COMP9208
Artificial Intelligence and Society

Lecture 1: 8 August 2025

Prof. Mikhail Prokopenko
Dr. Sheryl Chang
TA: Miss Christina Jamerlan

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

http://sydney.edu.au/campus-life/safety-security.html

Prof. Mikhail Prokopenko

Director, Centre for Complex Systems

Research Leader, Modelling and Simulation, School of Computer Science

MSc (Applied Math, 1988), MA (Economics, 1992), PhD (Computer Science, 2002)

PhD, Macquarie University: "A Unifying Semantics for Causal Reasoning about Action“ (2002)

CSIRO – NASA Ageless Aerospace Vehicles (AAV) project (2002 – 2005)

RoboCup Soccer Simulation League 2D:

The Japanese Society for Artificial Intelligence (JSAI) Award for scientific contribution (2002)

Best paper award at The 17th International RoboCup Symposium (Eindhoven, 2013)

World Champion: 2016 (Leipzig), team Gliders2016

World Champion: 2019 (Sydney), team Fractals2019

Specialty Chief Editor, Computational Intelligence, *Frontiers of Robotics and AI* (2014 – 2022)

Lord Robert May Prize for the best paper, *Journal of Biological Dy*namics (2019 – 2020)

Nature Communications: Top 50 SARS-CoV-2 articles from 2020

“Modelling transmission and control of the COVID-19 pandemic in Australia”

Email: mikhail.prokopenko@sydney.edu.au

Office: J12 (Computer Science), level 3W, room 324

Dr Sheryl Chang

PhD; MPM; BE (Aero) Hons. (USYD)

PhD: Non-linear effects of social connectivity on epidemic dynamics (2021)

Research interests and applications:

- Epidemic modelling influenced by social dynamics and human behaviours
- Recent publication in *Royal Society Interface Focus*, theme issue on "Mathematical Modeling for the Design of Ethical Infectious Disease Interventions"

Awards:

- Dean's Commendation 2023 for a Commendation in Teaching, Faculty of Engineering (2024)
- Lord Robert May Prize for the best paper biannual award (2019-2020)
- Top 50 SARS-CoV-2 articles in *Nature Communications* (2020)

Email:
sheryl.chang@sydney.edu.au

Before we begin the proceedings, I would like to acknowledge and pay respect to the traditional owners of the land on which we meet; the Gadigal people of the Eora Nation. It is upon their ancestral lands that the University of Sydney is built.

As we share our own knowledge, teaching, learning and research practices within this university may we also pay respect to the knowledge embedded forever within the Aboriginal Custodianship of Country.

Introductions

?

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): opinions

- the word “intellect” is derived from Latin “intellego” (inter-lego): to choose between, or to perceive a main message given some alternatives
- “A computer would deserve to be called intelligent if it could deceive a human into believing that it was human.” (Alan Turing)
- “Even as humans, we tend to acknowledge people capable of seeing the “big picture” — the full scope of future impacts of seemingly small changes — as more intelligent.” (Terry B. Bollinger)
- "There's a great phrase, written in the '70s: 'The definition of today's AI is a machine that can make a perfect chess move while the room is on fire.' It really speaks to the limitations of AI. In the next wave of AI research, if we want to make more helpful and useful machines, we've got to bring back the contextual understanding." (Fei-Fei Li)
- "Early AI was mainly based on logic. You're trying to make computers that reason like people. The second route is from biology: You're trying to make computers that can perceive and act and adapt like animals." (Geoffrey Hinton)
- “The thing that's going to make artificial intelligence so powerful is its ability to learn, and the way AI learns is to look at human culture." (Dan Brown)
- “By their very nature, heuristic shortcuts will produce biases, and that is true for both humans and artificial intelligence, but the heuristics of AI are not necessarily the human ones.” (Daniel Kahneman)

Discussion

?

About this course

This unit introduces basic concepts and methods of Artificial Intelligence (AI) tailored to science, technology, engineering, mathematics, medicine, business, and humanities students.

AI is aimed to design machines that can replicate and surpass human intelligence.

AI agents include any artificial system that senses its environment, takes actions towards its goals, and improves its performance with experience (e.g., self-driving car, face recognition, automatic language translation, humanoid robot, fraud detection, virtual assistant, and so on).

This unit covers foundations of modern applications of AI, relating real-world scenarios to simple examples and essential AI methods, such as search, planning and machine learning.

Learning Objectives

1. Understand and analyse the behaviour of AI systems

2. Understand the nature, structure, function and evolution of AI systems in different fields

3. Understand limitations of AI solutions and their socio-economic impacts

4. Evaluate AI systems that satisfy functional criteria within given application domains

5. Examine AI methods using system-level thinking

6. Select different approaches to designing AI systems in diverse real-world settings

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)

What will you do?

- Learn about different AI approaches, methods, applications and future trends
- Discuss societal impacts and limitations of modern AI, including biases and risks
- Review and analyse case studies, report and present

Assignments?

There are 3 assignment activities for this subject:

- Individual Report: article review (30%) – week 7: 18/09, Thursday, 11:59 pm
- Group Presentation (40%) – week 11: 23/10, Thursday, 11:59 pm
- Individual Report: data analysis (30%) – week 14: 17/11, Monday, 11:59 pm

The 2nd assignment will be carried out by groups of four or more, as determined by lecturer, depending on student numbers.

TURNITIN will be used for all text-based written assignments.

In accordance with University policy, penalties apply when work is submitted after 11:59pm on the due date (Sydney time): a penalty of 5% per day; after ten calendar days late, a mark of zero will be awarded

Marking individual report (article review)

	Weight	0-49%	50-64%	65-74%	75-84%	85-100%	Mark
Background and contextual analysis	20%	Not placed in the context of AI research and applications	Partially placed in the context of AI research and applications	Placed in the context of AI research and applications, with a brief or unstructured description	Placed in the context of AI research and applications, with a brief but structured description	Placed in the context of AI research and applications, with a detailed and structured description	
Novelty and societal implications	20%	Significantly mis-identified novelties and implications	Some novelties and implications identified with a brief or unstructured description	Majority of novelties and implications identified with a brief or unstructured description	Majority of novelties and implications identified with a brief but structured description	Majority of novelties and implications identified with a detailed and structured description	
Comparative research impact	20%	No clear or coherent structure or organisation	Poor or incomplete structure and organisation	Structure and organisation are partially appropriate	Structure and organisation are appropriate	Structure and organisation are appropriate and effective	
Future impact	20%	Impact described inadequately	Impact described incompletely	Impact described but without sufficient detail, structure or argumentation	Impact described in sufficient detail, but without strong argumentation	Impact described adequately, with sufficient detail, structure and argumentation	
Ethics of AI: limitations: biases and risks	20%	Significantly mis-identified limitations	Some limitations identified with a brief or unstructured description	Majority of limitations identified but without sufficient detail, structure or argumentation	Majority of limitations identified in sufficient detail but without strong argumentation	Majority of limitations fully identified, with sufficient detail, structure and argumentation	
Total	100%						

