COMP9208
Artificial Intelligence and Society

Lecture 3: 22 August 2025

Prof. Mikhail Prokopenko
Dr. Sheryl Chang

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

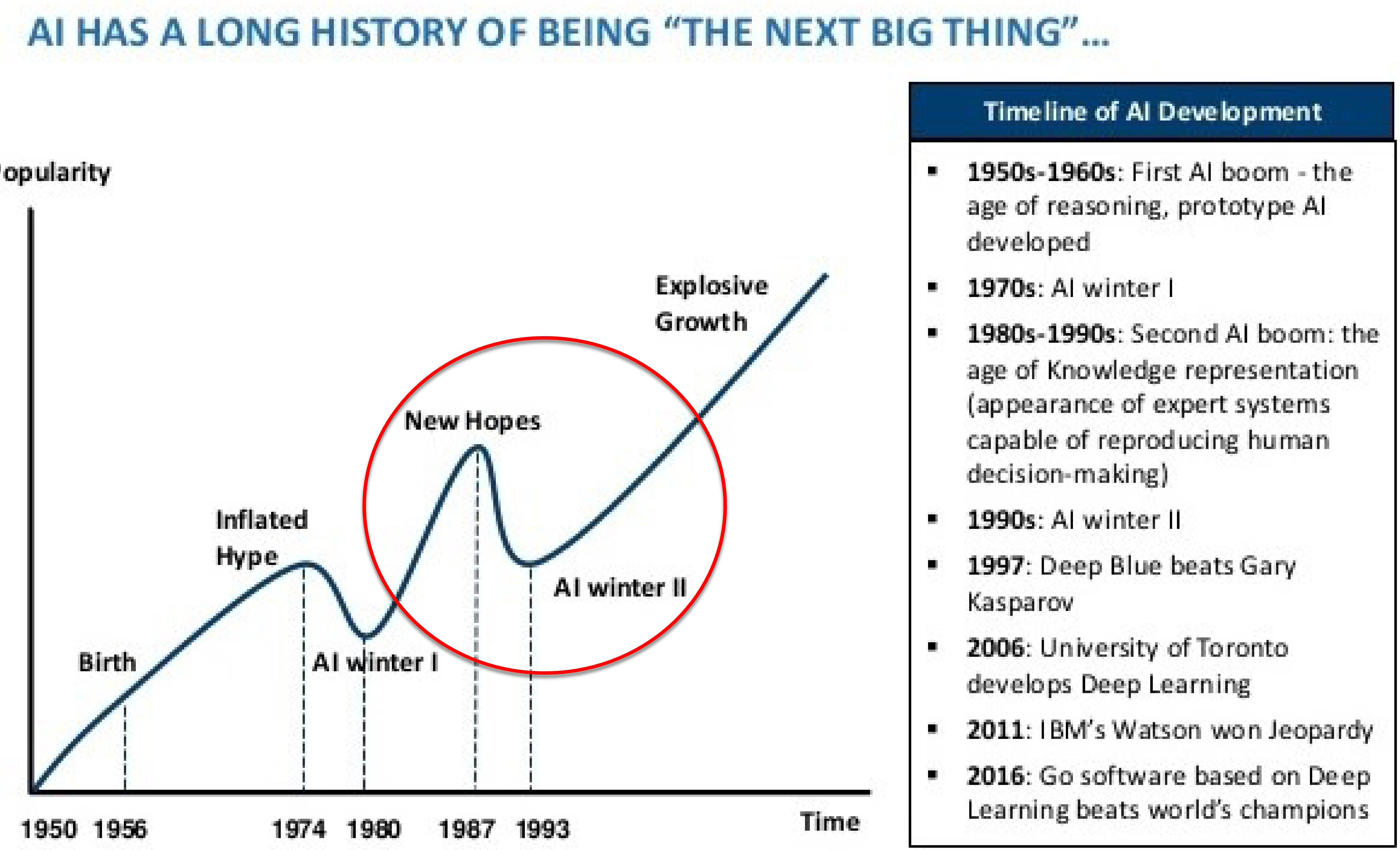
Email: mikhail.prokopenko@sydney.edu.au
Office: J12 (Computer Science), level 3W, room 324

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

?

History of AI: main achievements and setbacks

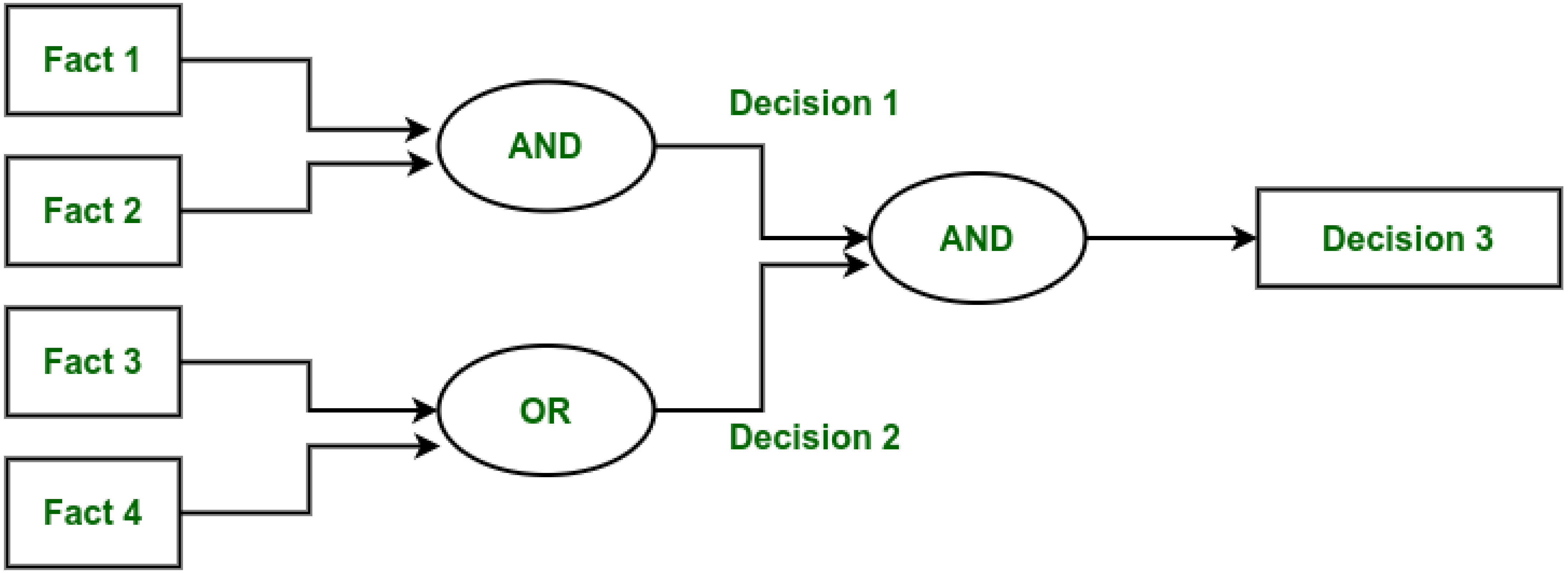


https://www.actuaries.digital/2018/09/05/history-of-ai-winters/

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)

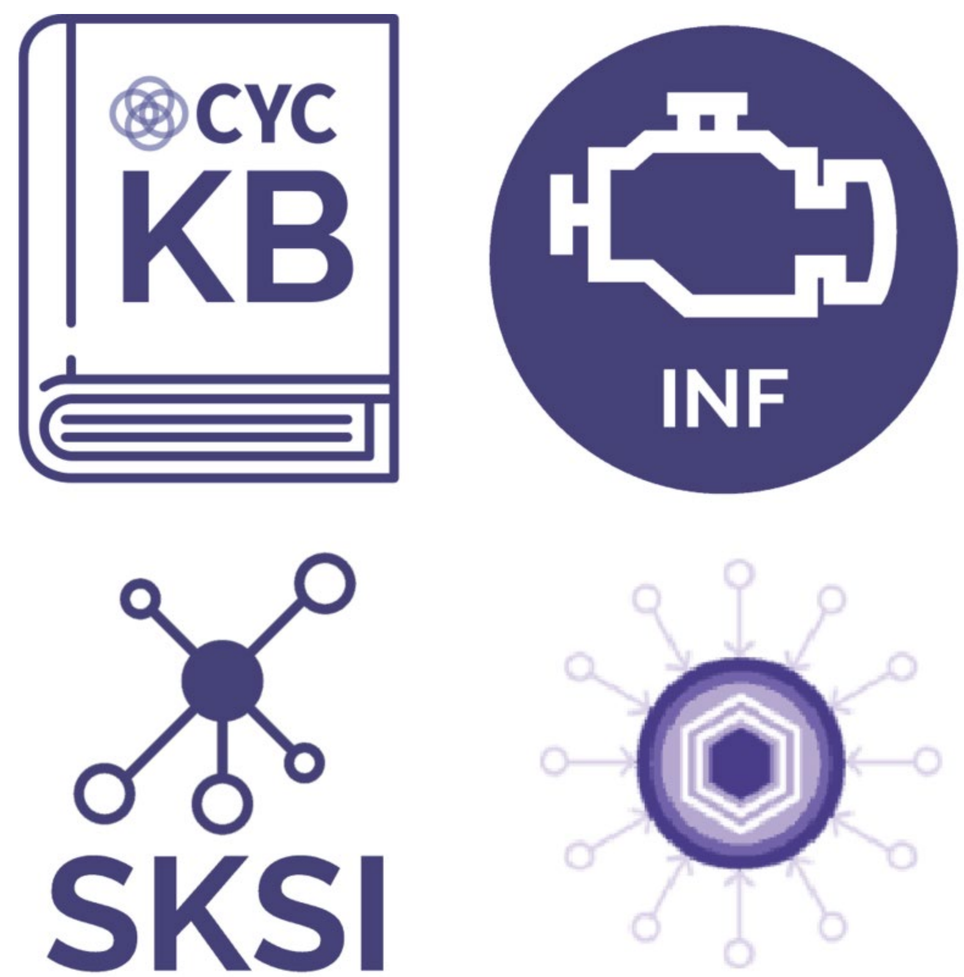
- Expert systems limitations:
    - ❖ domain-specific
    - ❖ knowledge acquisition problem: require knowledge engineers to distil and input the data
    - ❖ no real understanding of concepts and their relationships, no commonsense
    - ❖ simple task can potentially become computationally expensive
    - ❖ 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)

- 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

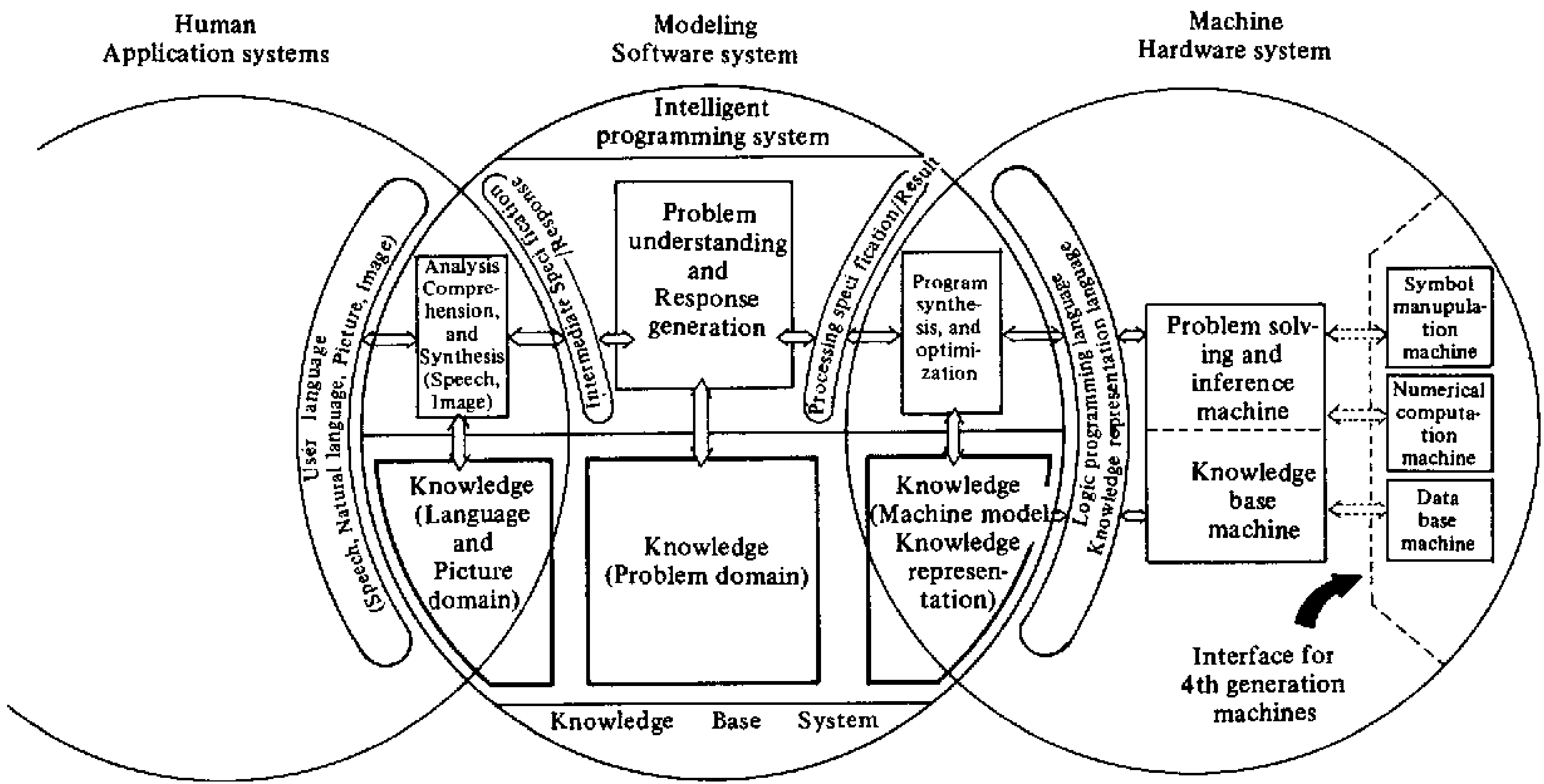


https://en.wikipedia.org/wiki/Cyc

History of AI: main achievements and setbacks

Boom (1980 – 1987)

