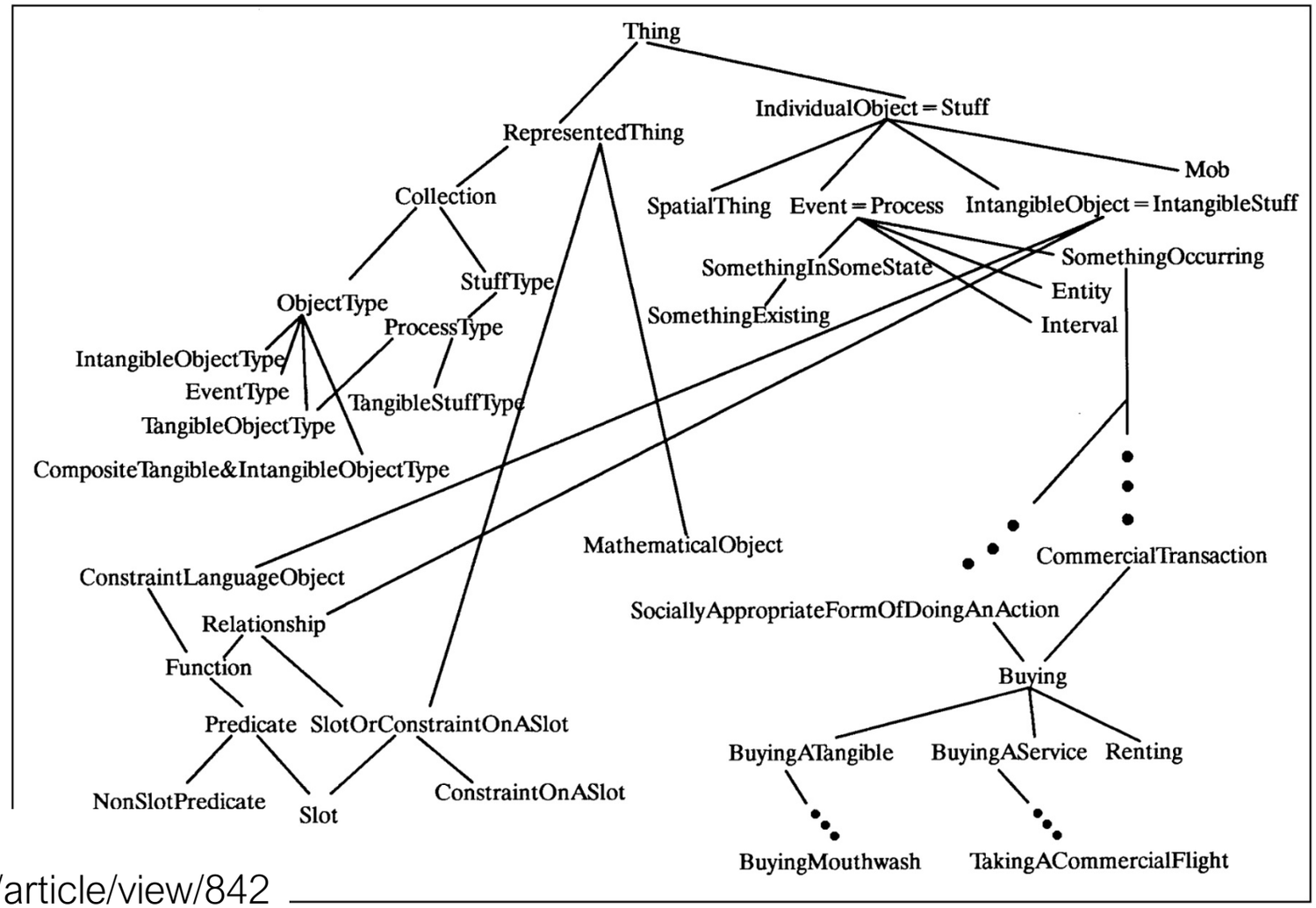


https://en.wikipedia.org/wiki/Cyc

History of AI: main achievements and setbacks

Boom (1980 – 1987)

- Cyc project's criticisms:
  - lack of scalability: the unending amount of data required to produce viable results
  - knowledge acquisition bottleneck: the inability of Cyc to evolve on its own
  - logic-based reasoning without learning
  - formalisation limitations: Cyc relies on a formal language (CycL) to represent knowledge and logical relationships
  - focus on answering over acting
  - lack of practical applications



https://en.wikipedia.org/wiki/Cyc
https://ojs.aaai.org/aimagazine/index.php/aimagazine/article/view/842