Marking group presentation (case studies)

	Weight	0-49%	50-64%	65-74%	75-84%	85-100%	Mark
Structure and organisation	30%	No clear or coherent structure or organisation	Poor or incomplete structure and organisation	Structure and organisation are partially appropriate	Structure and organisation are appropriate	Structure and organisation are appropriate and effective	
Topic development	50%	Development of content limited; may be incomplete or unclear. Inadequate supporting information, illustration or explanations. Societal impact and limitations described inadequately or mis-identified	Development of content adequate but lacks clearly stated propositions, argument or supporting information. Societal impact and limitations described incompletely. Some use of data, graphical or other visual material to support content but meaning and rationale for use may be questionable.	Clear and complete development of content and support for propositions and arguments. Societal impact and limitations described, but without sufficient detail, structure or argumentation. Content partially supported with relevant and well explained data, graphical or other visual material.	Full development of content and support for propositions and arguments, with purpose, relevance, focus, explanations. Societal impact and limitations described in sufficient detail, but without strong argumentation. Content supported with some data, graphical or other visual material.	Full and rich development of content and support for propositions and arguments, with purpose, relevance, focus, explanations. Societal impact and limitations described adequately, with sufficient detail, structure and argumentation. Content supported effectively with relevant and well explained data, graphical or other visual material.	
Presentation (slides)	20%	Lack of fluency in expression makes it difficult to comprehend	Lack of fluency in expression but still relatively easy to follow	Good level of fluency and clarity in expression and evidence of logical progression of ideas	High level of fluency and clarity in expression and evidence of logical progression of ideas	Sophistication in fluency of expression	
Total	100%						

Marking individual report (data analysis)

	Weight	0-49%	50-64%	65-74%	75-84%	85-100%	Mark
Structure and organisation	30%	No clear or coherent structure or organisation	Poor or incomplete structure and organisation	Structure and organisation are partially appropriate	Structure and organisation are appropriate	Structure and organisation are appropriate and effective	
Data analysis	50%	Significantly incomplete results. Societal impact and limitations described inadequately or mis-identified	Partially complete results. Societal impact and limitations described incompletely.	Complete results. Societal impact and limitations described, but without sufficient detail, structure or argumentation.	Complete results. Societal impact and limitations described in sufficient detail, but without strong argumentation.	Complete results. Societal impact and limitations described adequately, with sufficient detail, structure and argumentation.	
Written description	20%	Lack of fluency in expression makes it difficult to comprehend	Lack of fluency in expression but still relatively easy to follow	Good level of fluency and clarity in expression and evidence of logical progression of ideas	High level of fluency and clarity in expression and evidence of logical progression of ideas	Sophistication in fluency of expression	
Total	100%						

Assignments (example)

There are 3 assignment activities for this subject:

(1)	Individual Report (30%)	– week 7	55/100 → 16.5 / 30
(2)	Group Presentation (40%)	– week 11	64/100 → 25.6 / 40
(3)	Individual Report (30%)	– week 14	78/100 → 23.4 / 30
	Total (100%)		**65.5 ≈ 66**

- Text matching software (Turnitin) will be used for plagiarism detection in (1) and (3)

Questions

?

Generative Artificial Intelligence

1. Do not enter confidential, personal, proprietary or otherwise sensitive information into ChatGPT or other cloud-based generative AI platforms
    - Once information is fed into a cloud-based generative AI tool, it can be incorporated into the datasets that the tool uses to generate content for all users and could therefore be presented to another user at a later date (including those maliciously seeking to collect and aggregate it)

2. Do not rely on the accuracy of outputs
    - Be aware that generative AI tools like ChatGPT often produce incorrect information
    - Generative AI tools use probabilistic models that are trained on enormous datasets to generate plausible new content: they use existing, uncorroborated content from many sources and can produce inaccurate, biased or creatively fictitious content. Repeated use of a generative AI tool can create different content each time it is run
    - You remain responsible for your work: you must independently verify and edit AI-generated content to ensure the integrity, accuracy, and suitability of the output

3. Openly acknowledge your use of generative AI

4. Check for restrictions and conditions

https://intranet.sydney.edu.au/strategy-governance/gen-ai.html

Typical advice on using generative AI

1. Assistance purely with the language
    - paraphrasing or polishing the author's original content, rather than for suggesting new content is typically acceptable

2. Literature search
    - generative AI may be used for search assistance, e.g. to identify relevant literature
    - authors must read and discuss such references
    - ensure citation accuracy
    - consider possible biases in suggested citations

3. Low-novelty text
    - provide relevant citations for both the source of the text and and the source of the idea, including quotes for verbatim copying

4. New ideas or content
    - acknowledge the use of the generative AI model
    - check for known sources for the ideas and acknowledge them too
    - ensure that new content is "correct, coherent, original and does not have missing citations"

https://2023.aclweb.org/blog/ACL-2023-policy/

ChatGPT: Avianca airlines legal case

1. A passenger of Colombia-based Avianca Airlines alleged that during a 2019 flight he was “struck by a metal serving cart” and suffered personal injuries

2. A member of the passenger’s legal team, Steven Schwartz, used ChatGPT to conduct legal research for a federal court filing

3. ChatGPT found at least six other cases to show precedent, including Varghese v. China Southern Airlines and Shaboon v. Egypt Air, and assured the lawyer (Steven Schwartz) that the cases were real

4. The federal court found that the cases did not exist and had “bogus judicial decisions with bogus quotes and bogus internal citations”, and the judge considered sanctions

5. Steven Schwartz admitted he used ChatGPT but argued he did not intend to deceive the court, saying that he “did not understand it was not a search engine, but a generative language-processing tool.”

6. Scwhartz’s team wrote: “Mr. Schwartz and the Firm have already become the poster children for the perils of dabbling with new technology; their lesson has been learned.”

https://www.forbes.com/sites/mollybohannon/2023/06/08/lawyer-used-chatgpt-in-court-and-cited-fake-cases-a-judge-is-considering-sanctions/?sh=4b7aa3f7c7f3

Discussion

?

Responsible Generative AI Principles (Elsevier)

1. Consider the real-world impact of AI solutions on people
   - “are people’s health, livelihood (including career prospects) or rights affected in any way?”