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

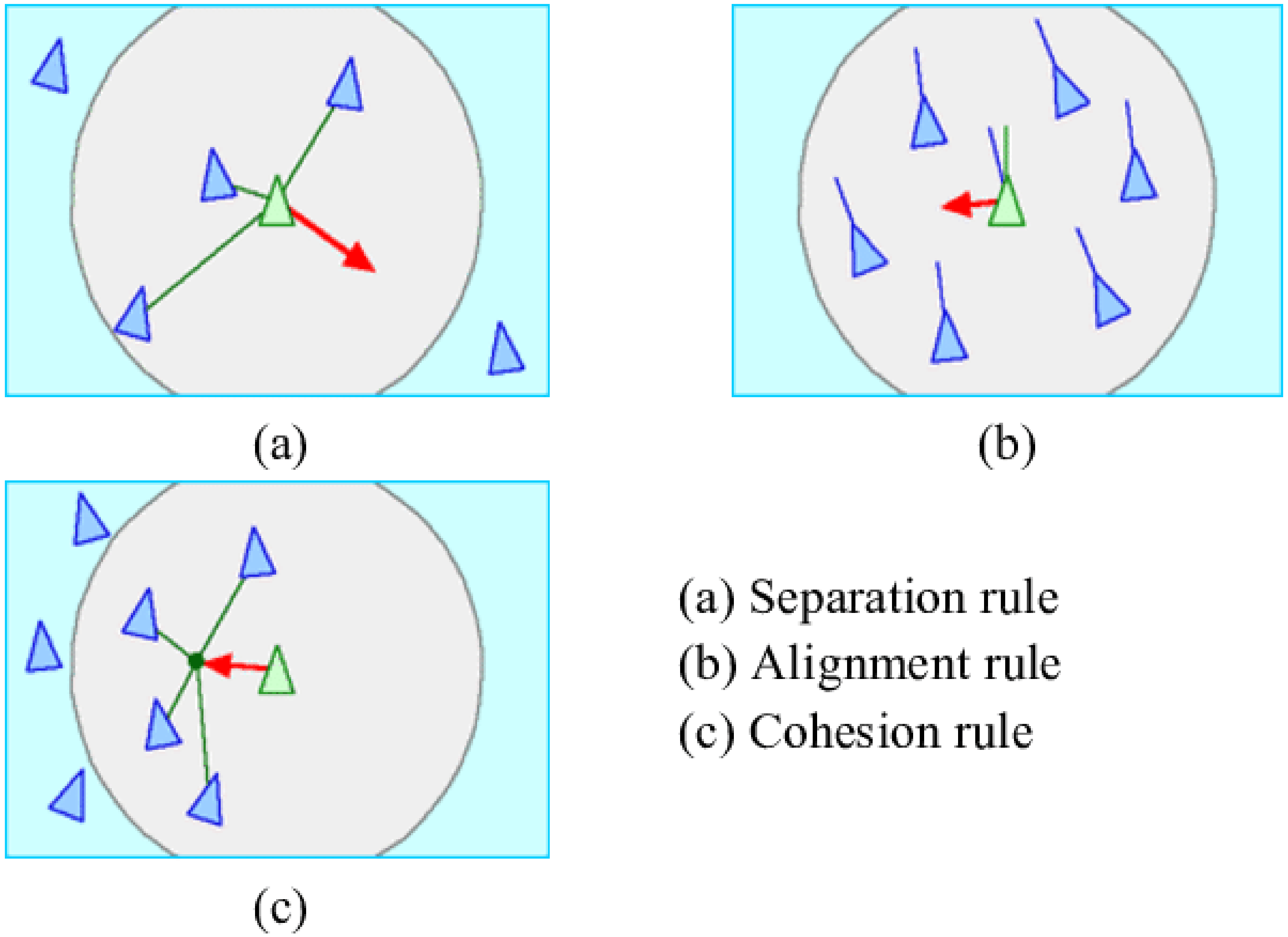


https://www.semanticscholar.org/paper/Overview-to-the-Fifth-Generation-Computer-System-Moto-Oka/27a1033e6bdeb71982198fed20970b819b6e290d

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)

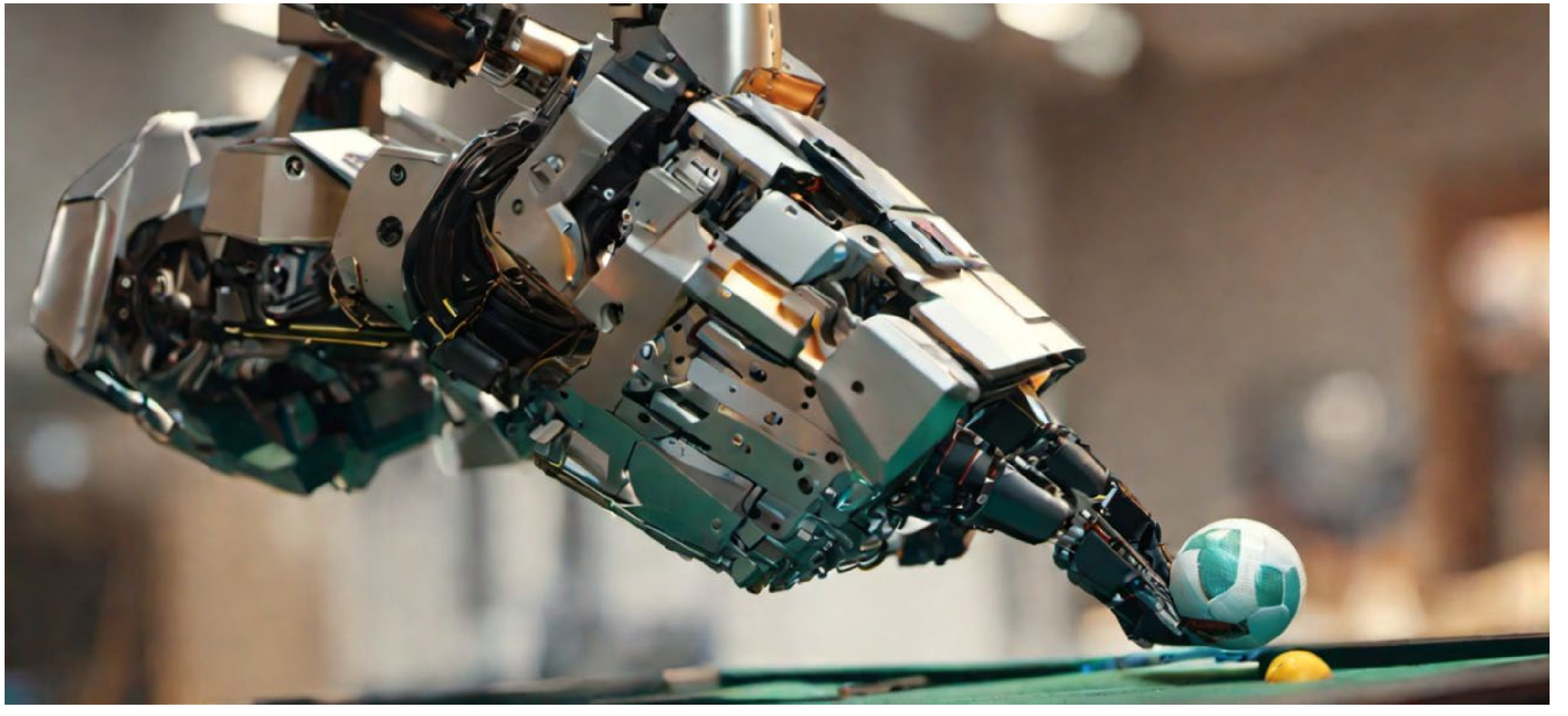


(a) Separation rule
(b) Alignment rule
(c) Cohesion rule

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

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



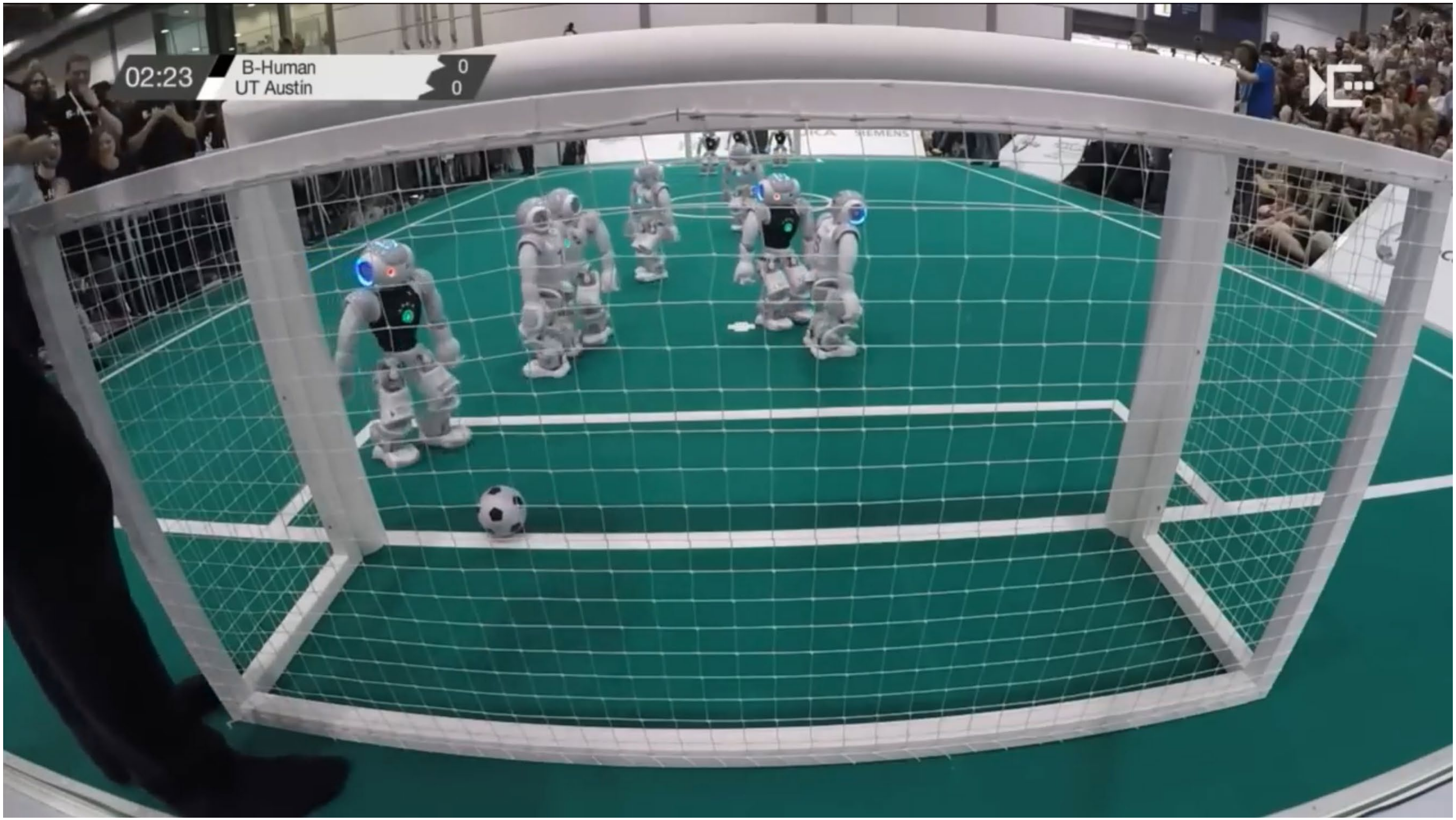
the robot needs to

- 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

Standard Platform League: image segmentation

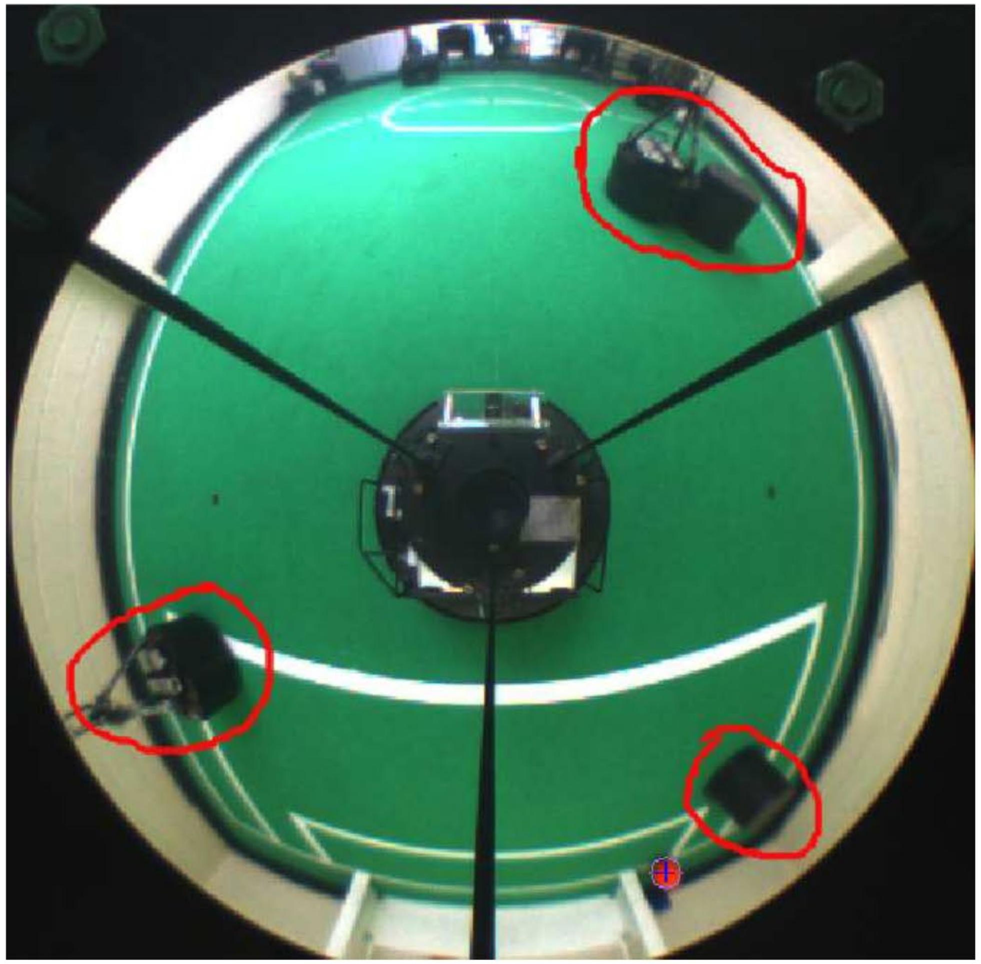


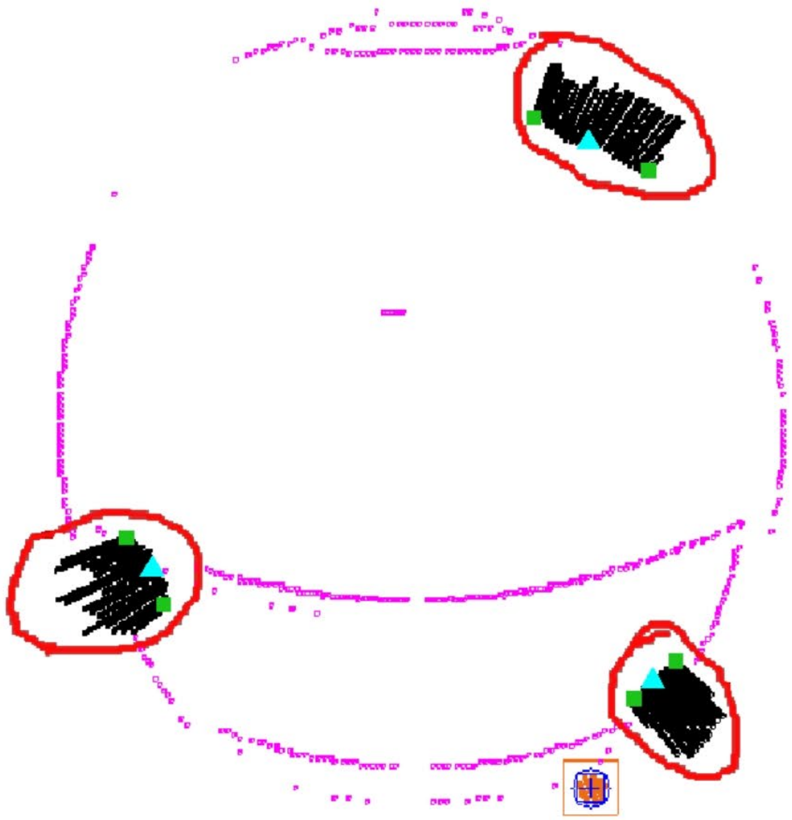
https://www.youtube.com/watch?v=XgRw42oHN-Y