History of AI: main achievements and setbacks

Boom (1980 – 1987)

- Fifth Generation Computer Project (Japan, 1980 – 1990s):

  - hardware
    - first generation (mid-1940s): computers using vacuum tubes
    - second generation (mid-1950s): computers using transistors and diodes
    - third generation (mid-1960s): computers using integrated circuits
    - fourth generation (1970s): computers using microprocessors

  - software
    - first generation: machine languages
    - second generation: low-level programming languages (Assembly)
    - third generation: structured high-level programming languages (Fortran, C)
    - fourth generation: non-procedural high-level programming languages (C++)

https://en.wikipedia.org/wiki/Fifth_Generation_Computer_Systems

History of AI: main achievements and setbacks

Boom (1980 – 1987)

- Fifth Generation Computer Project (Japan, 1980 – 1990s):
  - 10-year project to build a new kind of computer based on parallel processing and AI
  - build machines that could reason like humans, engage in dialogue, translate languages, interpret images, diagnose diseases, serve financial advisers, and so on
  - use logic programming languages, such as Prolog (Programming in Logic) designed for symbolic reasoning and problem-solving
    - program = set of axioms
    - computation = proof of a statement from axioms
  - create extensive knowledge bases with domain-specific information and encoded rules that computers could use autonomously
  - develop intuitive user interfaces and natural language interfaces

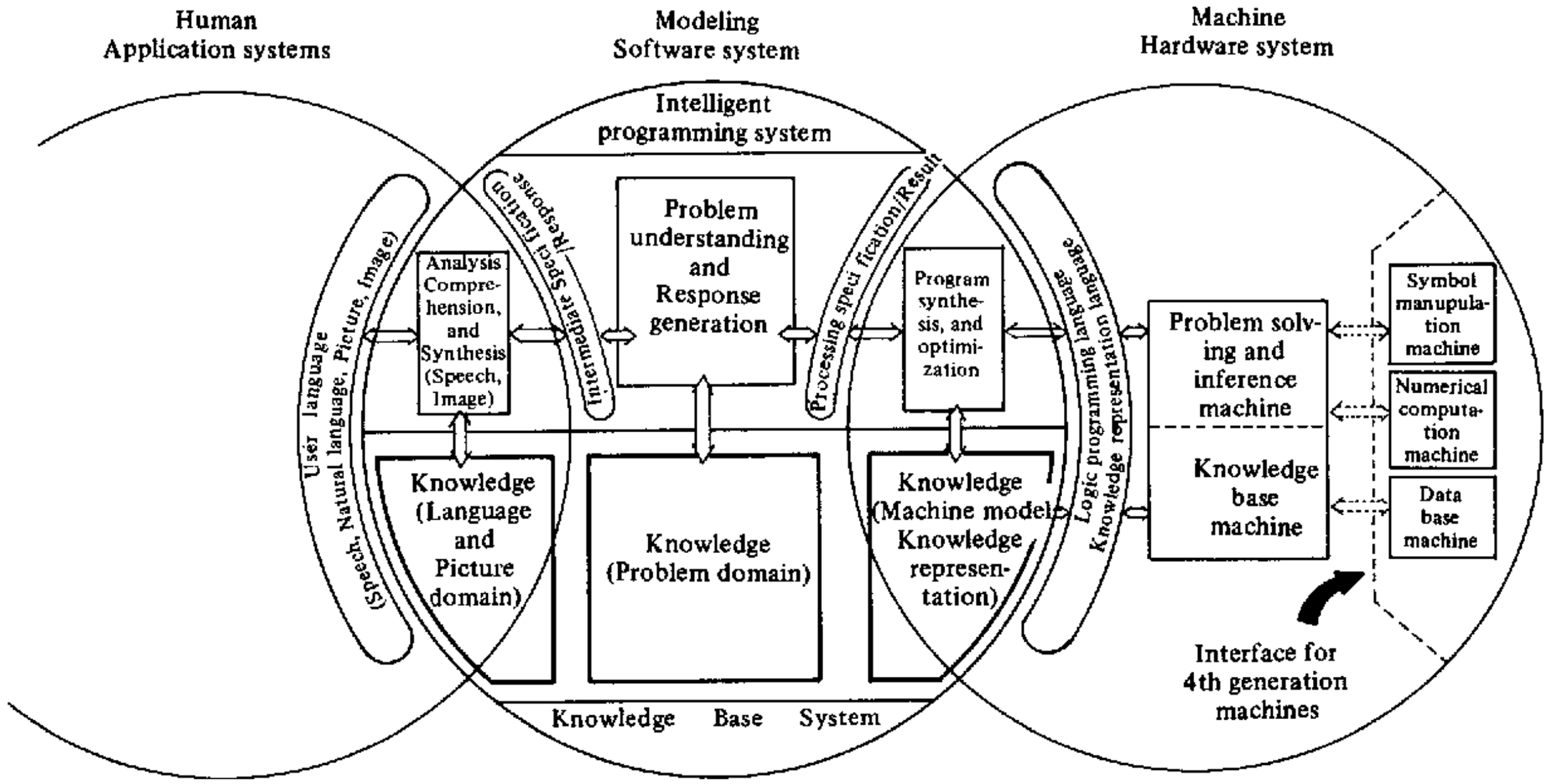
- ❖ faced technical challenges, budgetary constraints, and shifting priorities, leading to its eventual discontinuation in the early 1990s