2. Take action to prevent the creation or reinforcement of unfair bias
   - output may “lead to less favourable outcomes for individuals or groups based on gender, ethnicity, socio-economic status, and other personal attributes”

3. Explain how AI solutions work
   - “appropriate level of transparency for each application and use-case to ensure that different users can understand and trust the output”

4. Create accountability through human oversight
   - “important that humans have ownership and accountability over the development, use, and outcomes of AI systems”

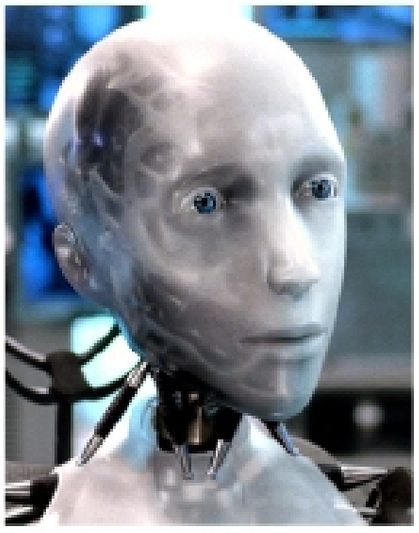
5. Respect privacy and champion robust data governance
   - ensure “robust data management and security policies and procedures”

https://www.elsevier.com/about/policies-and-standards/responsible-ai-principles

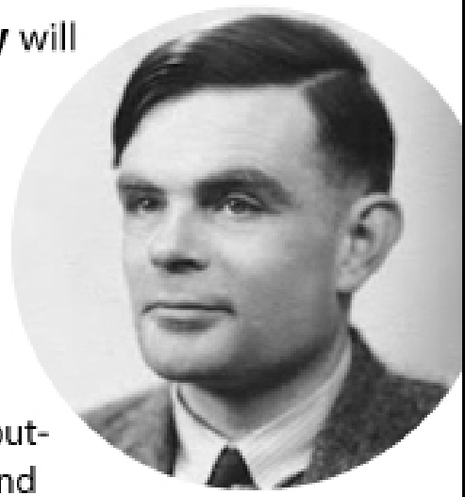
Timeline of AI: science and science fiction

HAL 9000 FROM "2001: A SPACE ODYSSEY" (CREDIT: WARNER BROS. STUDIOS)

Advances in artificial intelligence (AI) have given the world computers that can beat people at chess and "Jeopardy!," as well as drive cars and manage calendars. But despite the progress, engineers are still years away from developing machines that are self-aware. Some believe the resulting **technological singularity** will eradicate poverty and disease, while others warn it could endanger human survival.



1950: Isaac Asimov publishes the influential sci-fi story collection "**I, Robot.**" (Left: 2004 film version of "I, Robot")



1950: Alan Turing introduces the **Turing test** in his paper "Computing Machinery and Intelligence."(Credit: National Portrait Gallery, London)

1950s

Summer of 1956: Dartmouth conference launches the field of AI and **coins the term "artificial intelligence."** (Right: room-filling IBM-702 computer, as used by first AI researchers)

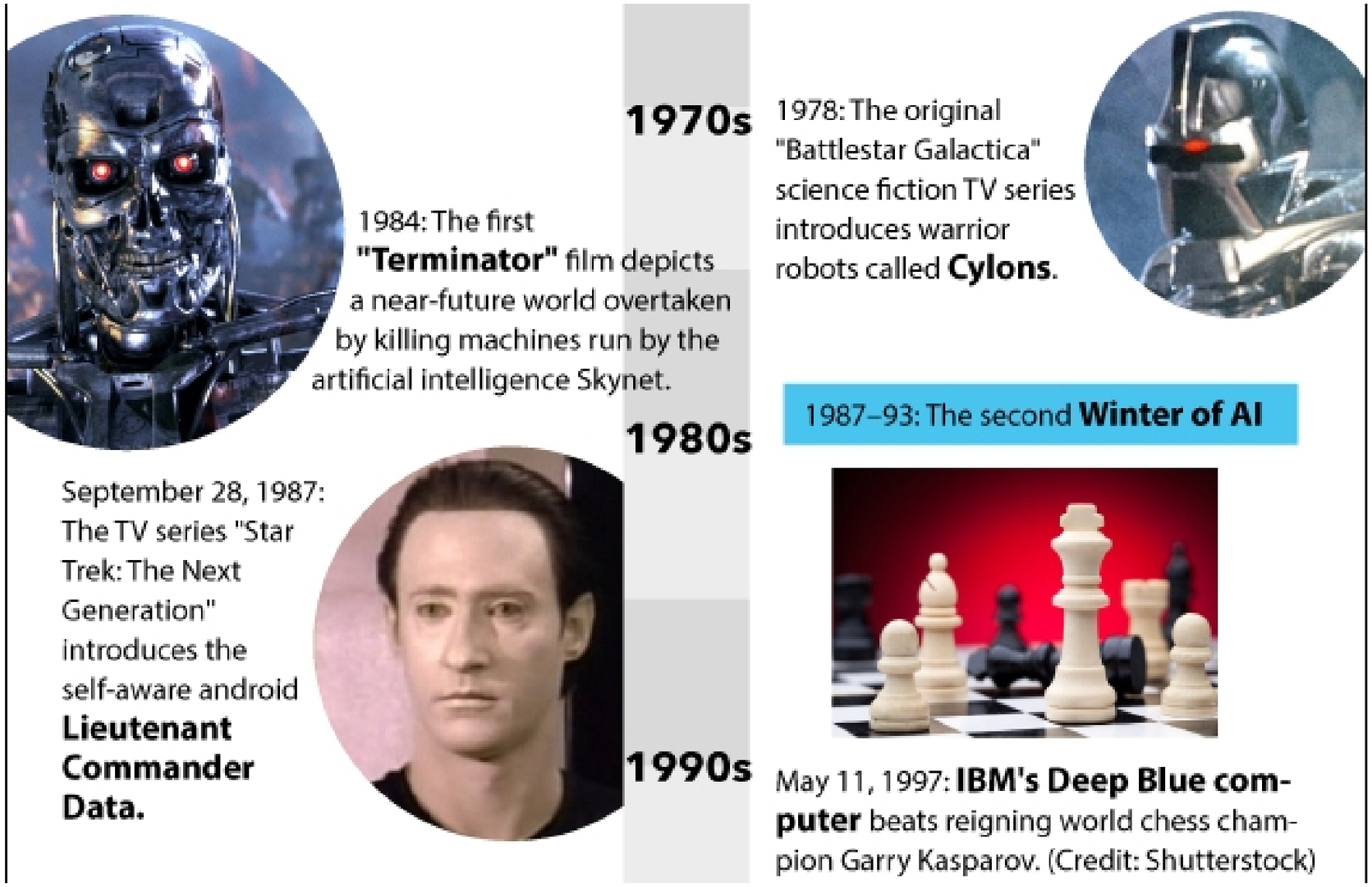
1968: "2001: A Space Odyssey," the book by Arthur C. Clarke and film by Stanley Kubrick, features the sentient and deadly computer **HAL 9000.**