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



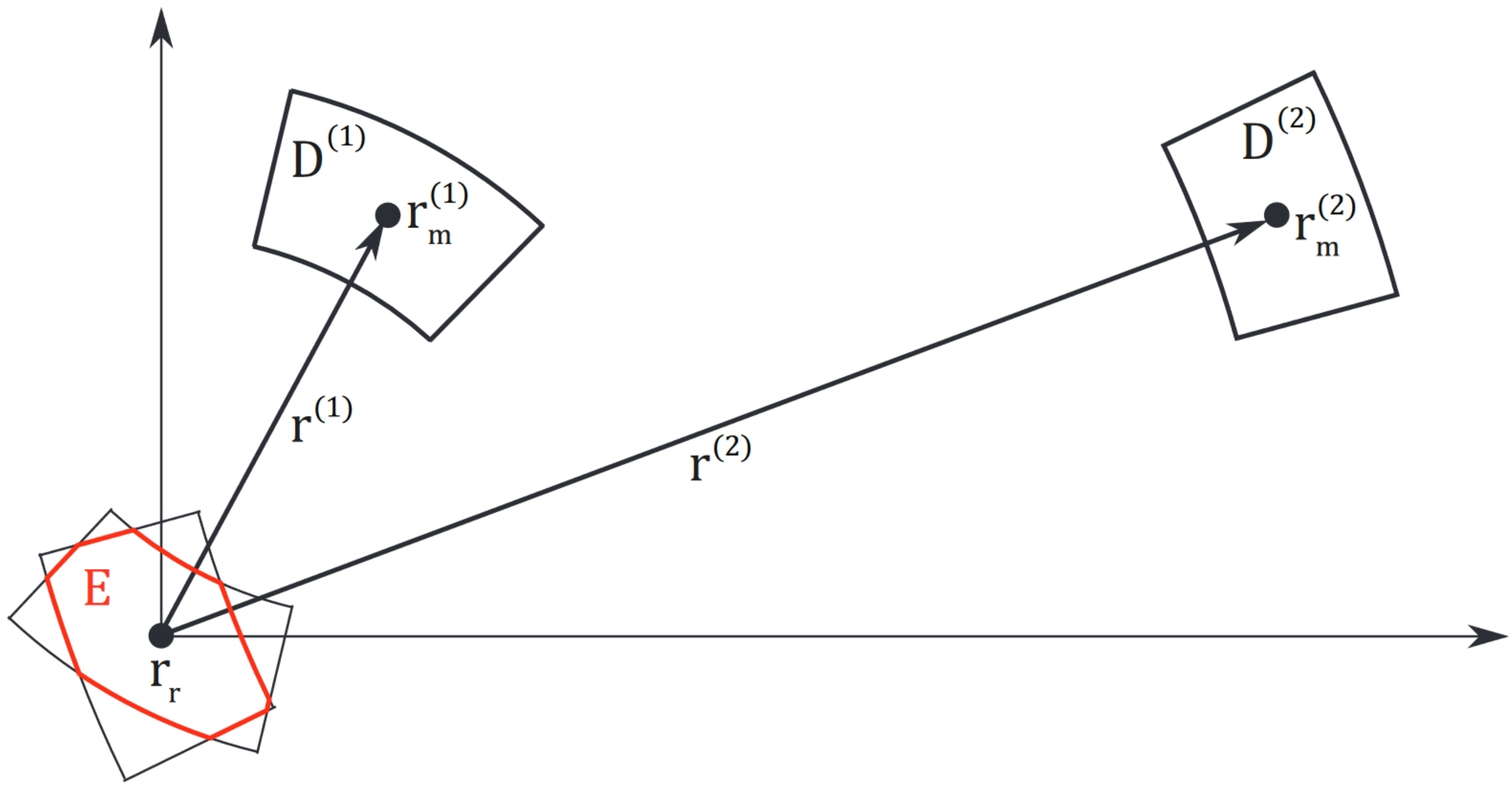


https://link.springer.com/chapter/10.1007/978-3-642-04686-5_29

Perception and Actuation

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

Questions

?

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?

?

Perception and Actuation

Integration of perception and actuation

https://www.youtube.com/watch?v=fn3KWM1kuAw

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
  - ○
  - ○

?

- impact on employment
  - ○
  - ○

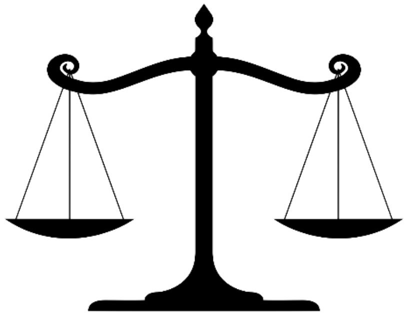
?

- consent and autonomy
  - ○
  - ○

?

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
    -
    -

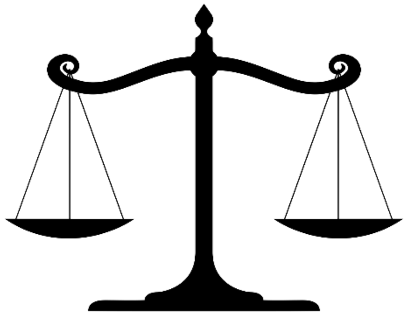
?

- consent and autonomy
    -
    -

?

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
  -
  -

?

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

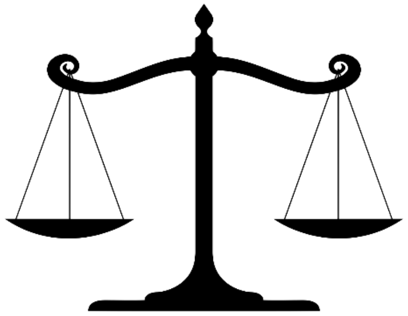
Perception and Actuation

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

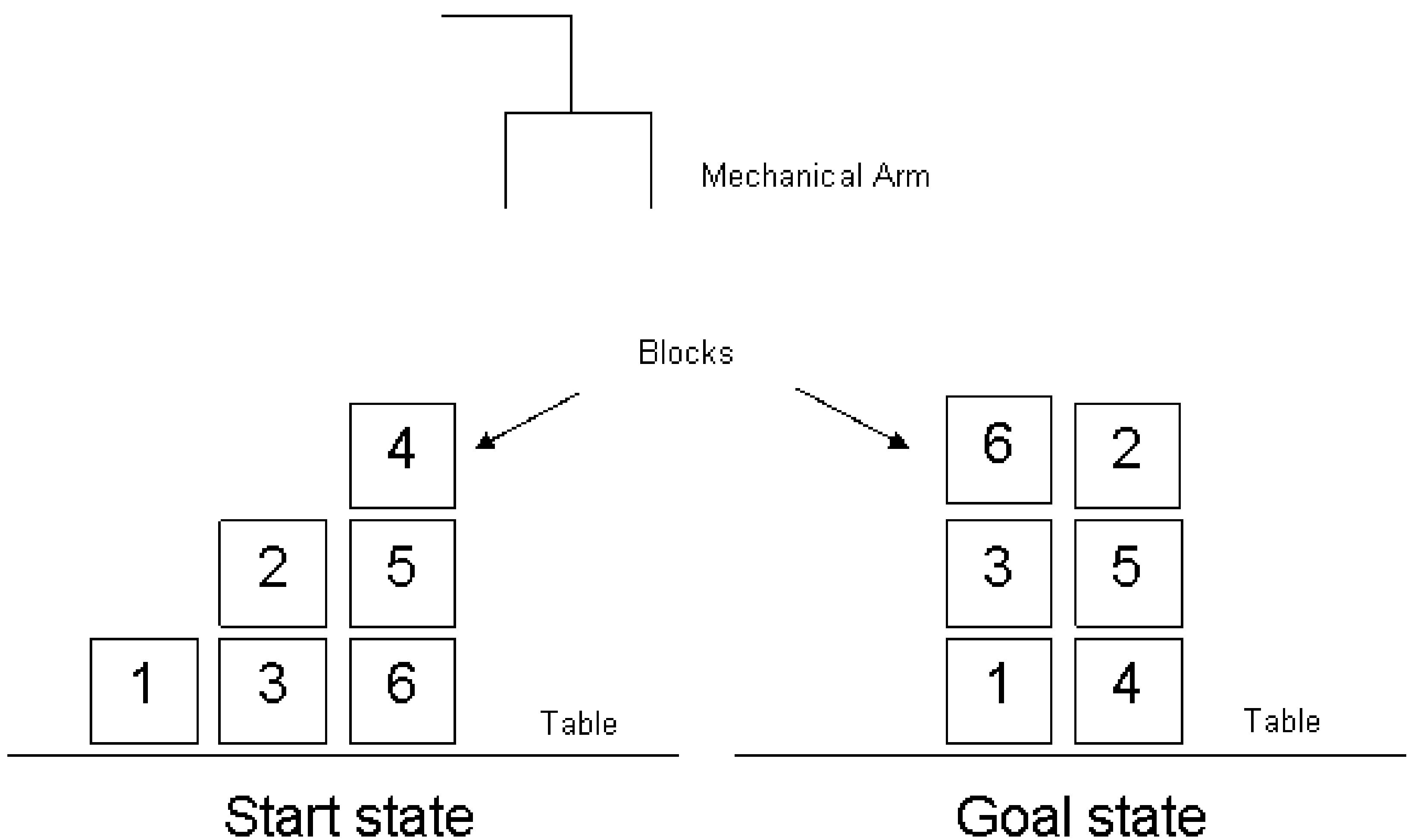
Discussion

?

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

Knowledge Representation and Reasoning (KR&R)

Logical Representation

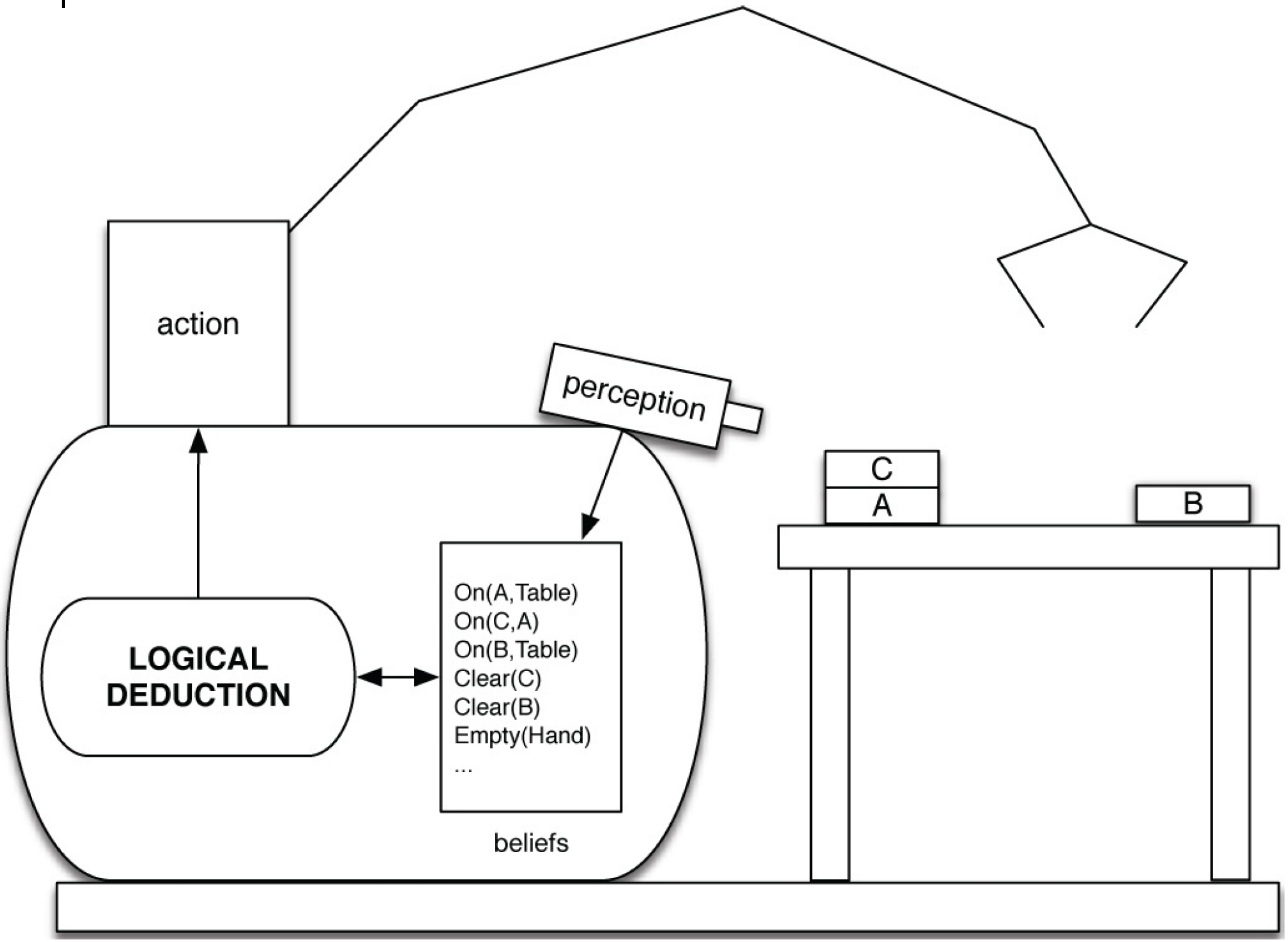


Figure 5: McCarthy's vision for logic-based AI. A logic-based AI system explicitly represents its beliefs about the environment via logical statements, and intelligent decision-making is broken down into a process of logical reasoning.

M. Wooldridge, A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going, Flatiron Books, 2021.

Knowledge Representation and Reasoning (KR&R)

Logical Representation

- Propositional logic:
  - uses propositions about the world, which can be either true or false
  - example: $A \wedge B \rightarrow C$ (if A and B are true, then C is true)

Propositional symbols (atomic formulas): A, B, C

Logical connectives: $\neg, \wedge, \vee, \rightarrow, \leftrightarrow$

Build up formulas recursively—if f and g are formulas, so are the following:

- Negation: $\neg f$
- Conjunction: $f \wedge g$
- Disjunction: $f \vee g$
- Implication: $f \rightarrow g$
- Biconditional: $f \leftrightarrow g$

https://medium.com/towards-nesy/logic-study-note-propositional-fol-22f353b0262f

Knowledge Representation and Reasoning (KR&R)

Logical Representation

- First-order logic (FOL):
  - extends propositional logic by including variables, relations, and quantifiers
  - example: ∀x (Human(x) → Mortal(x))   (All humans are mortal)

English	**First-Order**
At least one x is P	$\exists x P(x)$
All x are P	$\forall x P(x)$
Some x are P	$\exists x P(x)$
Not all x are P	$\exists x \neg P(x)$
No x are P	$\forall x \neg P(x)$

https://www.baeldung.com/cs/first-order-logic

Knowledge Representation and Reasoning (KR&R)

Logical Representation

- First-order logic (FOL):
  - extends propositional logic by including variables, relations, and quantifiers
  - example: $\forall x\ (\ \text{Human}(x) \rightarrow \text{Mortal}(x)\ )$ (All humans are mortal)

All humans are mortal.

Emma is human.

Therefore, Emma is mortal.

Sound reasoning

https://www.baeldung.com/cs/first-order-logic
M. Wooldridge, A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going, Flatiron Books, 2021.

Knowledge Representation and Reasoning (KR&R)

Logical Representation

- First-order logic (FOL):
  - extends propositional logic by including variables, relations, and quantifiers
  - example: $\forall x$ (Human(x) $\rightarrow$ Mortal(x)) (All humans are mortal)

All humans are mortal.

Emma is human.

Therefore, Emma is mortal.

Sound reasoning

Unsound reasoning

All students are hardworking.

Sophie is a student.

Therefore, Sophie is rich.

https://www.baeldung.com/cs/first-order-logic
M. Wooldridge, A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going, Flatiron Books, 2021.