https://en.wikipedia.org/wiki/Fifth_Generation_Computer_Systems

History of AI: main achievements and setbacks

Boom (1980 – 1987)

- Fifth Generation Computer Project (Japan, 1980 – 1990s):



https://www.semanticscholar.org/paper/Overview-to-the-Fifth-Generation-Computer-System-Moto-Oka/27a1033e6bdeb71982198fed20970b819b6e290d

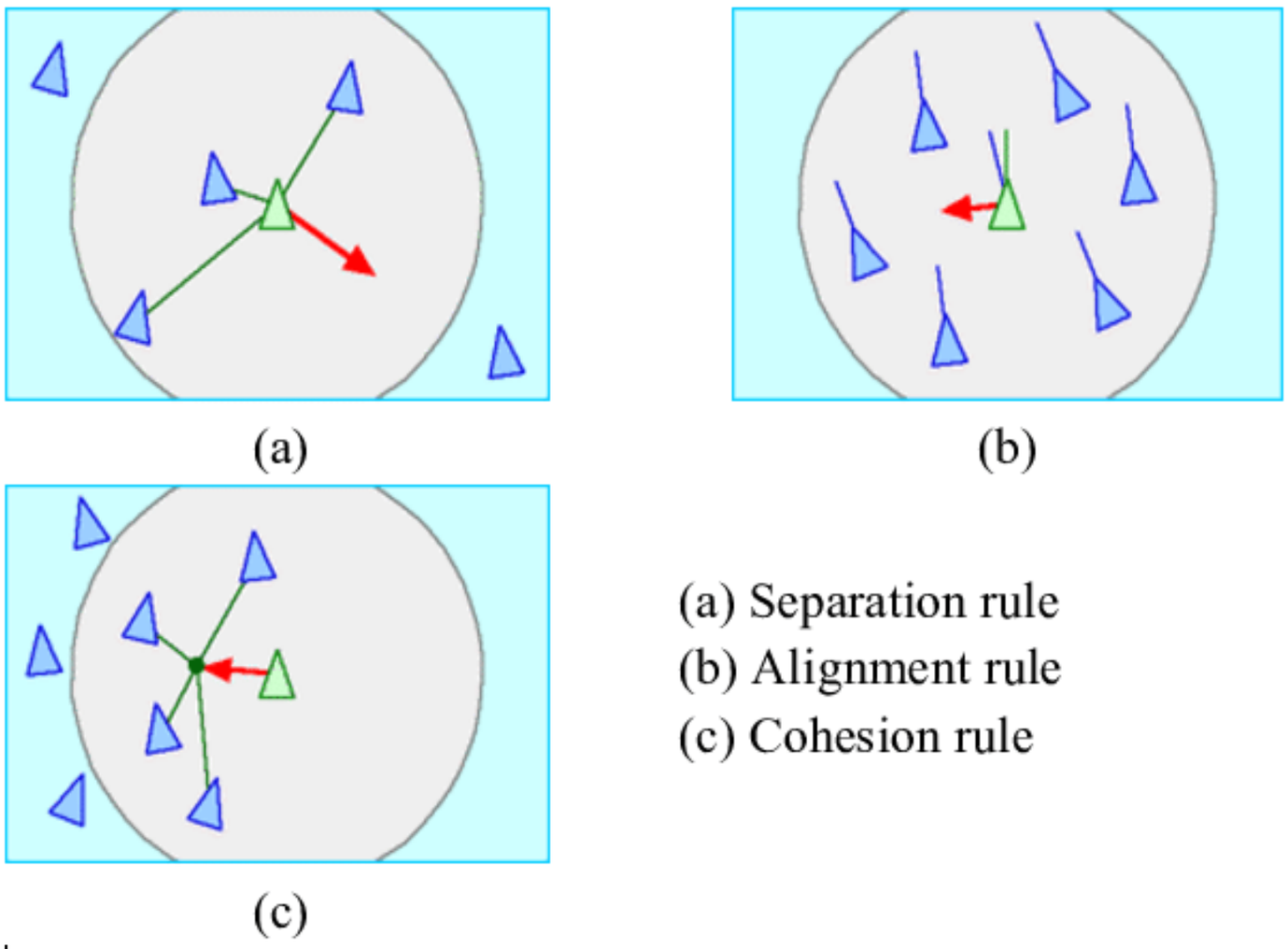
Questions

?

History of AI: main achievements and setbacks

Boom (1980 – 1987)

- Swarm Intelligence:
  - collective behaviour of decentralised, self-organized systems