1960s

1974-early 1980s: The first **Winter of AI**, a period of reduced funding and lowered interest in the field as hype turned to disappointment.

https://www.livescience.com/47544-history-of-a-i-artificial-intelligence-infographic.html

Timeline of AI: science and science fiction



https://www.livescience.com/47544-history-of-a-i-artificial-intelligence-infographic.html

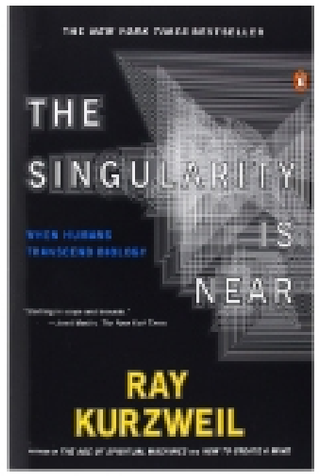
Timeline of AI: science and science fiction

June 29, 2001: Steven Spielberg releases his version of a film – originally developed by Stanley Kubrick – about a robot boy: "**A.I.: Artificial Intelligence**."

2000s

2005: A Stanford vehicle wins the **DARPA grand challenge**, driving autonomously across the desert for 131 miles (211 kilometers).

2005: Inventor and futurist Ray Kurzweil predicts an event he calls **the Singularity** will occur around 2045, when the intelligence of artificial minds exceeds that of the human brain.



2010s

2011: **IBM's Watson wins "Jeopardy!**," beating former champions Brad Rutter and Ken Jennings. (Credit: "Jeopardy!" screengrab from Wikimedia)

October 14, 2011: Apple introduces intelligent personal assistant **Siri** on the iPhone 4S.

June 2012: A Google Brain computer cluster **trains itself to recognize a cat** from millions of images in YouTube videos. (Credit: Shutterstock)

https://www.livescience.com/47544-history-of-a-i-artificial-intelligence-infographic.html

Timeline of AI: science and science fiction

recognize a cat from millions of images in YouTube videos. (Credit: Shutterstock)

December 18, 2013: The movie "Her" (left), stars Joaquin Phoenix as a man who **falls in love with his artificially intelligent computer operating system**, voiced by Scarlett Johansson.

April 10, 2014: The film "Transcendence" (below) stars Johnny Depp as an AI researcher whose **mind is uploaded to a computer** and develops into a super-intelligence.

June 7, 2014: Chatbot Eugene Goostman is said to have **passed the Turing test** in University of Reading competition, launching controversy.

August, 2014: Researchers call for creation of a **new Turing test**, to be decided at 2015 workshop.

Sources:
http://aitopics.org/misc/brief-history
http://loebner.net/Prizef/TuringArticle.html
http://www.pcworld.com/article/219893/ibm_watson_vanquishes_human_jeopardy_foes.html
http://www.engadget.com/2011/10/04/iphone-4s-hands-on/
http://www.reading.ac.uk/news-and-events/releases/PR583836.aspx
http://www.livescience.com/47296-turing-test-needs-an-update.html
http://www.nytimes.com/2012/06/26/technology/in-a-big-network-of-computers-evidence-of-machine-learning.html?pagewanted=all&_r=0

KARL TATE, TANYA LEWIS / © LiveScience.com

https://www.livescience.com/47544-history-of-a-i-artificial-intelligence-infographic.html

Break

?

Artificial Intelligence (AI): predictions

- “Artificial intelligence will reach human levels by around 2029. Follow that out further to, say, 2045, and we will have multiplied the intelligence – the human biological machine intelligence of our civilization – a billion-fold.” (Ray Kurzweil)
- “The pace of progress in artificial intelligence (I’m not referring to narrow AI) is incredibly fast. Unless you have direct exposure to groups like Deepmind, you have no idea how fast — it is growing at a pace close to exponential. The risk of something seriously dangerous happening is in the five-year time frame. 10 years at most.” (Elon Musk)
- “AI will probably most likely lead to the end of the world, but in the meantime, there'll be great companies." (Sam Altman)
- “Success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last, unless we learn how to avoid the risks.” (Stephen Hawking)
- "By the end of the decade, AI data centers could consume as much as 20% to 25% of US power requirements. Today, that's ~4% or less." (Rene Haas)
- AI is hitting the global labour market “like a tsunami”, and “is likely to impact 60% of jobs in advanced economies and 40% of jobs around the world in the next two years… It could bring tremendous increase in productivity if we manage it well, but it can also lead to more misinformation and, of course, more inequality in our society.” (Kristalina Georgieva)