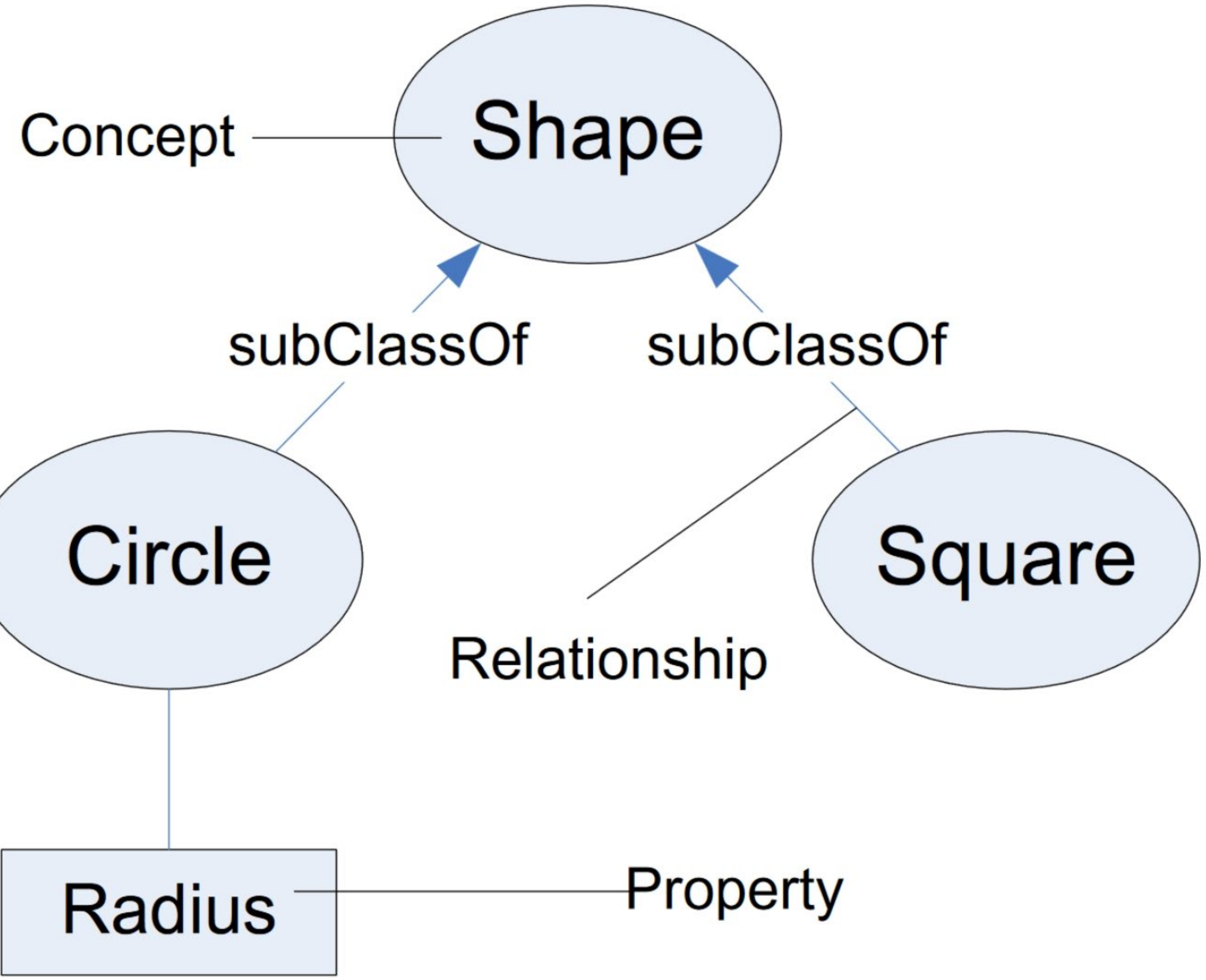
Questions

?

Knowledge Representation and Reasoning (KR&R)

Logical Representation

- Description logic:
  - formal language for representing structured knowledge and reasoning about it
  - example: ontologies such as Web Ontology Language (OWL)

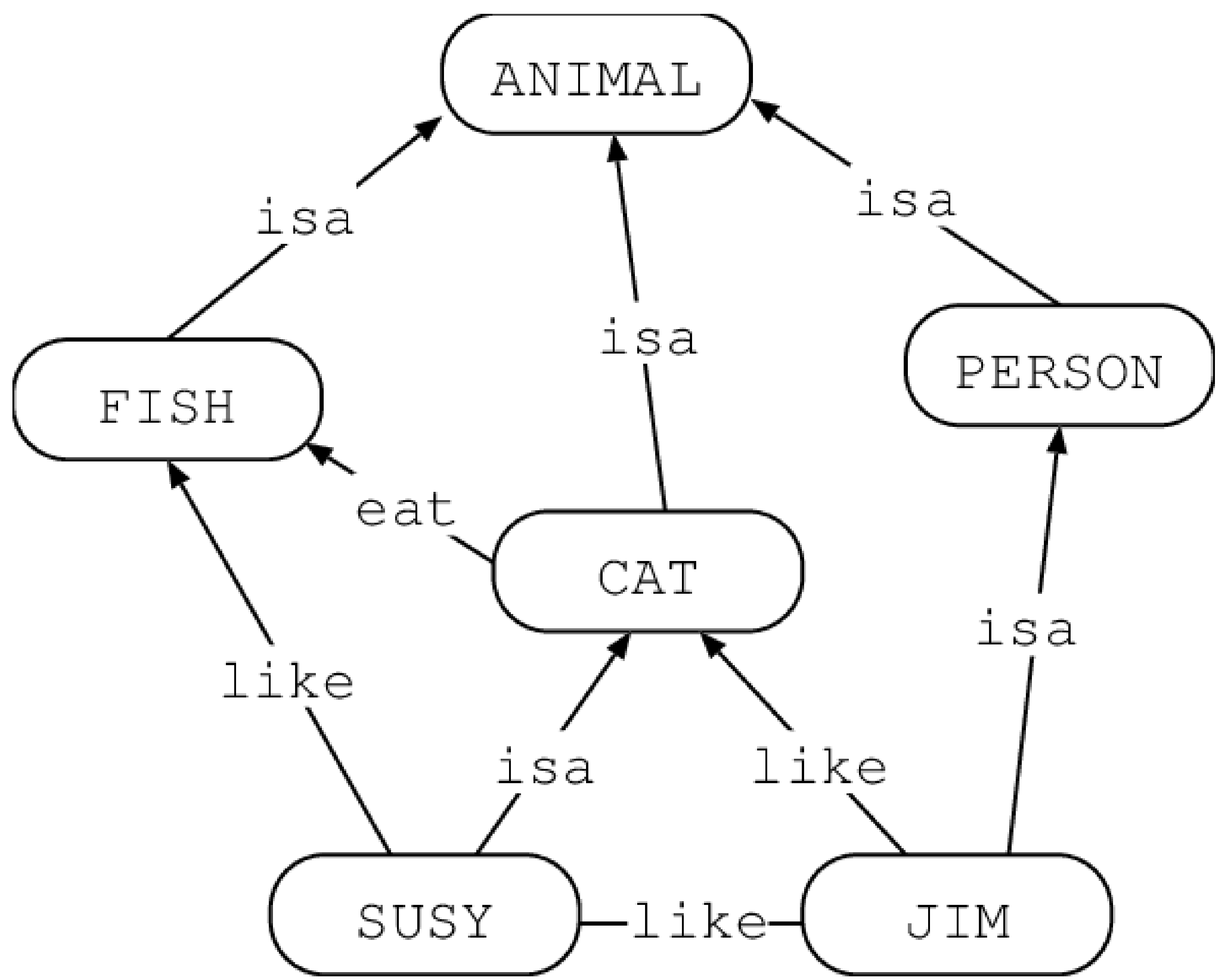


https://www.researchgate.net/publication/4186813_A_model-driven_approach_for_specifying_semantic_Web_services

Knowledge Representation and Reasoning (KR&R)

Semantic networks

- graphs consisting of nodes (concepts) and edges (relationships)
- useful for visualizing relationships and hierarchies
- example: concept maps and word nets

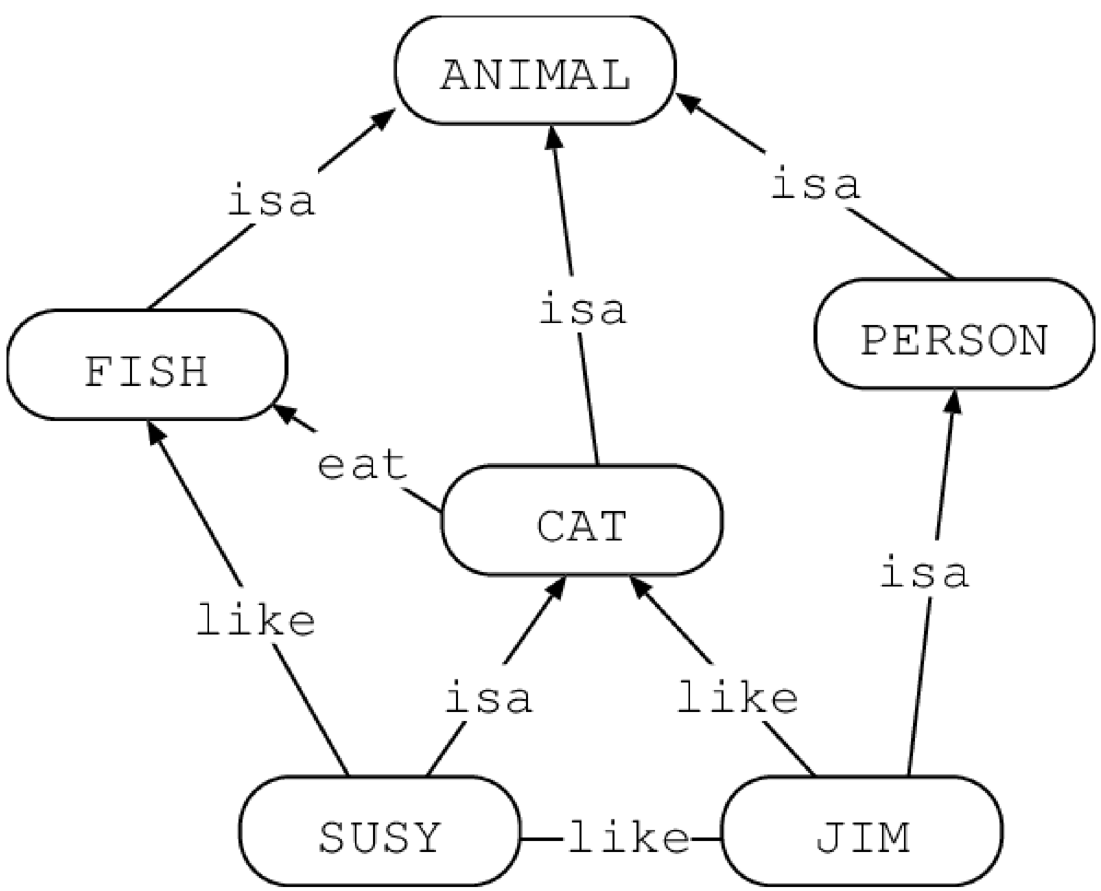


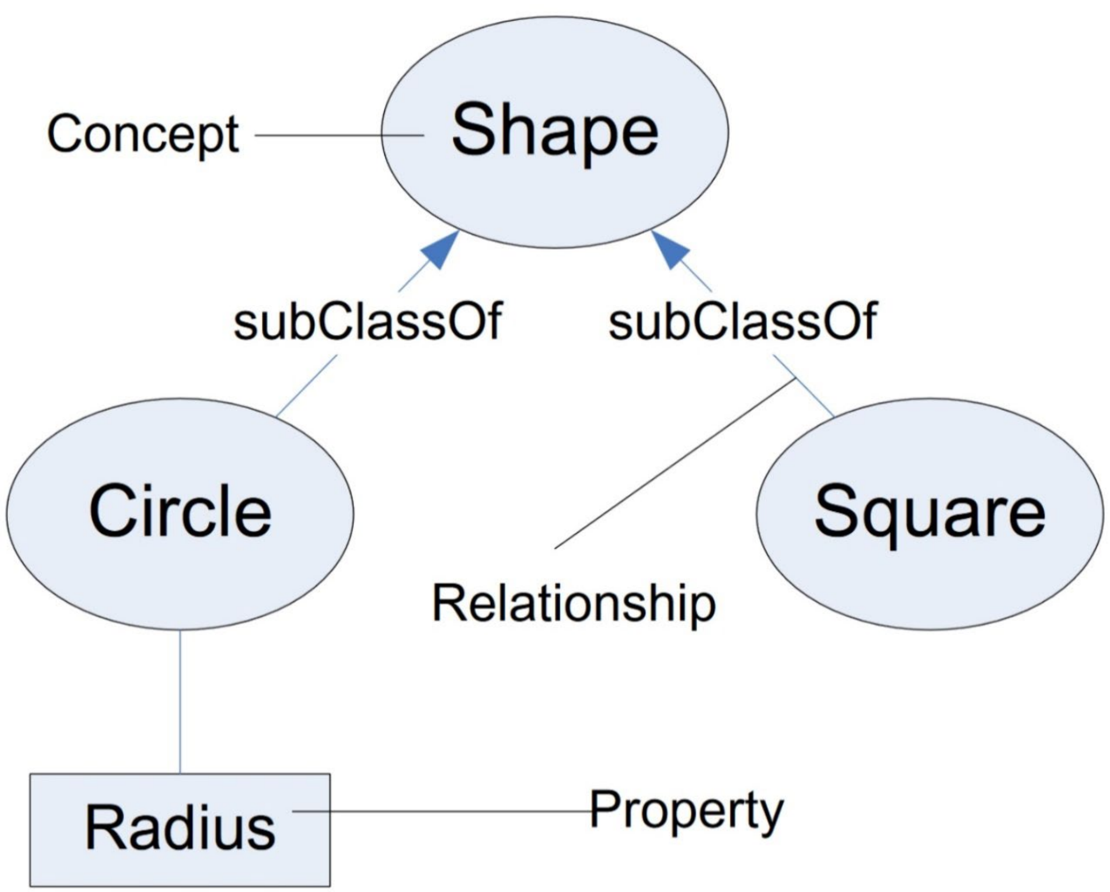
https://www.researchgate.net/publication/2701045_On_Generating_Semantic_Dispositions_in_a_Given_Subject_Domain

Knowledge Representation and Reasoning (KR&R)

Semantic networks vs ontologies

- semantic networks are more intuitive and more simple, less focused on interoperability
- ontologies are more formal and rigorous, often defined using formal languages, to represent complex and domain-specific knowledge
- ontologies are designed to enable interoperability and integration across different systems and domains