  - boids: an artificial life program simulating the flocking behaviour of birds (Craig Reynolds, 1986)



https://en.wikipedia.org/wiki/Boids
https://www.scs-europe.net/dlib/2014/ecms14papers/simo_ECMS2014_0062.pdf

History of AI: main achievements and setbacks

https://en.wikipedia.org/wiki/Boids
https://www.youtube.com/watch?v=QbUPfMXXQIY

Discussion

?

Perception and Actuation

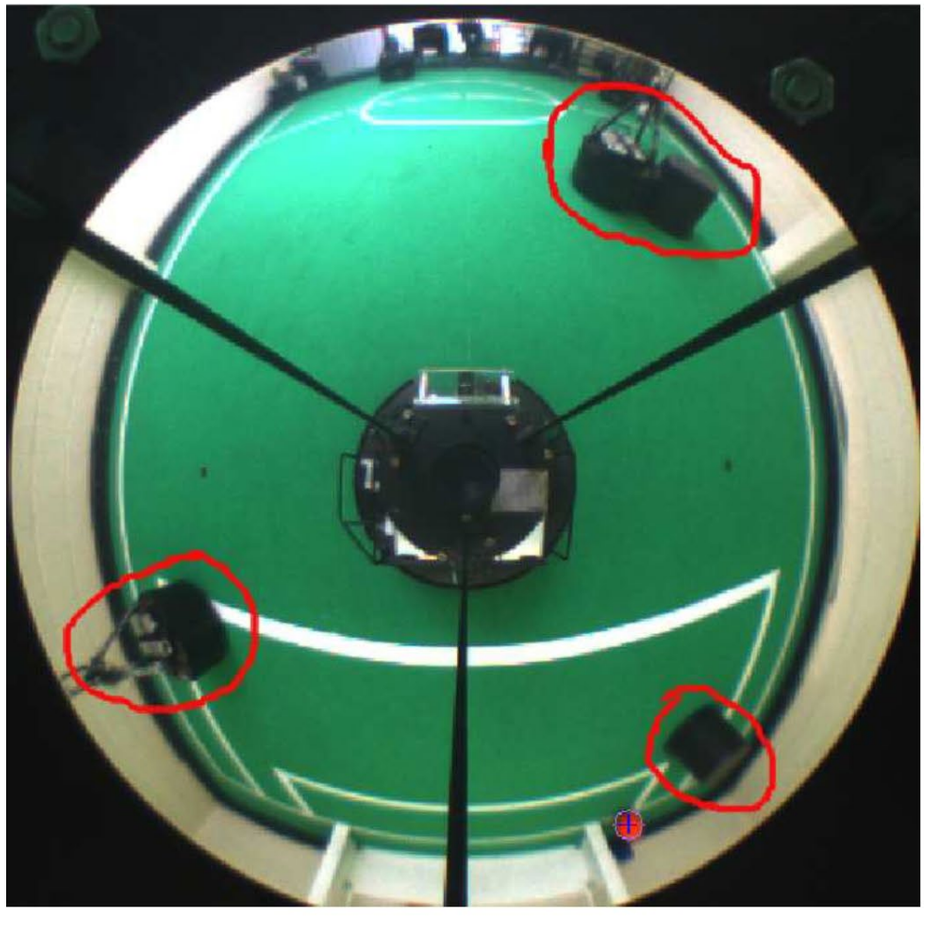
RoboCup-2019 (Sydney, Australia): Middle Size League