Artificial Intelligence (AI): socio-economic impacts

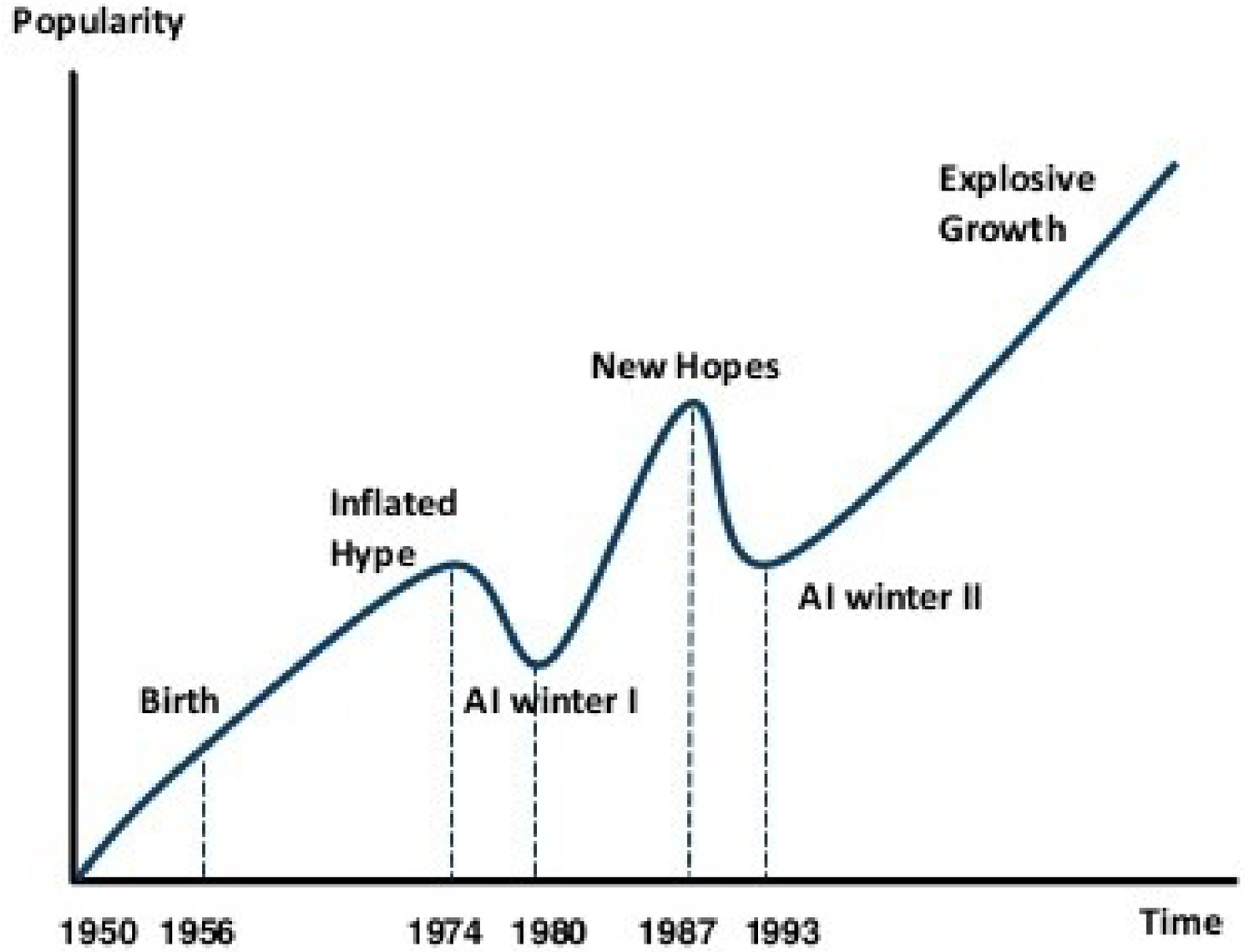
- "It is difficult to think of a major industry that AI will not transform. This includes healthcare, education, transportation, retail, communications, and agriculture. There are surprisingly clear paths for AI to make a big difference in all of these industries." (Andrew Ng)
- "It's very clear that 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)

Discussion

?

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

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".

A. M. Turing (1950) Computing Machinery and Intelligence. *Mind 49:* 433-460.

COMPUTING MACHINERY AND INTELLIGENCE

By A. M. Turing

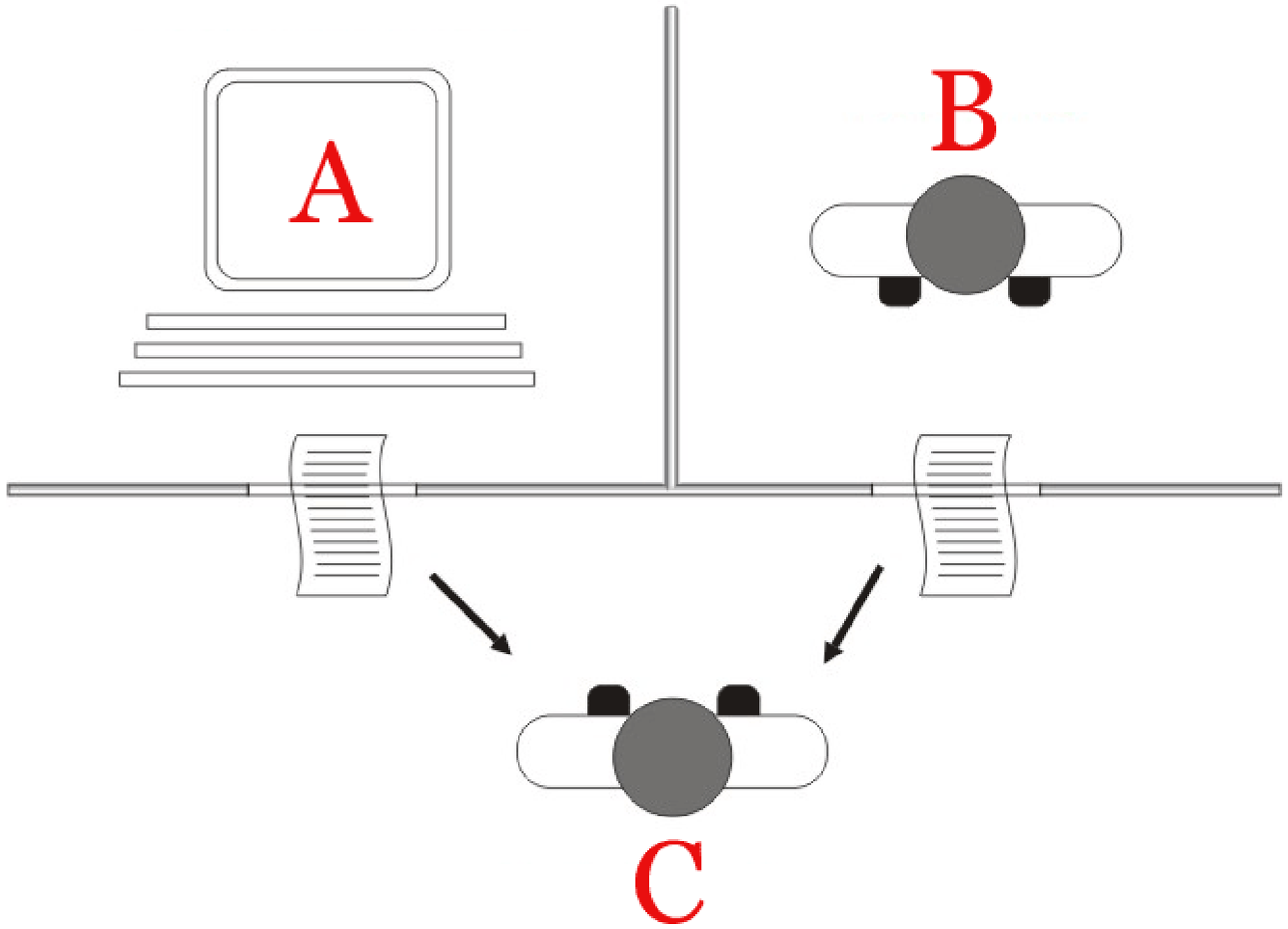
1. The Imitation Game

I propose to consider the question, "Can machines think?" This should begin with definitions of the meaning of the terms "machine" and "think." The definitions might be framed so as to reflect so far as possible the normal use of the words, but this attitude is dangerous, If the meaning of the words "machine" and "think" are to be found by examining how they are commonly used it is difficult to escape the conclusion that the meaning and the answer to the question, "Can machines think?" is to be sought in a statistical survey such as a Gallup poll. But this is absurd. Instead of attempting such a definition I shall replace the question by another, which is closely related to it and is expressed in relatively unambiguous words.