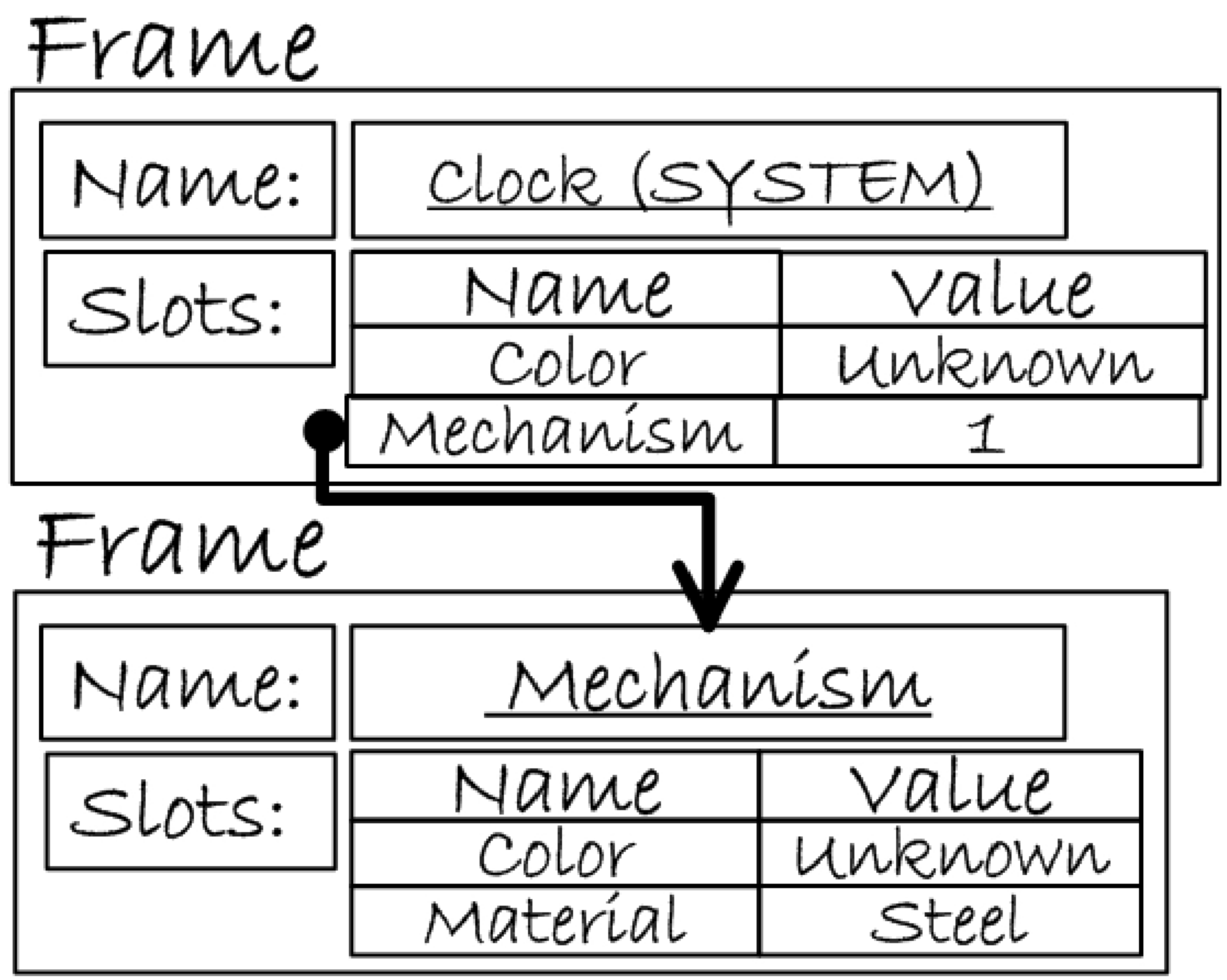




Knowledge Representation and Reasoning (KR&R)

Frame-based systems (semantic networks where nodes have structure):

- data structures dividing knowledge into substructures by representing "stereotyped situations“
- frames consist of slots (attributes) and slot fillers (values)
- example: a "dog" frame might have slots for "legs" (4), "sound" (bark), and so on

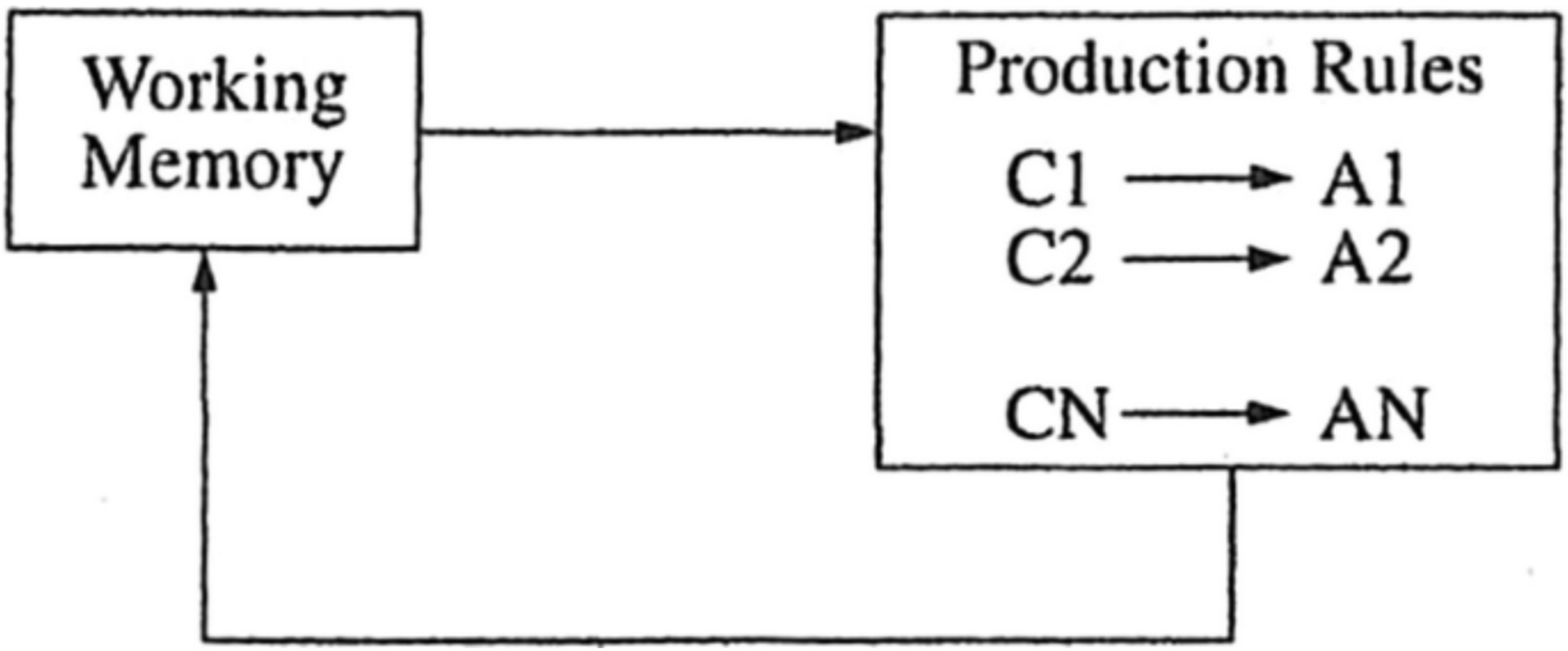


https://www.researchgate.net/publication/236647335_An_extension_of_frame-based_knowledge_representation_schema

Knowledge Representation and Reasoning (KR&R)

Production rules:

- knowledge is represented as a set of condition-action pairs (rules)
- when a condition is satisfied, the corresponding action is executed
- working memory contains currently produced results which can trigger new rules
- example: IF it is raining THEN carry an umbrella

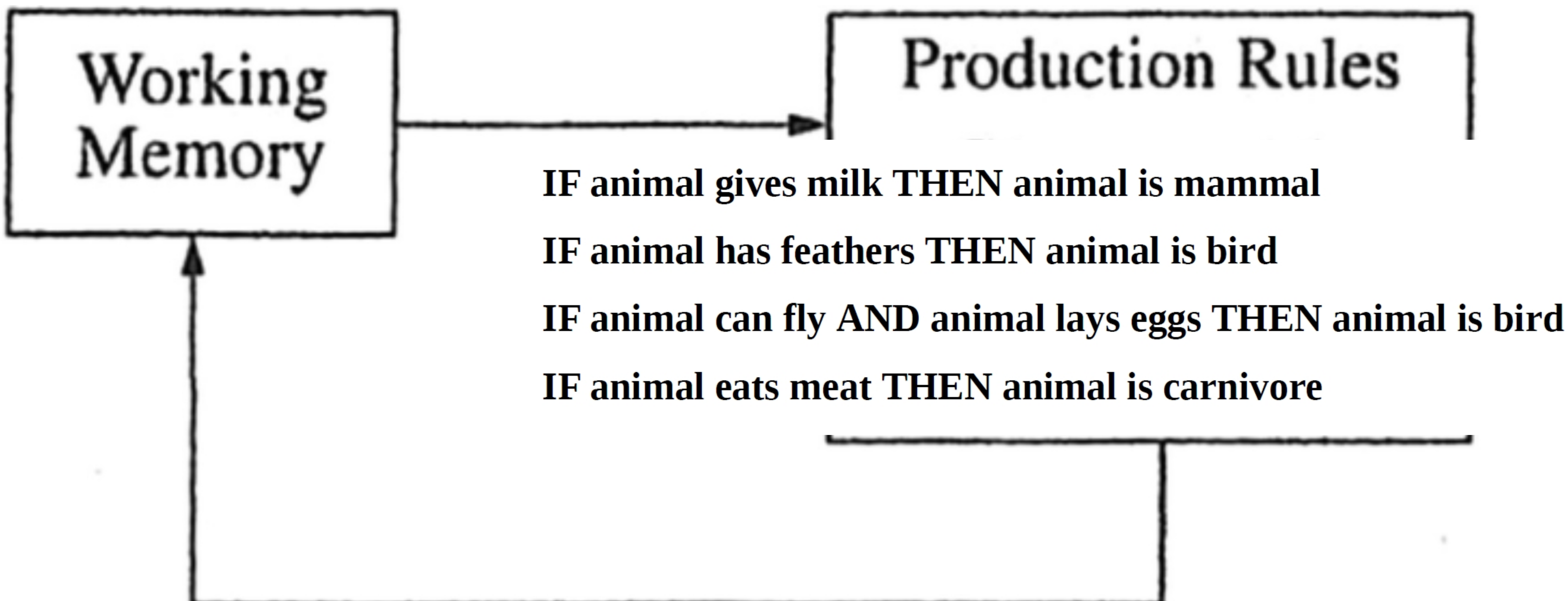


https://link.springer.com/chapter/10.1007/978-3-030-71873-2_1

Knowledge Representation and Reasoning (KR&R)

Production rules:

- knowledge is represented as a set of condition-action pairs (rules)
- when a condition is satisfied, the corresponding action is executed
- working memory contains currently produced results which can trigger new rules
- example: IF it is raining THEN carry an umbrella

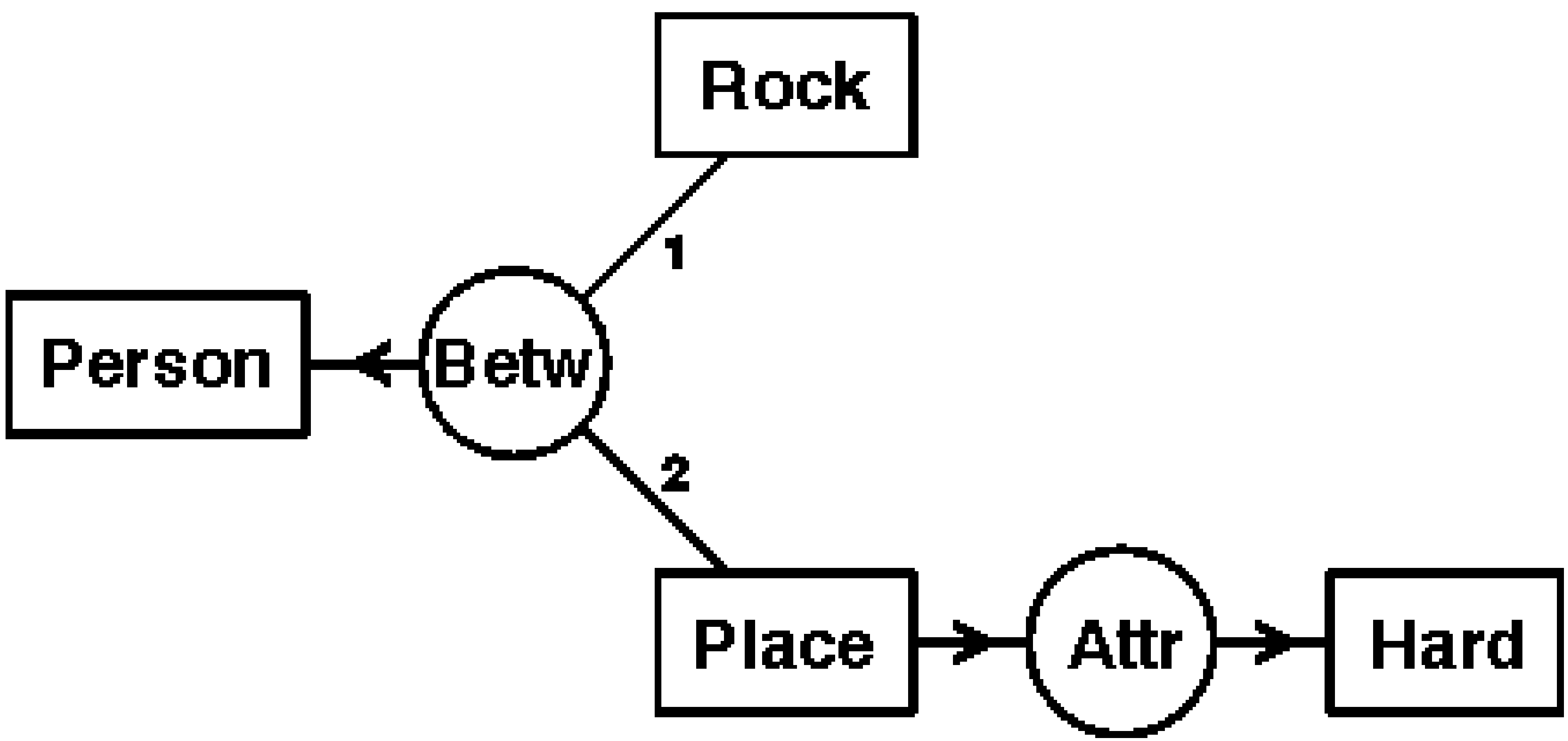


https://link.springer.com/chapter/10.1007/978-3-030-71873-2_1
M. Wooldridge, A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going, Flatiron Books, 2021.

Knowledge Representation and Reasoning (KR&R)

Conceptual graphs:

- semantic networks where nodes represent concepts and edges represent complex relations
- helps to visually represent and reason about complex relationships
- example: diagrams with sentences in natural language processing

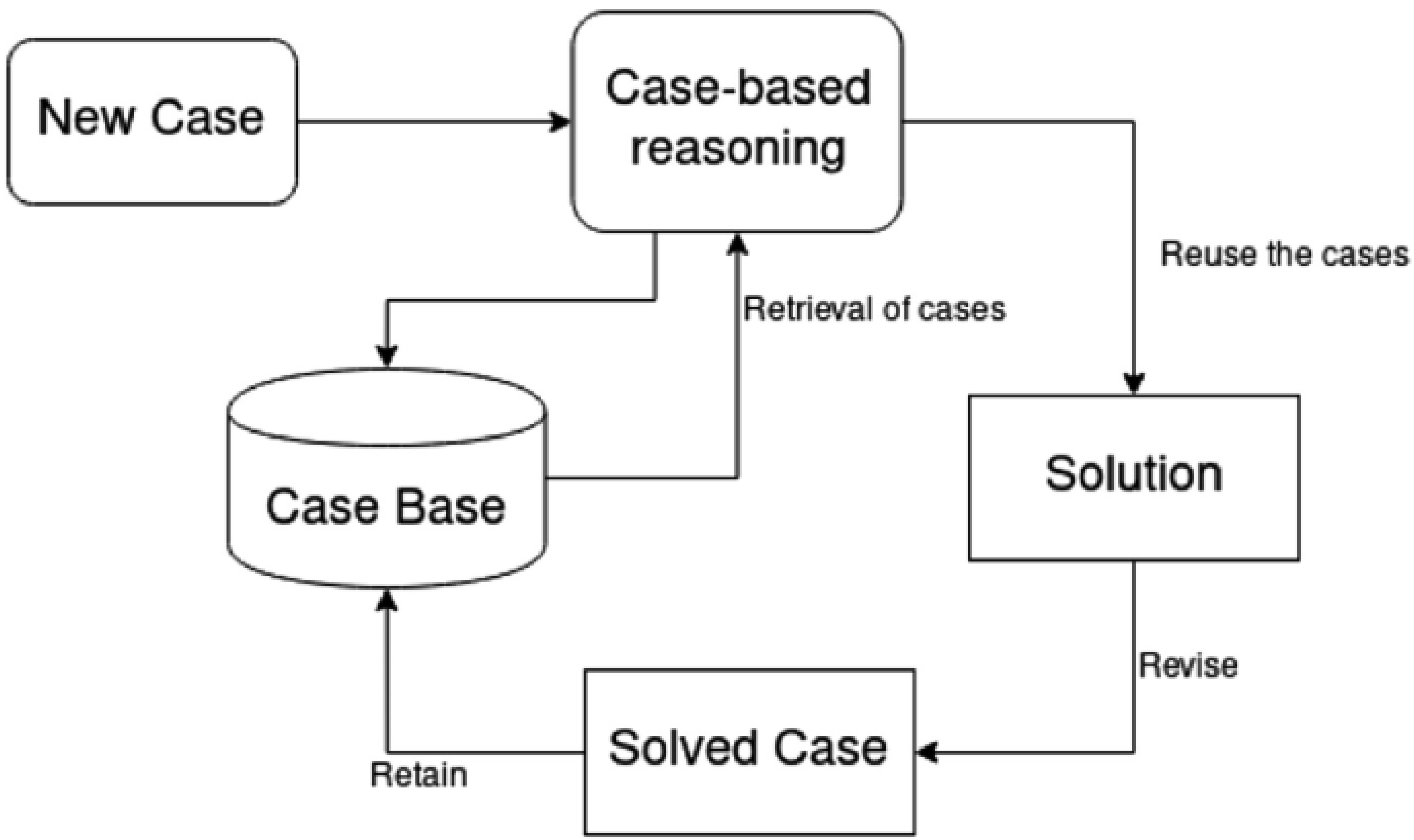


https://www.jfsowa.com/cg/cgexampw.htm

Knowledge Representation and Reasoning (KR&R)