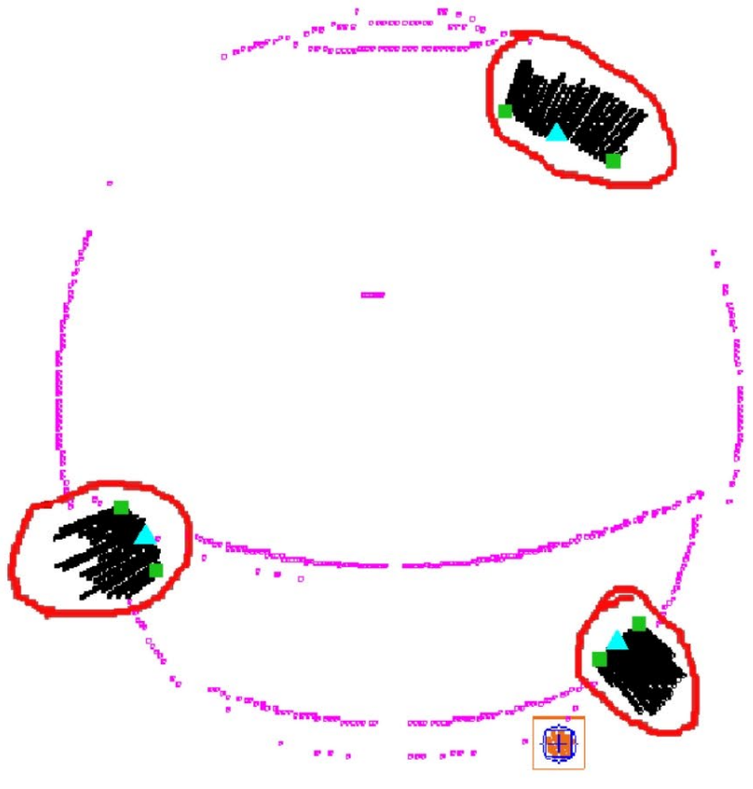
https://www.youtube.com/watch?v=_Y5_iGxWFrQ

Perception and Actuation

Computer vision:

- image processing: filtering, edge detection, and image segmentation
- object recognition: identifying and classifying objects within images
- face recognition: detecting and identifying human faces using features like facial landmarks
- scene understanding: combining object recognition and spatial context to understand the overall scene, for example, semantic segmentation