https://redirect.cs.umbc.edu/courses/471/papers/turing.pdf

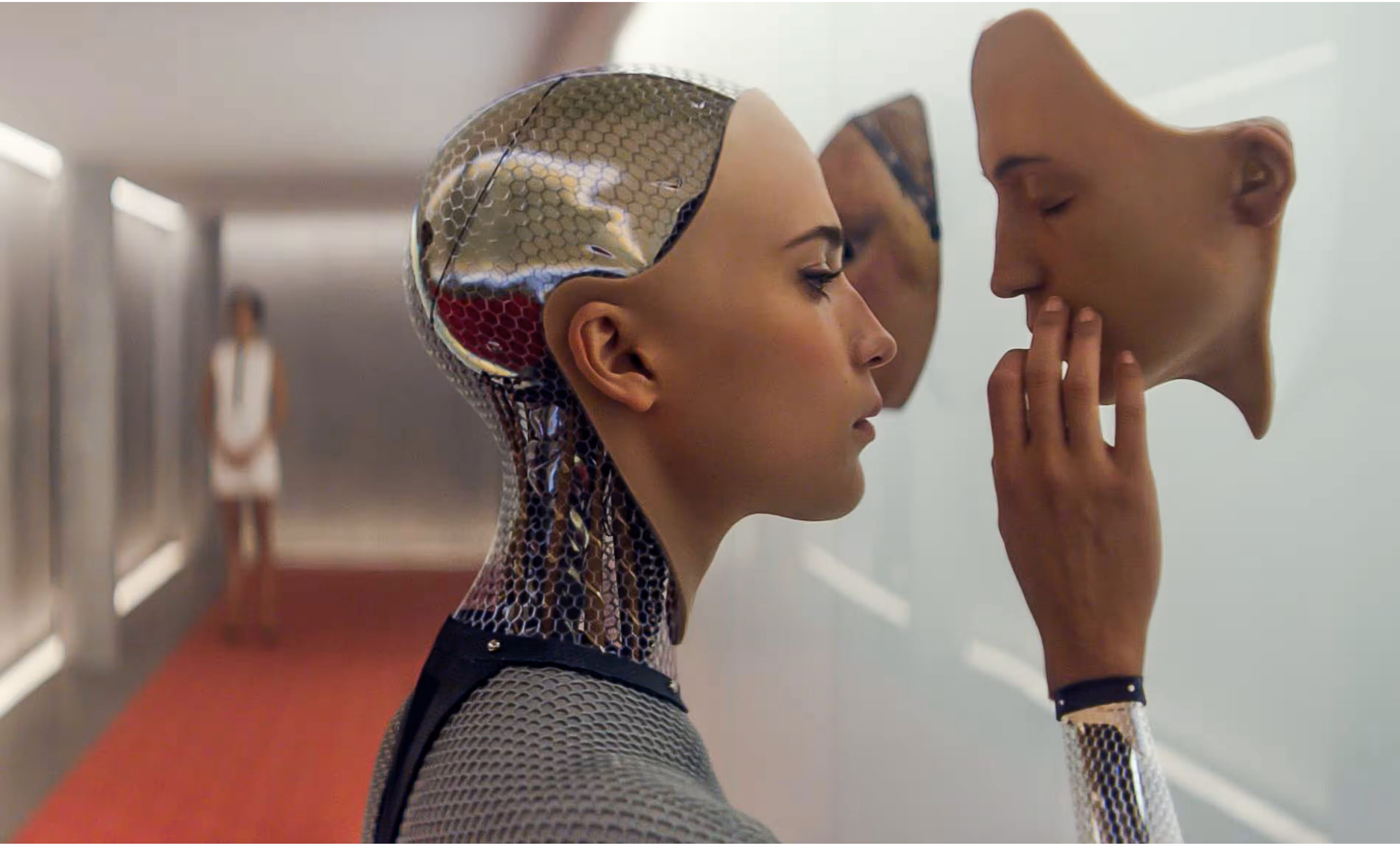
Turing test

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

Turing test: movie “Ex Machina”(2014)



https://www.theguardian.com/film/2022/may/03/ex-machina-when-you-see-this-sci-fi-thriller-youll-never-stop-thinking-about-it

IN THIS BUILDING DURING THE SUMMER OF 1956

JOHN McCARTHY (DARTMOUTH COLLEGE), MARVIN L. MINSKY (MIT)
NATHANIEL ROCHESTER (IBM), AND CLAUDE SHANNON (BELL LABORATORIES)
CONDUCTED

THE DARTMOUTH SUMMER RESEARCH PROJECT ON ARTIFICIAL INTELLIGENCE

FIRST USE OF THE TERM "ARTIFICIAL INTELLIGENCE"

FOUNDING OF ARTIFICIAL INTELLIGENCE AS A RESEARCH DISCIPLINE

"To proceed on the basis of the conjecture
that every aspect of learning or any other feature of intelligence
can in principle be so precisely described that a machine can be made to simulate it."