Case-based reasoning:

- knowledge is represented as a database of past cases
- solves new problems by finding similar past cases and adapting their solutions
- example: legal reasoning systems that use past court cases to inform new cases



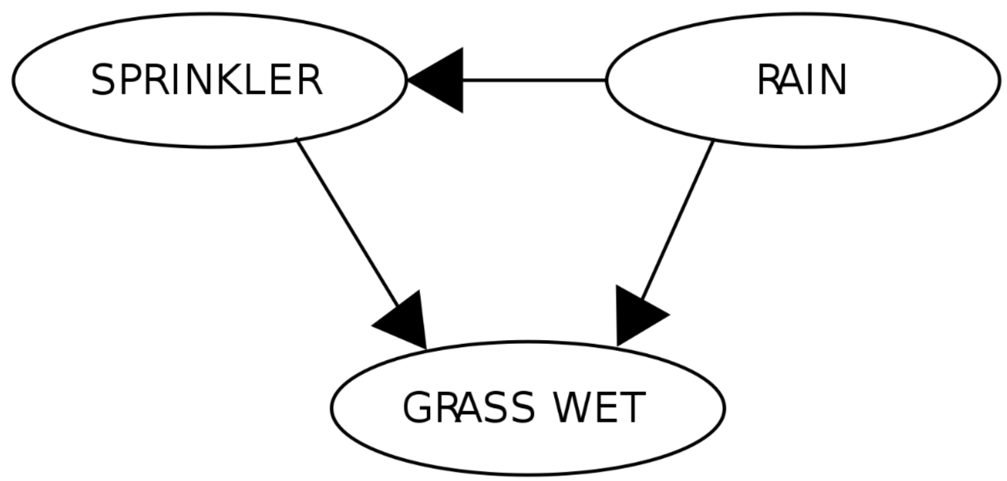
https://link.springer.com/chapter/10.1007/978-981-15-7031-5_62

Knowledge Representation and Reasoning (KR&R)

Bayesian networks:

- probabilistic graphical models representing a set of variables and their conditional dependencies
- useful for reasoning under uncertainty
- example: medical diagnosis with symptoms probabilistically linked to disease

	SPRINKLER	
RAIN	T	F
F	0.4	0.6
T	0.01	0.99



RAIN	
T	F
0.2	0.8

		GRASS WET	
SPRINKLER	RAIN	T	F
F	F	0.0	1.0
F	T	0.8	0.2
T	F	0.9	0.1
T	T	0.99	0.01

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

Break

?

Knowledge Representation and Reasoning (KR&R)

Reasoning techniques

- Deductive reasoning:
    - derives logically certain conclusions from given premises
    - example: All men are mortal; Socrates is a man; therefore, Socrates is mortal
- Inductive reasoning:
    - derives generalisations from specific observations
    - example: the sun rises every morning, therefore it will rise again tomorrow
- Abductive reasoning:
    - infers the best explanation for a set of observations
    - example: diagnosing a disease based on symptoms and medical knowledge
- Probabilistic reasoning:
    - handles uncertainty by using probability theory
    - example: Bayesian inference to update beliefs based on new evidence

Knowledge Representation and Reasoning (KR&R)

Ethical concerns:

- bias and fairness
  - ontologies and knowledge graphs that reflect and reinforce societal stereotypes (gender or racial)
  - reasoning systems may produce biased outcomes because they use incomplete or biased knowledge bases
- transparency and explainability
  - automated decision-making systems rely on complex reasoning processes, making it hard to explain their decisions to users or stakeholders.
  - knowledge-based systems in healthcare: understanding the reasoning behind a diagnosis is crucial
- accountability
  - legal decisions made by AI systems involve reasoning about laws, precedents and liabilities
  - financial decisions made by AI can have significant impacts on individuals’ lives, where accountability for mistakes or unfair decisions needs to be clear
- privacy
  - knowledge graphs with sensitive personal information, exposing individuals to privacy breaches
  - reasoning systems may infer sensitive information about individuals based on the available data

ChatGPT 4o

Knowledge Representation and Reasoning (KR&R)

Ethical concerns:

- misuse of information
  - AI systems may generate false or misleading information based on reasoning processes, contributing to misinformation campaigns
  - enhanced surveillance capabilities using knowledge representation to track and profile individuals
- intellectual property and ownership
  - disputes over the ownership of knowledge graphs created from proprietary data sources
  - legal challenges regarding the use of copyrighted material in building knowledge bases
- ethical reasoning
  - autonomous systems making decisions that involve ethical trade-offs, such as prioritising one life over another in critical situations
  - AI systems used in healthcare or social services need to consider ethical implications of their recommendations and actions

ChatGPT 4o

Knowledge Representation and Reasoning (KR&R)

Mitigation of ethical concerns:

- bias and fairness
  - ensuring diverse and representative data sources when building knowledge bases
  - regularly auditing and updating knowledge representations to identify and mitigate biases
  - developing algorithms that can detect and correct biases in knowledge and reasoning processes
- transparency and explainability
  - developing explainable AI (XAI) techniques, making the reasoning processes of AI systems more transparent and understandable
- accountability
  - establishing clear accountability frameworks that define who is responsible for the decisions made by AI systems
  - ensuring that human oversight is maintained in critical decision-making processes, with mechanisms for appeal and revision
- privacy
  - implementing strong data protection measures, such as anonymisation and encryption, to safeguard personal information
  - ensuring compliance with data privacy regulations and providing users with control over their data

ChatGPT 4o

Knowledge Representation and Reasoning (KR&R)

Mitigation of ethical concerns:

- misuse of information
  - developing ethical guidelines and regulations to prevent the misuse of AI technologies
  - building safeguards into AI systems to detect and prevent the generation and dissemination of false information
- intellectual property and ownership
  - establishing clear policies and agreements regarding the ownership and use of knowledge and data in AI systems
  - ensuring that the rights of data owners and creators are respected and fairly compensated
- ethical reasoning
  - incorporating ethical reasoning frameworks into AI systems to ensure they can handle moral dilemmas appropriately
  - engaging ethicists and stakeholders in the design and deployment of AI systems to ensure they align with societal values

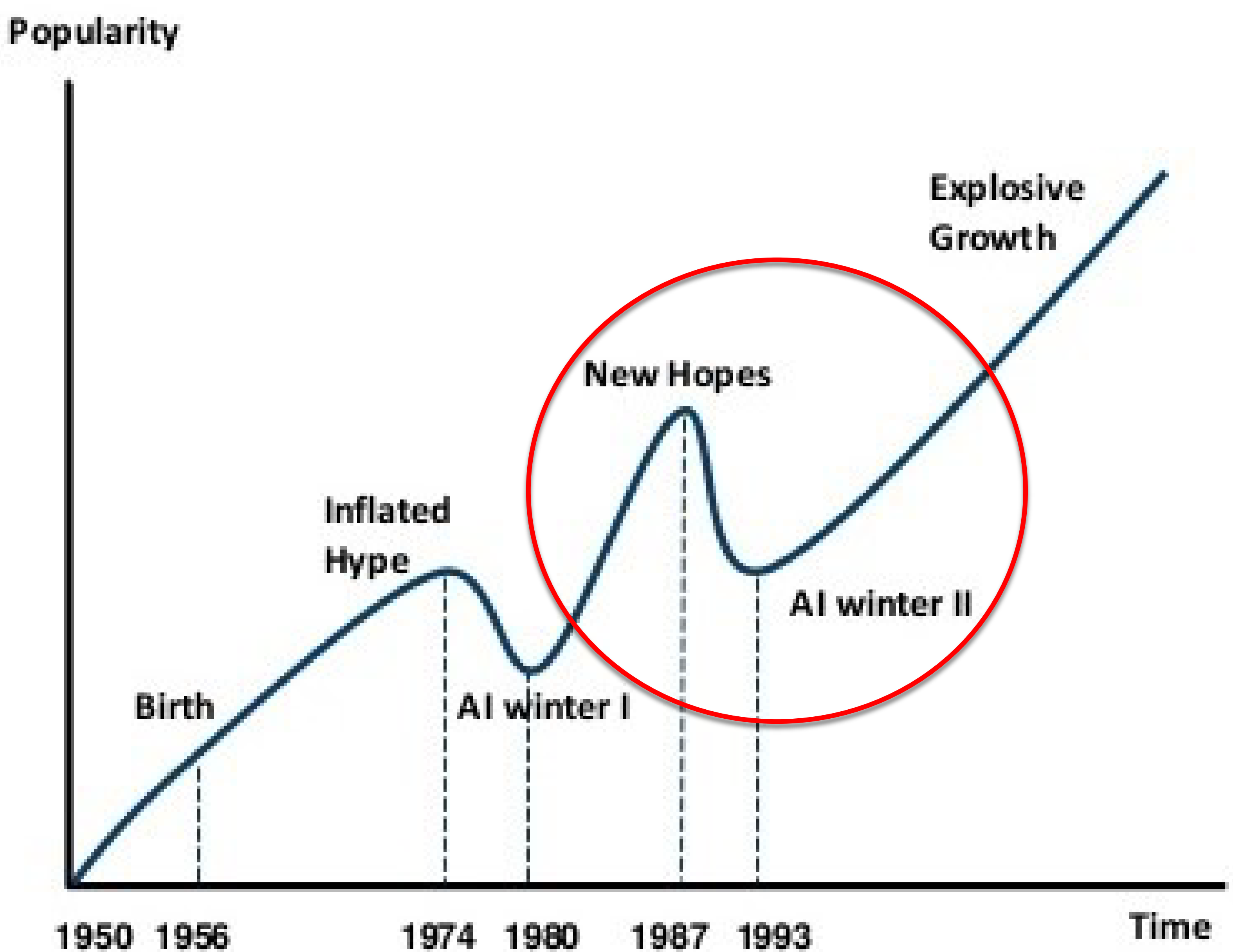
ChatGPT 4o

Break

?

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/

History of AI: main achievements and setbacks

Second AI winter (1987 – 1993)

- ❖ sudden collapse of the market for specialised AI hardware in 1987
- ❖ an entire AI industry worth half a billion dollars was crushed overnight
- ❖ expert systems were too expensive to maintain (difficult to update, could not learn, were "brittle“, could not resolve well-known problems)
- ❖ AI expectations were much higher than what was achievable
- ❖ The Defense Advanced Research Projects Agency (DARPA) decided that AI was not "the next wave"
- ❖ by the end of 1993, over 300 AI companies were shut down, became bankrupt, or were acquired by others

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

History of AI: main achievements and setbacks

New ideas and more computing power (1990s – 2010s)

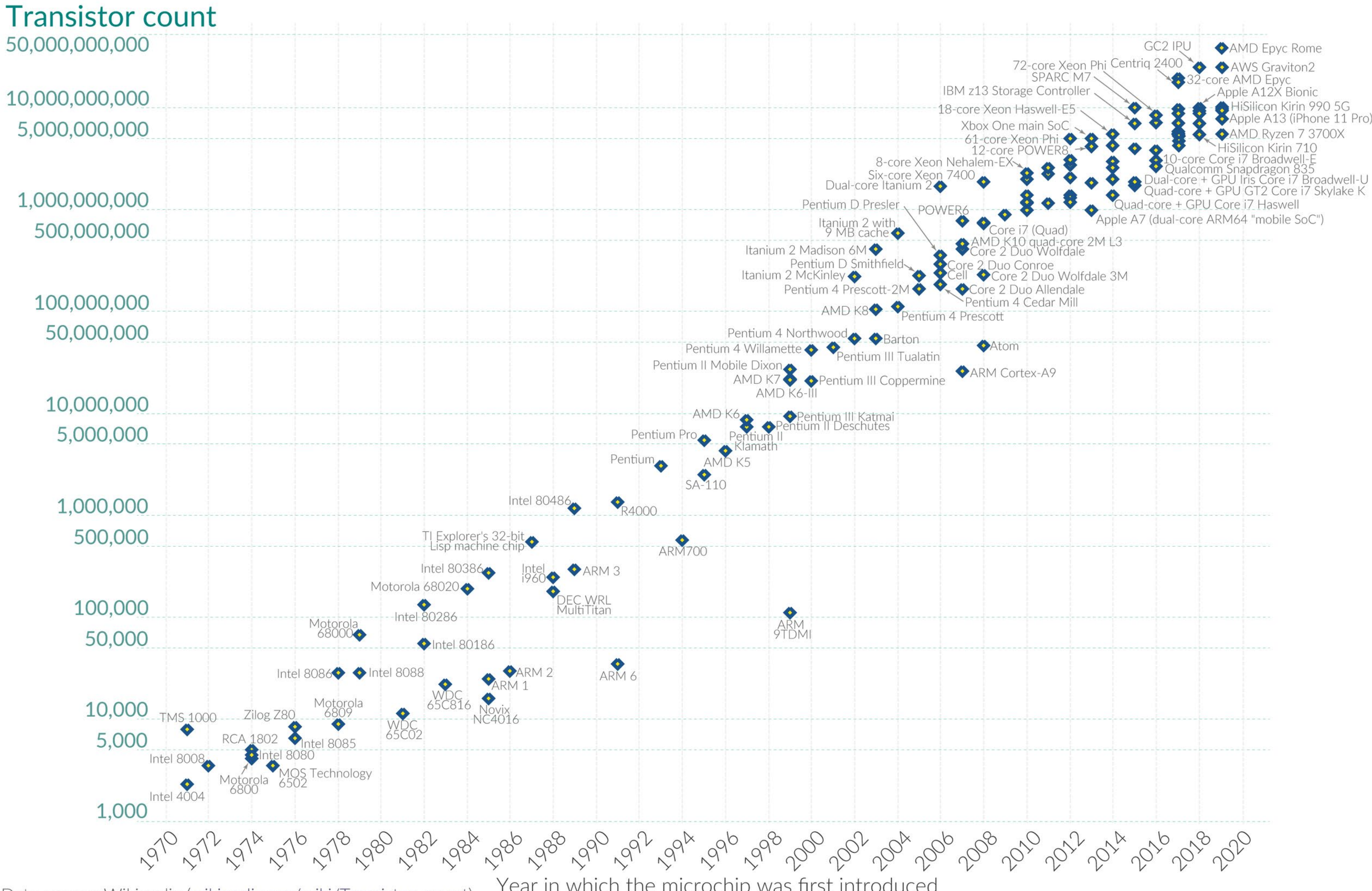
- 1997: Deep Blue (IBM) became the first computer program to beat the world chess champion, Garry Kasparov:
  - Deep Blue was processing up to 200,000,000 moves per second
- 2005: a Stanford robot won the DARPA Grand Challenge:
  - autonomous drive for 131 miles along "an unrehearsed desert trail"
- 2007: a team from CMU won the DARPA Urban Challenge:
  - autonomous navigation for 55 miles in "an urban environment while adhering to traffic hazards and all traffic laws"
- 2011: IBM's question answering system, Watson won a Jeopardy! quiz show exhibition match:
  - defeated two greatest Jeopardy! champions by a significant margin

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

Moore's Law: The number of transistors on microchips doubles every two years

Our World in Data

Moore's law describes the empirical regularity that the number of transistors on integrated circuits doubles approximately every two years.
This advancement is important for other aspects of technological progress in computing – such as processing speed or the price of computers.