https://link.springer.com/chapter/10.1007/978-3-642-04686-5_29

Perception and Actuation

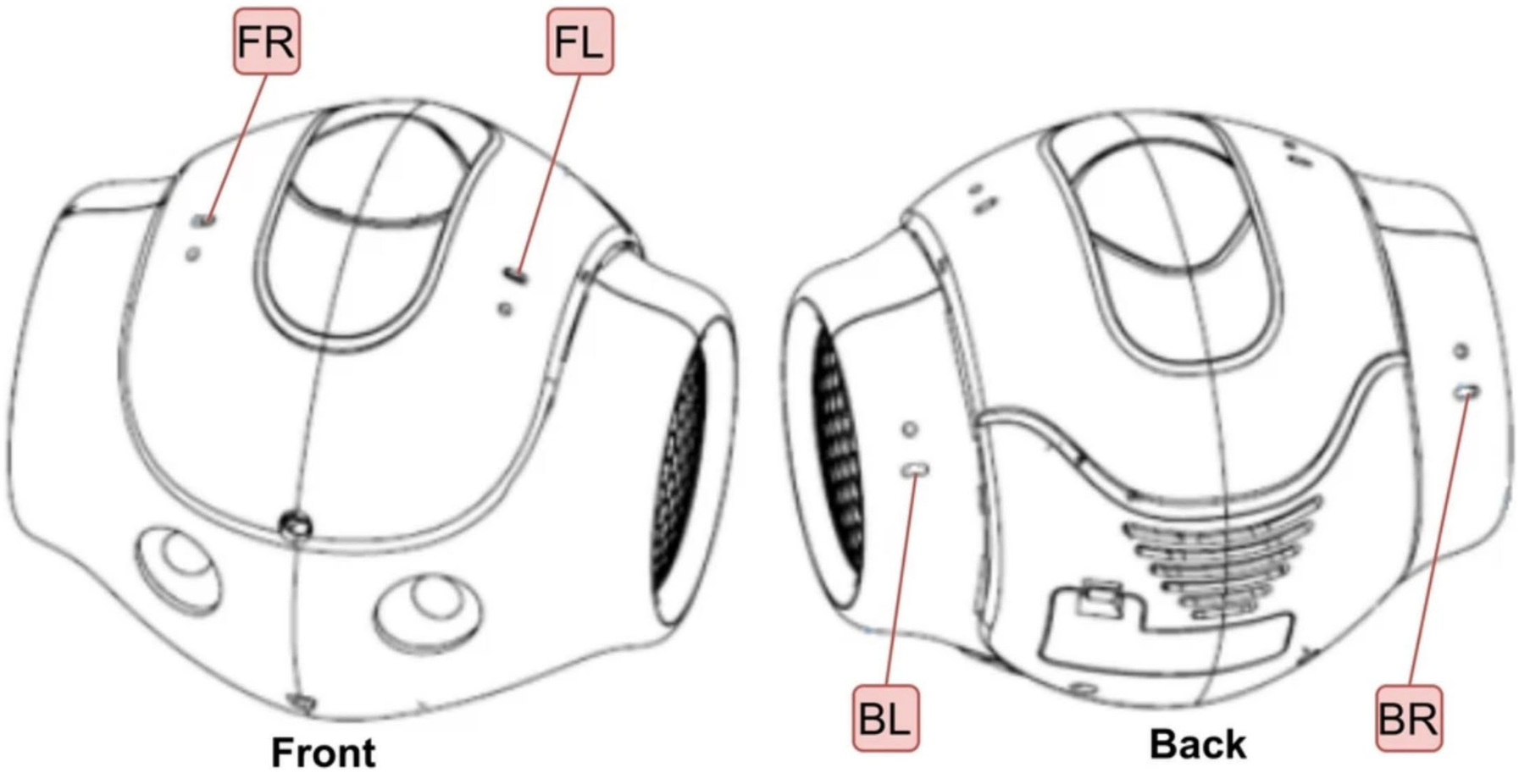
RoboCup Middle Size League: Detection of goal, robots and ball

https://www.youtube.com/watch?v=NuluD3HR_Iw

Perception and Actuation

Auditory processing:

- sound localization: determining the direction of a sound source using microphone arrays
- speech and speaker recognition: identifying spoken words and the speaker's identity
- environmental sound recognition: detecting and classifying sounds from the environment, such as sirens, alarms, or natural sounds