IN COMMEMORATION OF THE PROJECT'S 50th ANNIVERSARY
JULY 13, 2006

https://aaai.org/ojs/index.php/aimagazine/article/view/1911/1809

Photographer: Joe Mehling

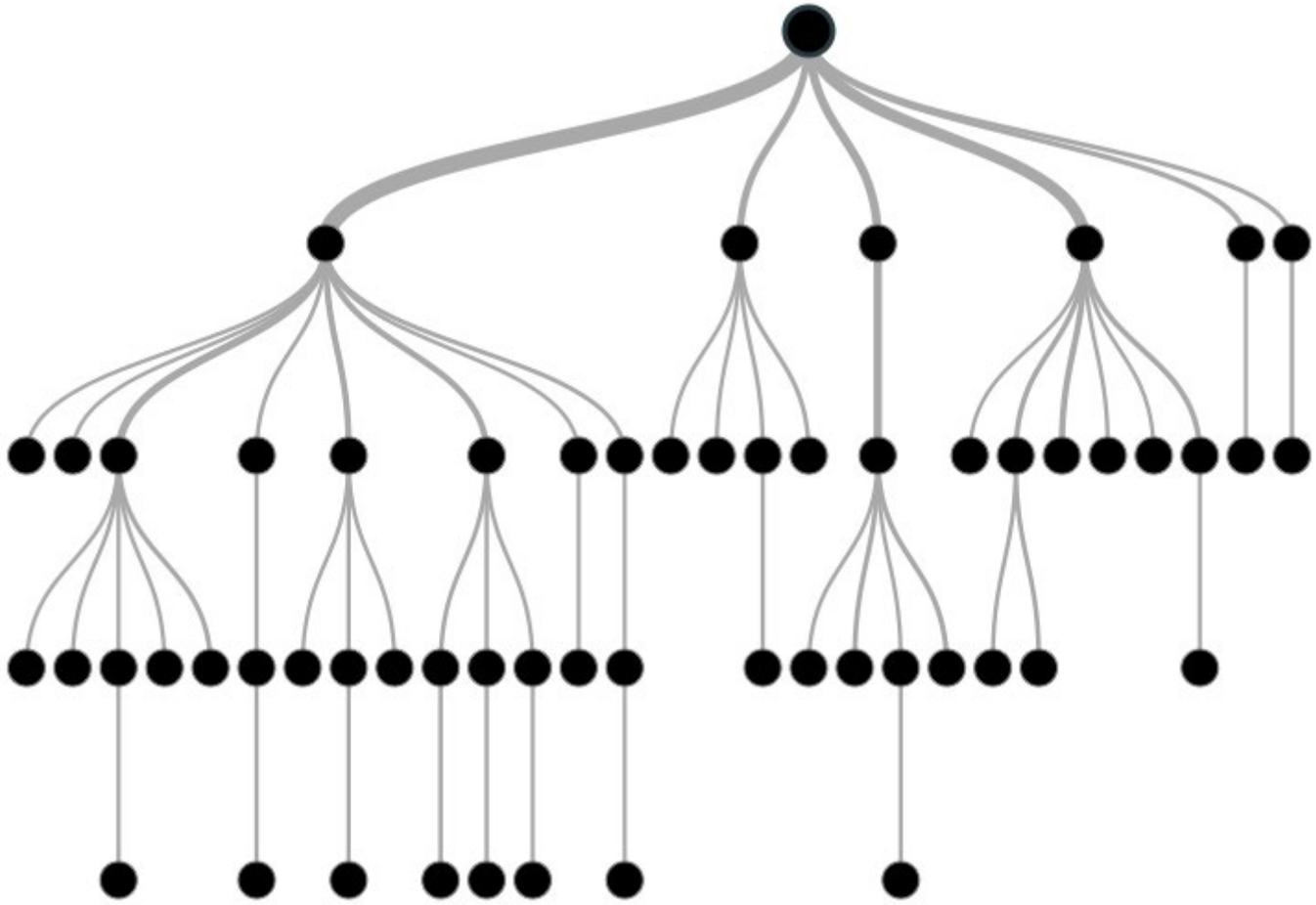
Figure 1. Trenchard More, John McCarthy, Marvin Minsky, Oliver Selfridge, and Ray Solomonoff.

https://aaai.org/ojs/index.php/aimagazine/article/view/1911/1809

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

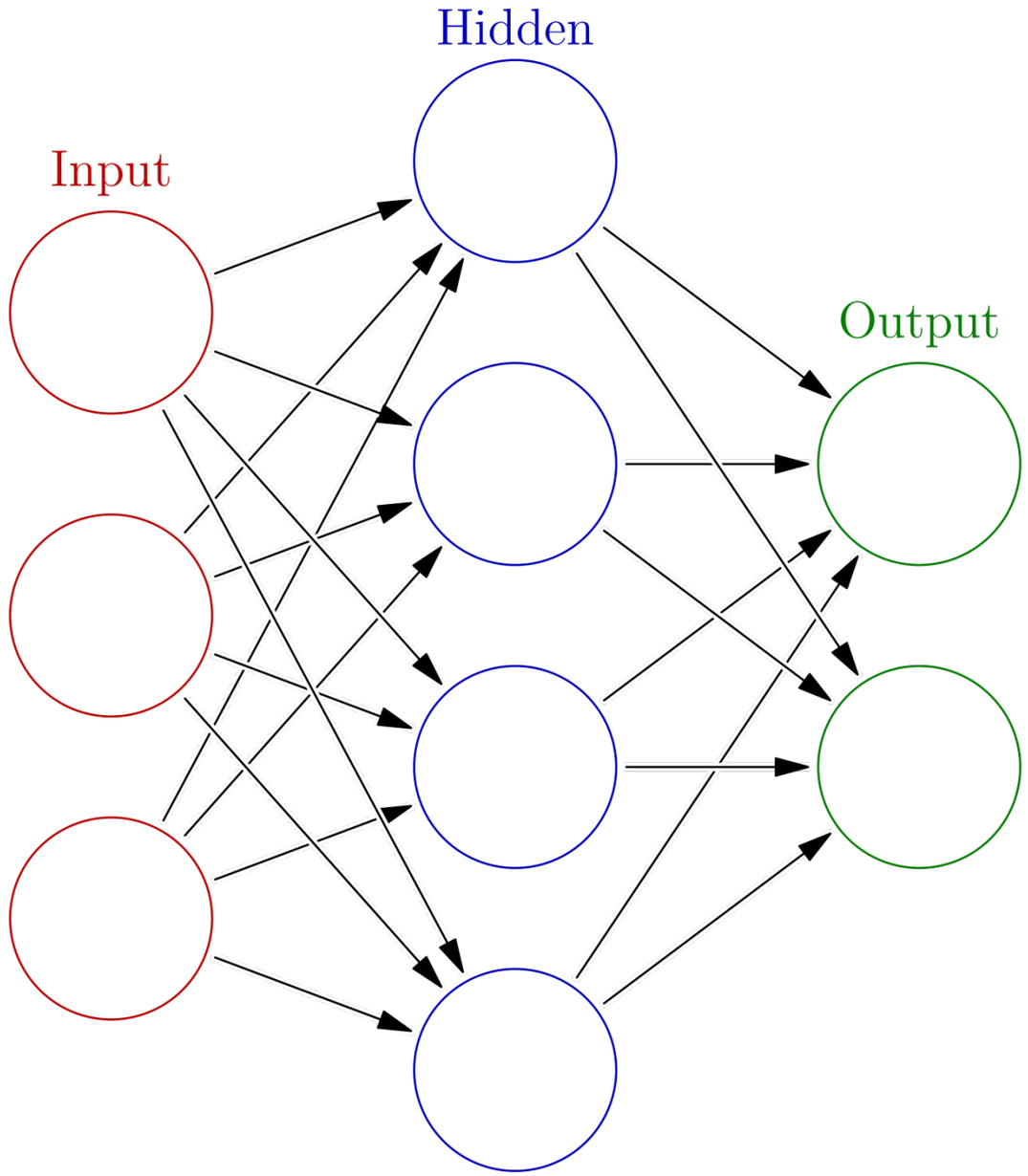


https://www.justinmath.com/the-first-wave-of-ai-reasoning-as-search/

History of AI: main achievements and setbacks

Early successes (1956 – 1974)