Data source: Wikipedia (wikipedia.org/wiki/Transistor_count)

OurWorldinData.org – Research and data to make progress against the world's largest problems.

https://en.wikipedia.org/wiki/Moore's_law

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

New ideas (1990s – 2010s)

- Nouvelle AI and behaviour-based robotics:
  - “Elephants Don't Play Chess”, Rodney A. Brooks (1990): "the world is its own best model. It is always exactly up to date. It always has every detail there is to be known. The trick is to sense it appropriately and often enough."
- Probabilistic reasoning:
  - “Probabilistic Reasoning in Intelligent Systems: Networks of Plausible Inference”, Judea Pearl (1988)
  - “Causality: Models, Reasoning, and Inference”, Judea Pearl (2000)
- Neural networks and “artificial curiosity”:
  - adversarial neural networks: one network models a probability distribution over outputs, while another network learns (by gradient descent) to predict the reactions of the environment to these outputs, Juergen Schmidhuber (1991)
  - self-supervised learning: a recurrent neural network (RNN) is trained by generating supervisory signals itself, Juergen Schmidhuber (1992)
  - long short-term memory: an additional RNN module that learns when to remember or forget some information, Sepp Hochreiter & Juergen Schmidhuber (1997)
- Ant colony optimisation (Marco Dorigo, 1992)

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

History of AI: main achievements and setbacks

New ideas (1990s – 2010s)

- Nouvelle AI and behaviour-based robotics:
  - subsumption architecture
    - **Avoid obstacles:** If I detect an obstacle, then change direction, choosing a new direction at random.
    - **Shut down:** If I am at the docking station and have a low battery, then shut down.
    - **Empty dirt:** If I am at the docking station and am carrying dirt, then empty dirt container.
    - **Head to dock:** If the battery is low or dirt container is full, then head to docking station.
    - **Vacuum:** If I detect dirt at the current location, then switch on vacuum.
    - **Walk randomly:** Choose a direction at random and move in that direction.

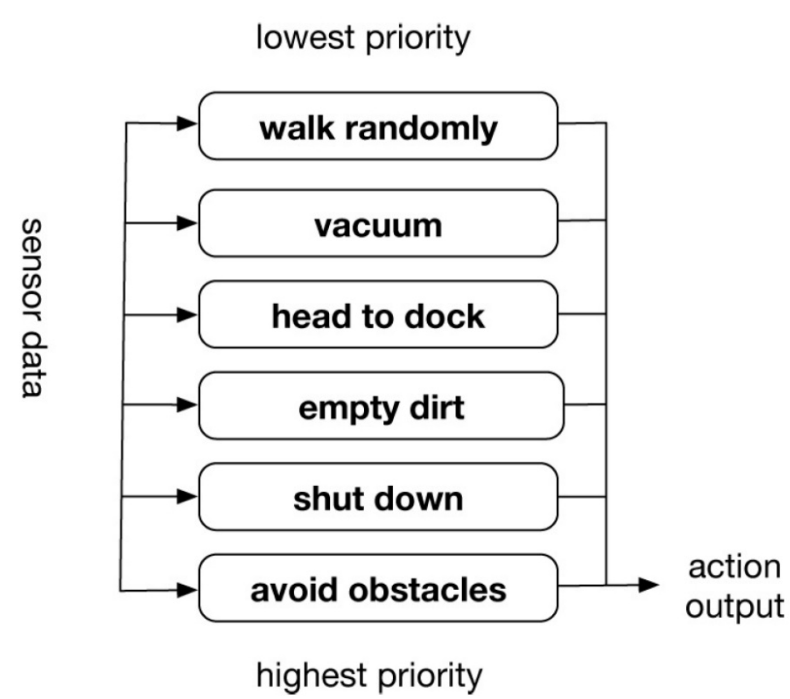
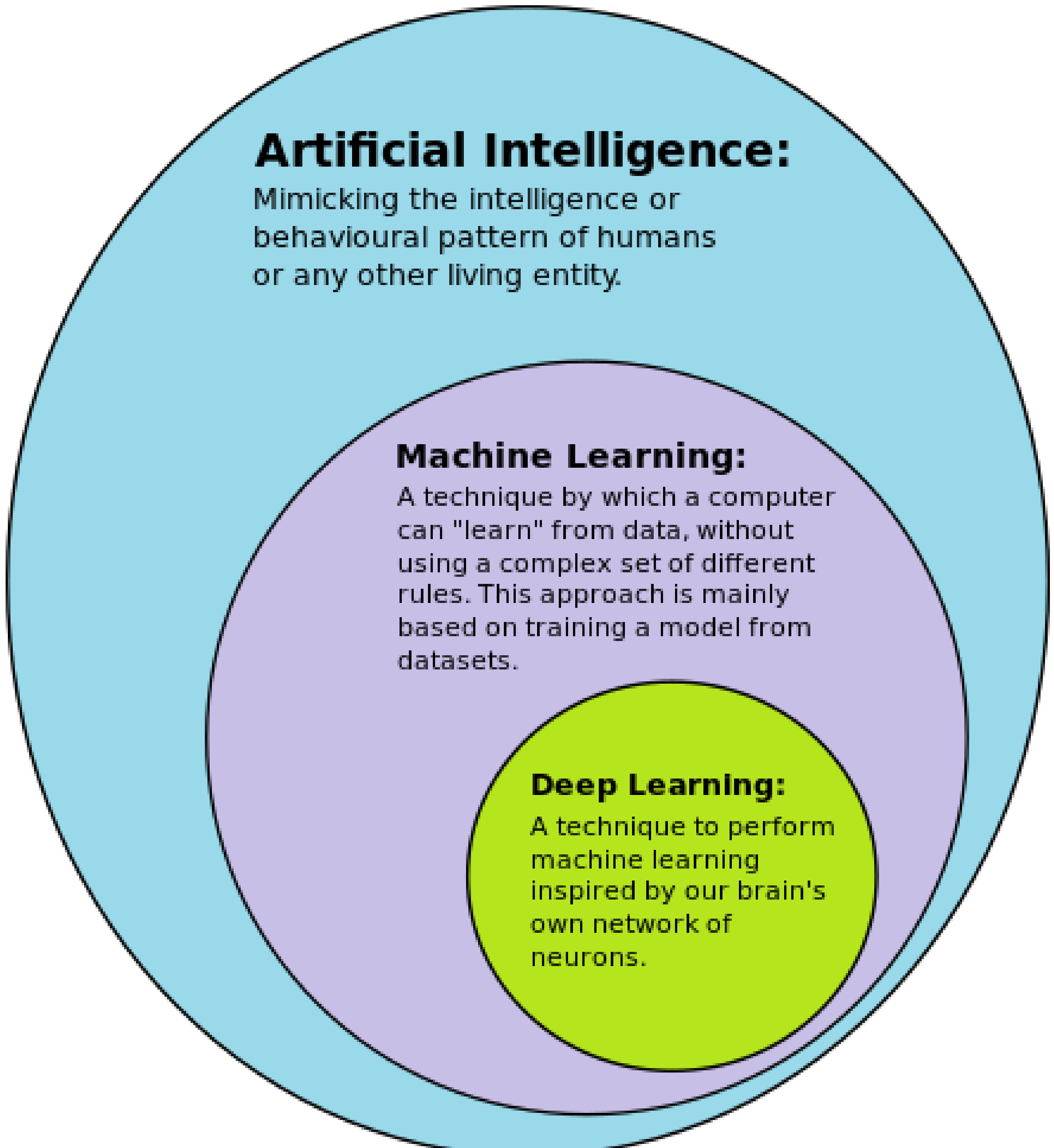


Figure 6: A simple subsumption hierarchy for a robot vacuum cleaner.

M. Wooldridge, A Brief History of Artificial Intelligence: What It Is, Where We Are, and Where We Are Going, Flatiron Books, 2021.

History of AI: main achievements and setbacks

Deep Learning: neural networks + representation learning

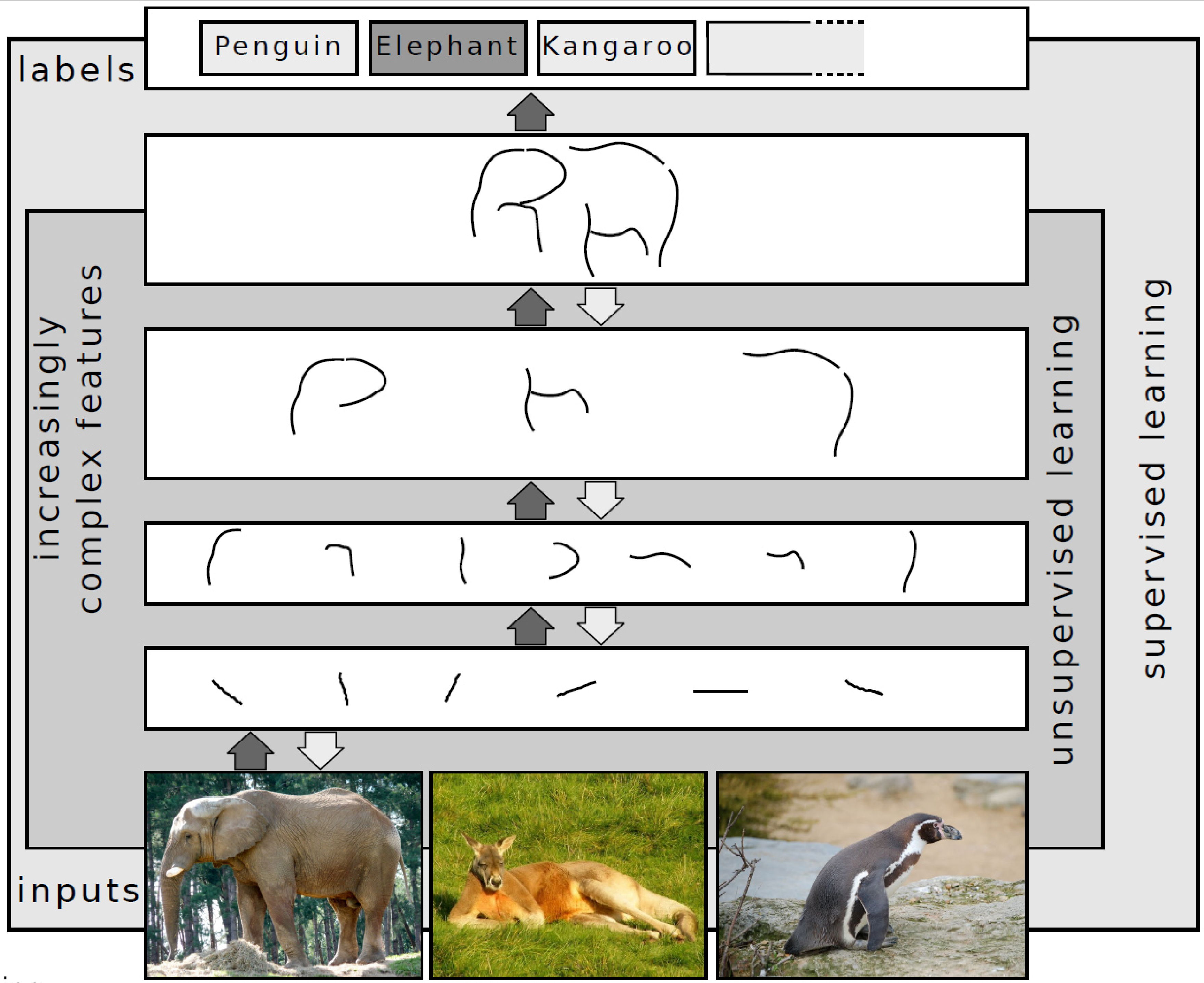


https://en.wikipedia.org/wiki/Deep_learning

History of AI: main achievements and setbacks

Deep Learning: neural networks + representation learning

- a hierarchy of layers transforming input data into composite abstractions

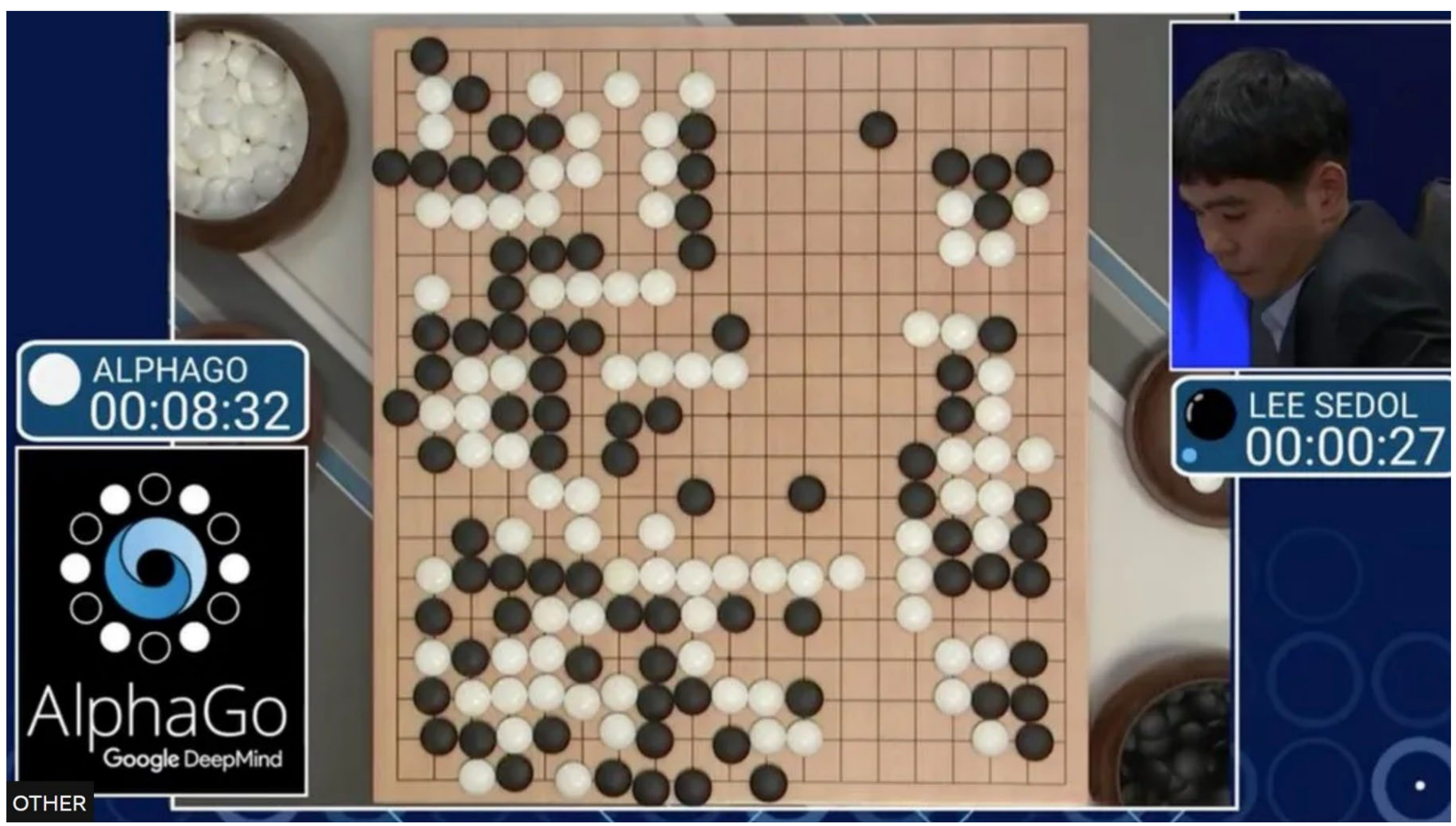


https://en.wikipedia.org/wiki/Deep_learning

History of AI: main achievements and setbacks

Deep learning and big data (2011 – 2020)