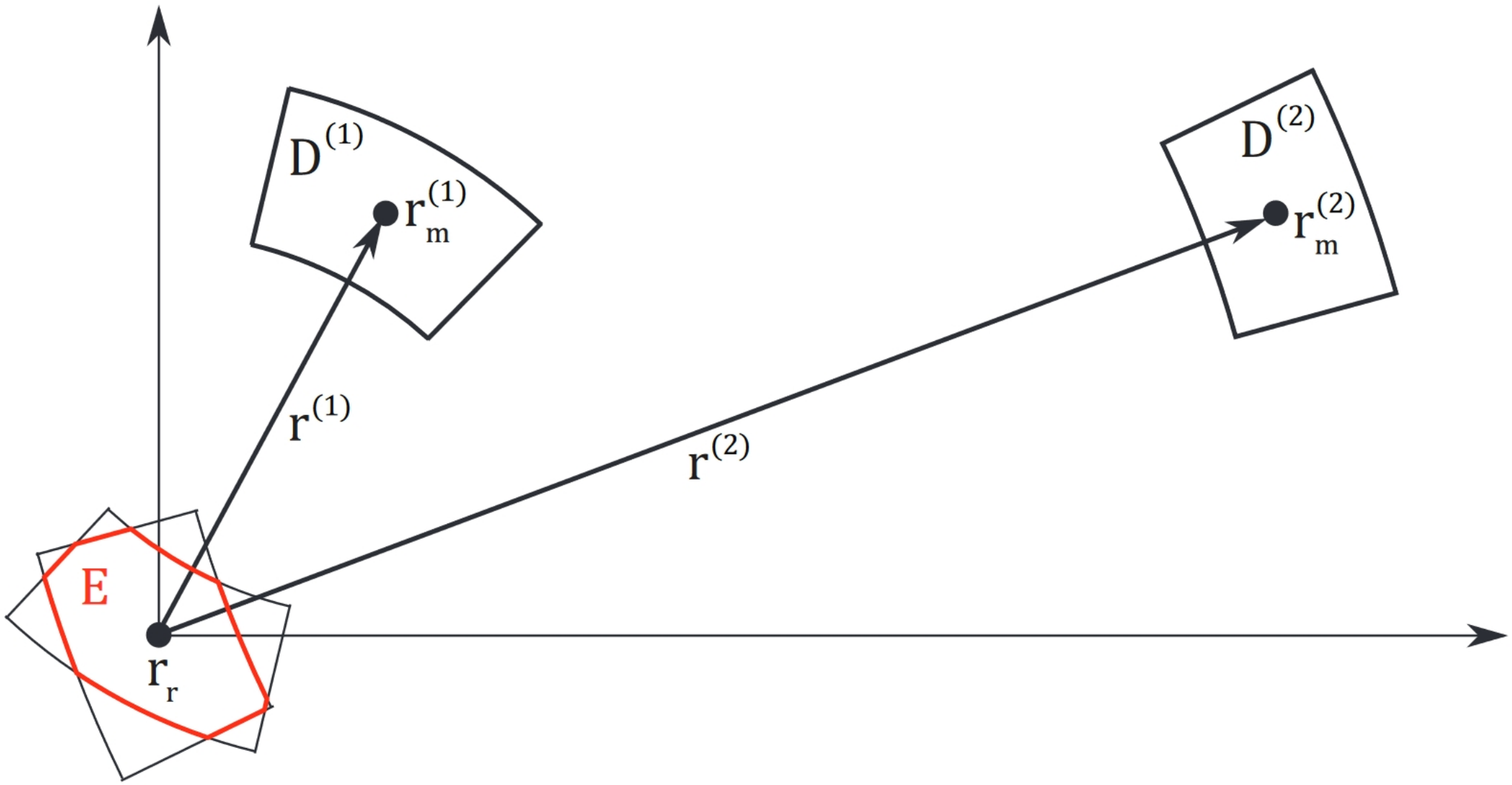


https://link.springer.com/article/10.1007/s00521-022-08047-x

Perception and Actuation

Sensor fusion:

- combining data from multiple sensors: cameras, LIDAR, RADAR, and inertial measurement units. to create a comprehensive perception of the environment
- specialised algorithms such as Kalman filters and particle filters to integrate and interpret data from various sources