- neural network with many processing layers model high-level relationships in data
- human accuracy in computer vision, for example, in traffic sign recognition
- the 5V characteristics of big data (IBM): Volume, Velocity, Variety, Value, Veracity
- 2016: Google's DeepMind AlphaGo program defeated a Go champion, Lee Se-dol
- by 2016, the market for AI-related products exceeded 8 billion dollars



https://en.wikipedia.org/wiki/History_of_artificial_intelligence
https://www.bbc.com/news/technology-35785875

History of AI: main achievements and setbacks

Deep learning and big data (2011 – 2020)

- AlphaGo: plays Go, outperforms human champions, using a Monte Carlo tree search to find moves based on an artificial neural network extensively trained by both from human and computer play

- AlphaZero: plays Go, outperforms human champions and AlphaGo, without learning from human games (trained solely via self-play)

- AlphaStar: plays the video game StarCraft II, outperforms best human players, initially learns to imitate the moves of best human players, then plays against its own variants using reinforcement learning

- AlphaFold: performs predictions of protein structure, accurate enough to predict even single-mutation effects and the structure of complexes created by proteins with DNA and RNA (AlphaFold 3, May 2024) – using an attention network, a deep learning technique

https://en.wikipedia.org/wiki/AlphaGo
https://en.wikipedia.org/wiki/AlphaZero
https://en.wikipedia.org/wiki/AlphaStar_(software)
https://en.wikipedia.org/wiki/AlphaFold

History of AI: main achievements and setbacks

Socio-economic impacts of AlphaGo, AlphaZero, AlphaStar, AlphaFold…

- ➢ increased investment in AI and machine learning technologies
- ➢ demonstrated potential for AI to manage and optimise complex, real-time decision-making processes, in industries such as defence and logistics
- ➢ improved understanding of protein structures and accelerated scientific discovery
- ➢ significantly increased the speed of drug discovery and development processes in biotechnology and pharmaceutical industries
- ➢ reduced R&D costs and time-to-market for new therapies
- ➢ attracted substantial investment into biotech startups and research initiatives on leveraging AI for biological and medical advancements
- ❖ highlighted ethical issues: data privacy, algorithmic transparency, and the potential for AI bias
- ❖ highlighted the need for policies to ensure broader access and equitable benefits

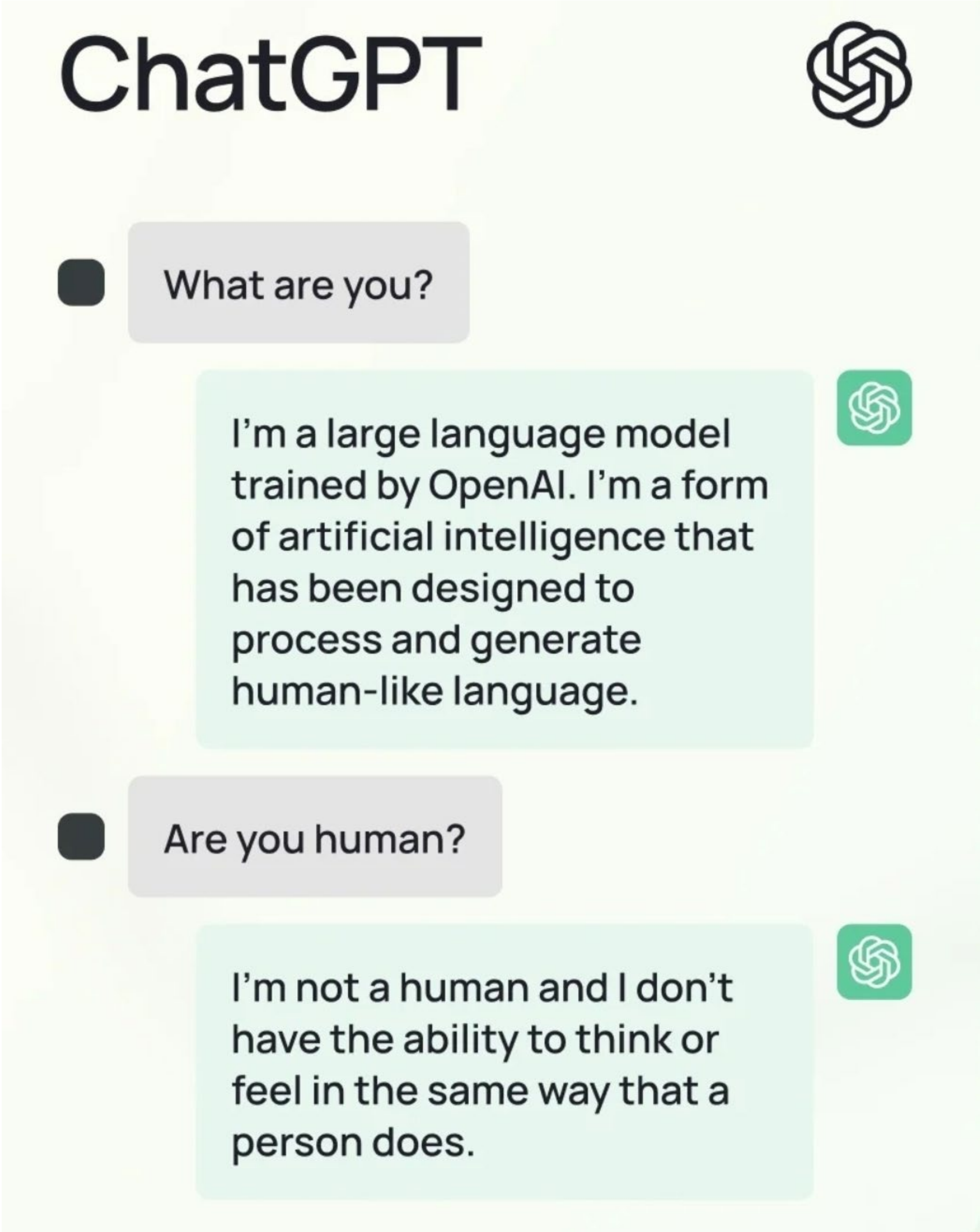
GPT-4o

History of AI: main achievements and setbacks

Large language models (2020 – present)

- 2020: GPT-3 released by OpenAI
- 2022: Gato released by DeepMind
- 2023: GPT-4 tested by Microsoft Research:

  "it could reasonably be viewed as an early (yet still incomplete) version of an artificial general intelligence (AGI) system"



https://en.wikipedia.org/wiki/History_of_artificial_intelligence
https://www.ultimate.ai/blog/ai-automation/chatgpt-the-tech-behind-the-hype-and-what-it-means-for-your-support

History of AI: main achievements and setbacks

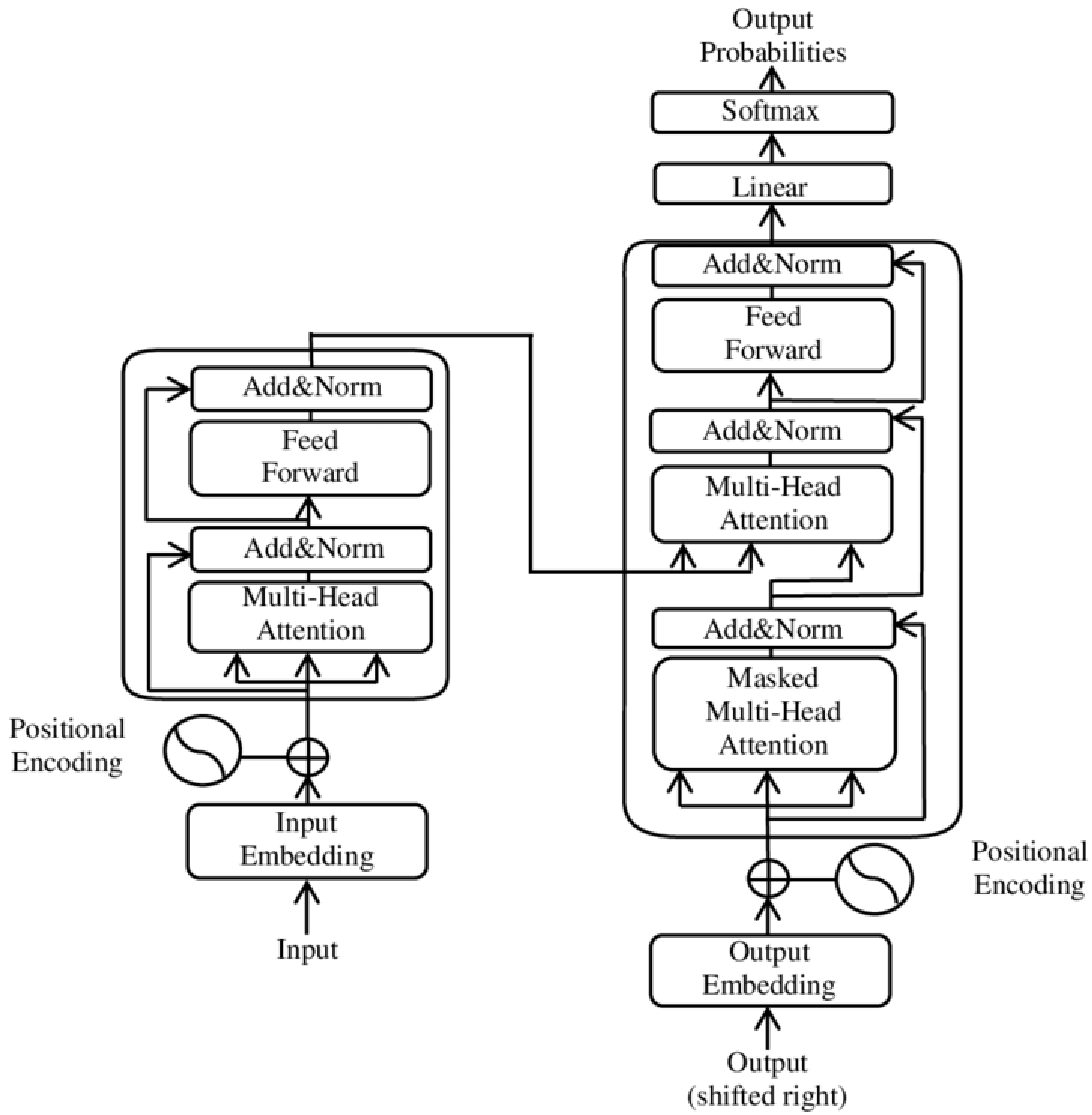
Large language models (2020 – present)

- "Attention Is All You Need" (2017, the deep learning transformer architecture proposed by Google):
  - text is converted to numerical representations (tokens), and each token is converted into a vector using a word embedding table
  - a word embedding table is a real-valued vector that encodes the meaning of the word: the words that are closer in the vector space are expected to be similar in meaning
  - transformer layers transform the vectors many times, extracting more and more linguistic information, by alternating attention and feedforward layers of neural network
  - at each layer, each token is contextualized relative to other tokens via a parallel multi-head attention mechanism: for key tokens the signal is amplified and for less important tokens the signal is diminished
  - the attention layer weighs all previous states according to a learned measure of relevance, providing relevant information about far-away tokens which can be essential for the meaning of a word in a sentence

https://en.wikipedia.org/wiki/History_of_artificial_intelligence
https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture)
https://en.wikipedia.org/wiki/Word_embedding

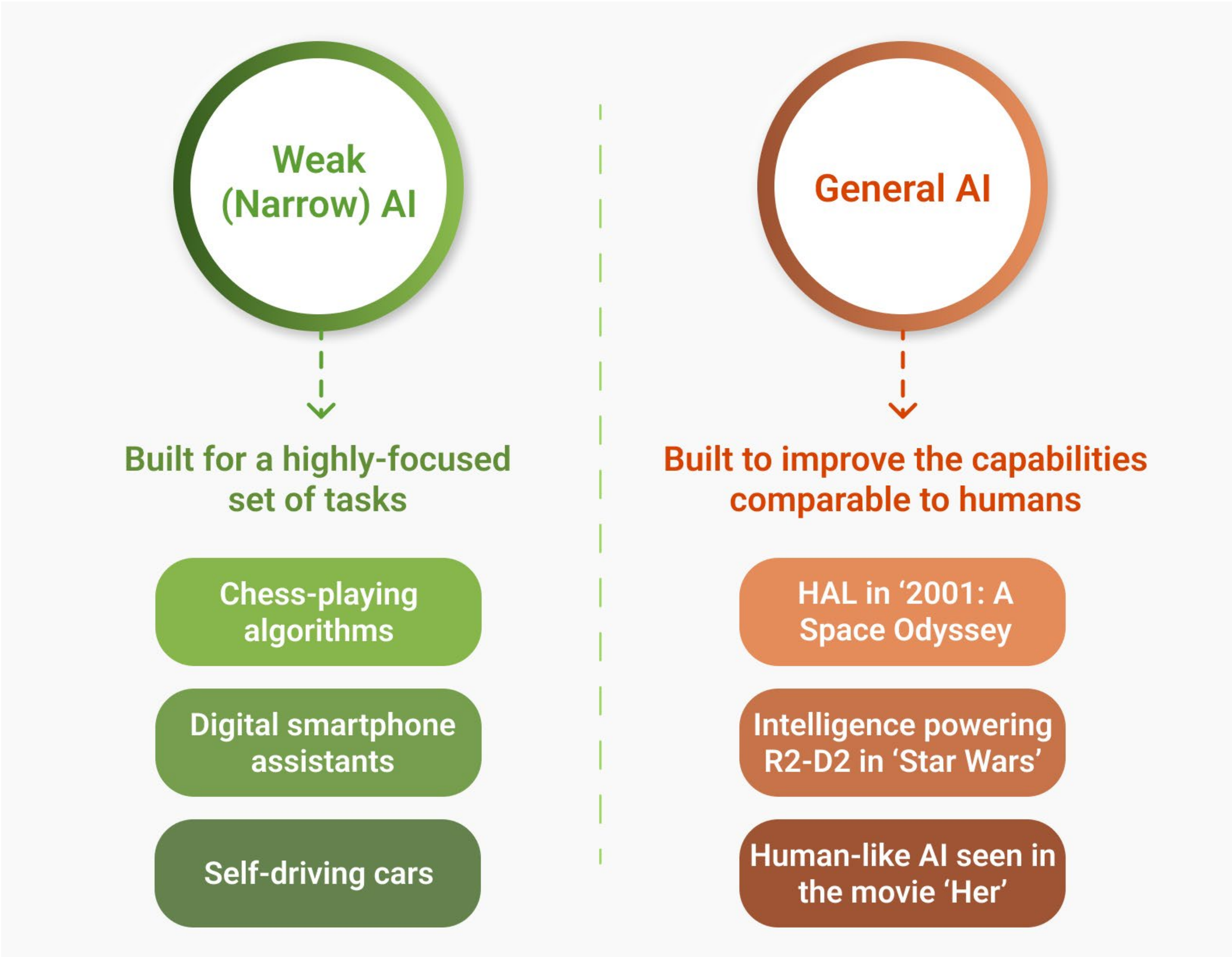
History of AI: main achievements and setbacks

"Attention Is All You Need" (2017, the transformer architecture)



https://en.wikipedia.org/wiki/Transformer_(deep_learning_architecture)

History of AI: main achievements and setbacks



https://www.plugger.ai/blog/general-ai-vs-narrow-ai-2022-guide

Homework

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

- “AlphaGo—The movie. Full documentary.”
  YouTube: https://www.youtube.com/watch?v=WXuK6gekU1Y

- please try to watch this documentary before the next lecture, and prepare to give a few brief comments about it:
  - 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?

?