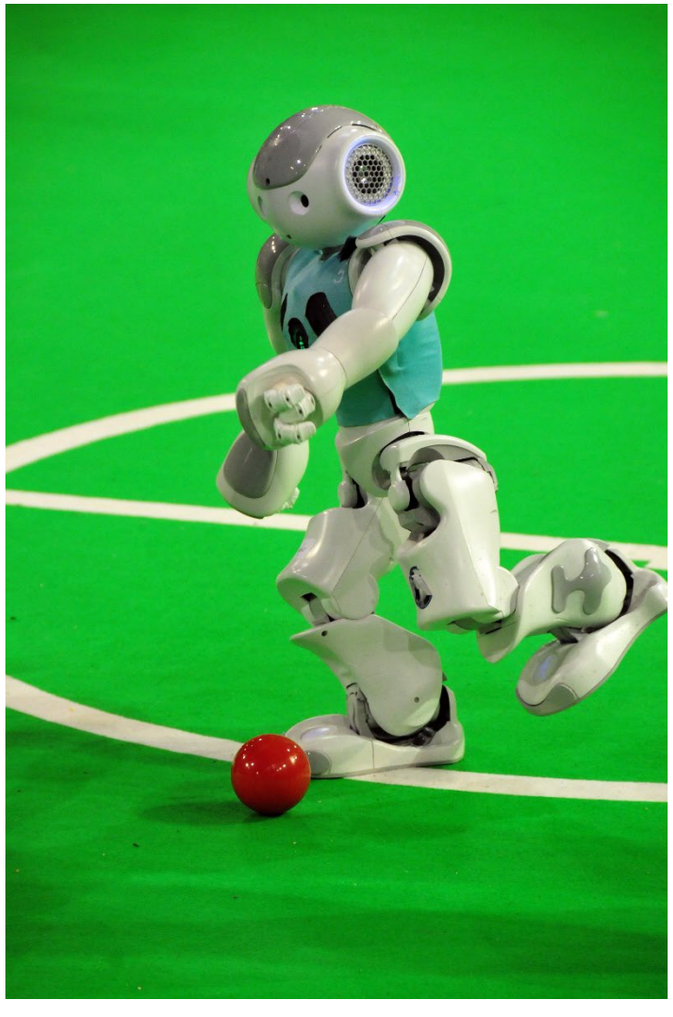


http://www.prokopenko.net/uploads/2/1/7/6/21762362/rc2013_budden_prokopenko.pdf

Perception and Actuation

Robotics and autonomous systems:

- motion planning: algorithms to determine a path or trajectory for a robot to follow
- control systems: methods to accurately execute planned motions within dynamic environments, manage acceleration, braking, and steering
- manipulation: techniques for controlling robotic arms and grippers
- feedback loops: use sensor readings to dynamically adjust actions (for example: robot may adjust its grip based on tactile sensor feedback)



https://en.wikipedia.org/wiki/RoboCup

Break

?

Perception and Actuation

Ethical concerns:

- privacy and surveillance
  - cameras and microphones in public and private spaces can track and record people's activities without their consent
  - facial recognition systems can be used to identify and monitor individuals without their knowledge
- bias and fairness
  - facial recognition systems that perform poorly on certain demographic groups, such as people with darker skin tones, leading to misidentification and potential harm
  - voice recognition systems that do not accurately understand accents or dialects, disadvantaging non-native speakers
- accountability and transparency
  - autonomous vehicles making decisions in complex traffic scenarios, where understanding the decision-making process is crucial for addressing accidents or malfunctions
- security and safety
  - an autonomous vehicle misinterpreting sensor data and causing an accident
  - industrial robots being hacked to perform dangerous actions

ChatGPT 4o

Perception and Actuation

Ethical concerns:

- ethical use of military and law enforcement AI
  - autonomous drones used for surveillance or targeted strikes, raising questions about the decision-making process for use of lethal force
  - predictive policing algorithms may reinforce existing biases and lead to discriminatory practices
- impact on employment
  - autonomous delivery robots and vehicles replacing delivery drivers
  - AI-powered diagnostic tools reducing the need for certain medical professionals
- consent and autonomy
  - smart home devices that constantly monitor household activities, potentially undermining individual autonomy and consent.
  - wearable health monitors that collect and analyse personal health data without explicit consent

ChatGPT 4o

Perception and Actuation

Mitigation of ethical concerns:

- privacy and surveillance
  - implementing strict regulations and transparency around data collection and usage.
  - developing and deploying privacy-preserving AI technologies, such as differential privacy and federated (collaborative) learning
- bias and fairness
  - ensuring diverse and representative training datasets
  - regularly auditing AI systems for bias and implementing corrective measures.
- accountability and transparency
  - developing explainable AI (XAI) techniques to make AI decision-making processes more transparent and understandable
  - establishing clear accountability frameworks for AI systems, ensuring that humans remain responsible for AI actions
- security and safety
  - implementing robust security measures to protect AI systems from cyber threats
  - rigorous testing and validation of AI systems to ensure safety and reliability
  - developing fail-safe mechanisms and emergency protocols

ChatGPT 4o

Perception and Actuation

Mitigation of ethical concerns:

- ethical use of military and law enforcement AI
  - establishing international regulations and ethical guidelines for the use of AI in military and law enforcement
  - ensuring that decisions involving the use of force remain under human control and oversight
- impact on employment
  - investing in retraining and education programs to help workers transition to new roles
  - implementing policies to ensure fair distribution of the economic benefits of AI
- consent and autonomy
  - clear and informed consent mechanisms for data collection and usage.
  - empowering individuals with tools and rights to control their data and how it is used by AI systems

ChatGPT 4o

Discussion

?

Homework

- check the information sources provided on different slides (at least some)

- Boston Dynamics
  YouTube: https://www.youtube.com/@BostonDynamics

- please try to watch some of the videos before the next lecture, and prepare to give a few brief comments:
  - what was most striking about the AI technology in the videos?
  - what societal impact did the AI make?
  - what were the main AI limitations, biases and risks?

?