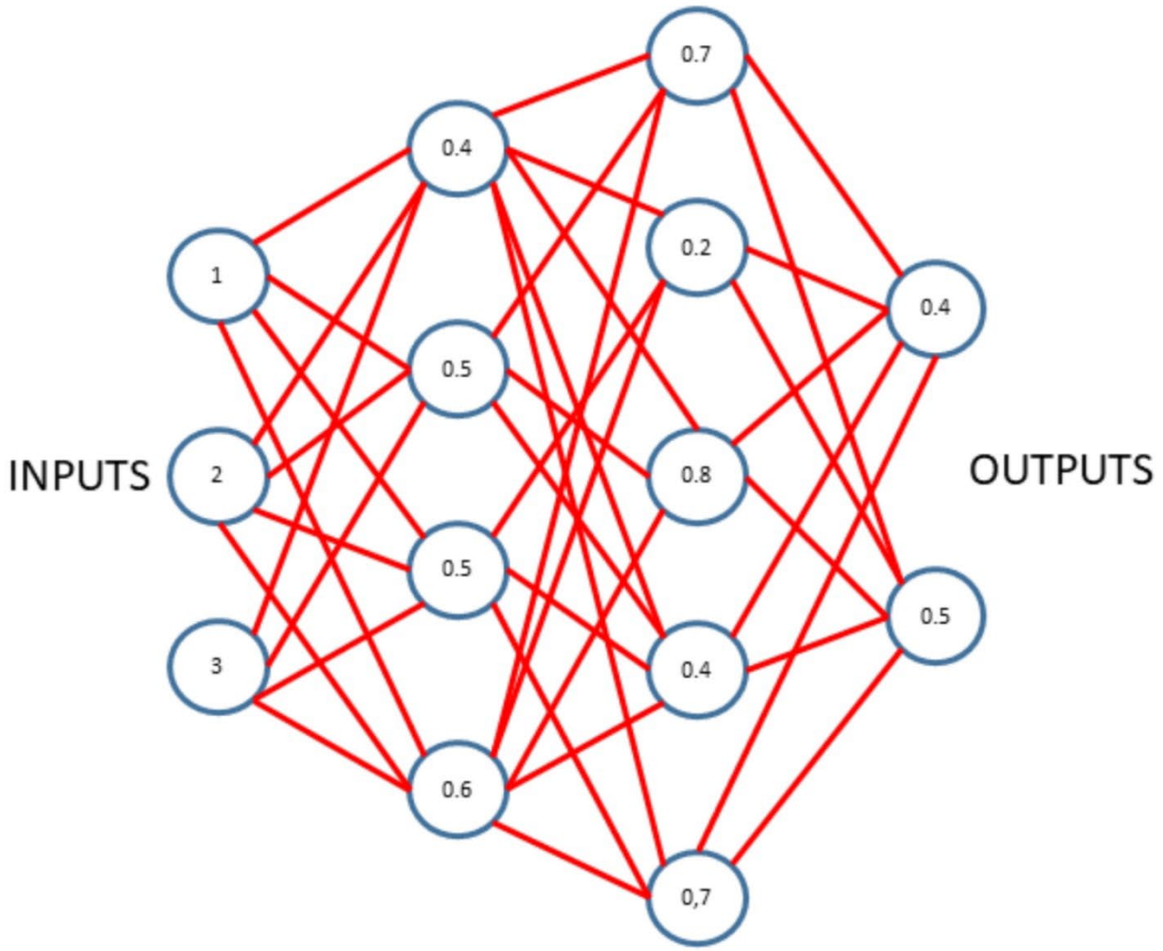
- Artificial neural networks:
  - interconnected nodes (artificial neurons), linked by edges,
  - each artificial neuron receives signals from neighbouring neurons, combines them using a non-linear function of the sum of its inputs, and sends the result (a number) to other neighbouring neurons.
  - the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



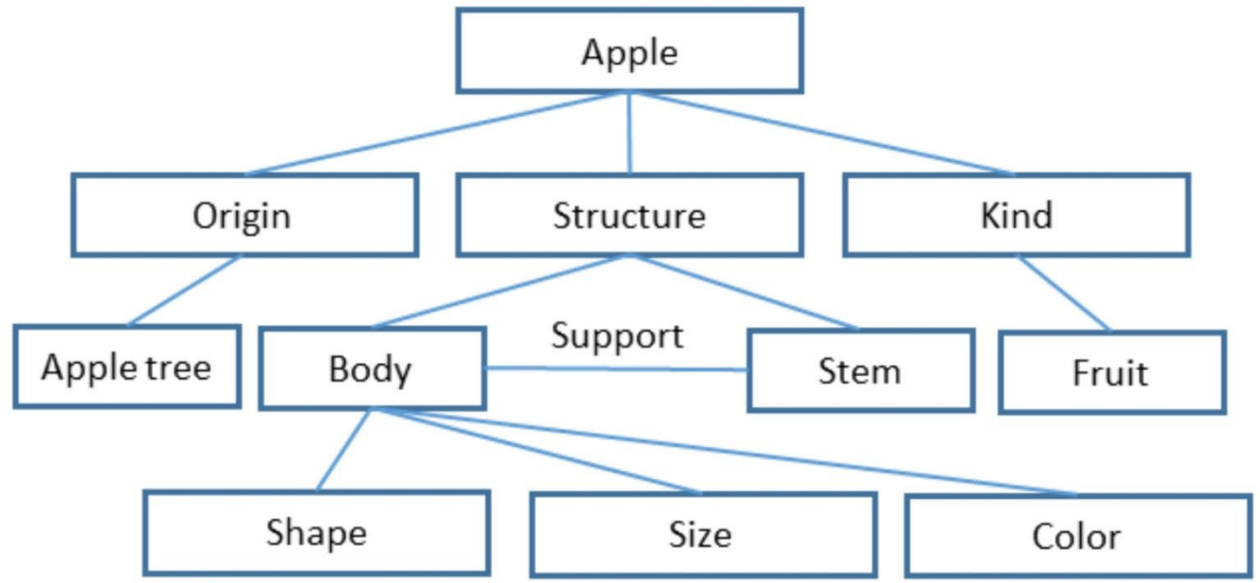
https://en.wikipedia.org/wiki/Neural_network_(machine_learning)

History of AI: main achievements and setbacks

Connectionist vs Symbolic AI approaches



(a)



(b1)

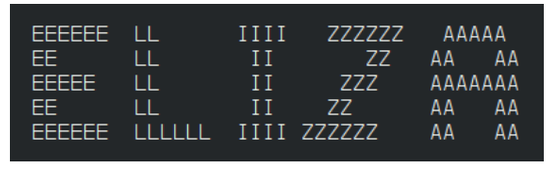
sibilings(X, Y) :- parent(Z, X) and parent)Z,Y)

(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"

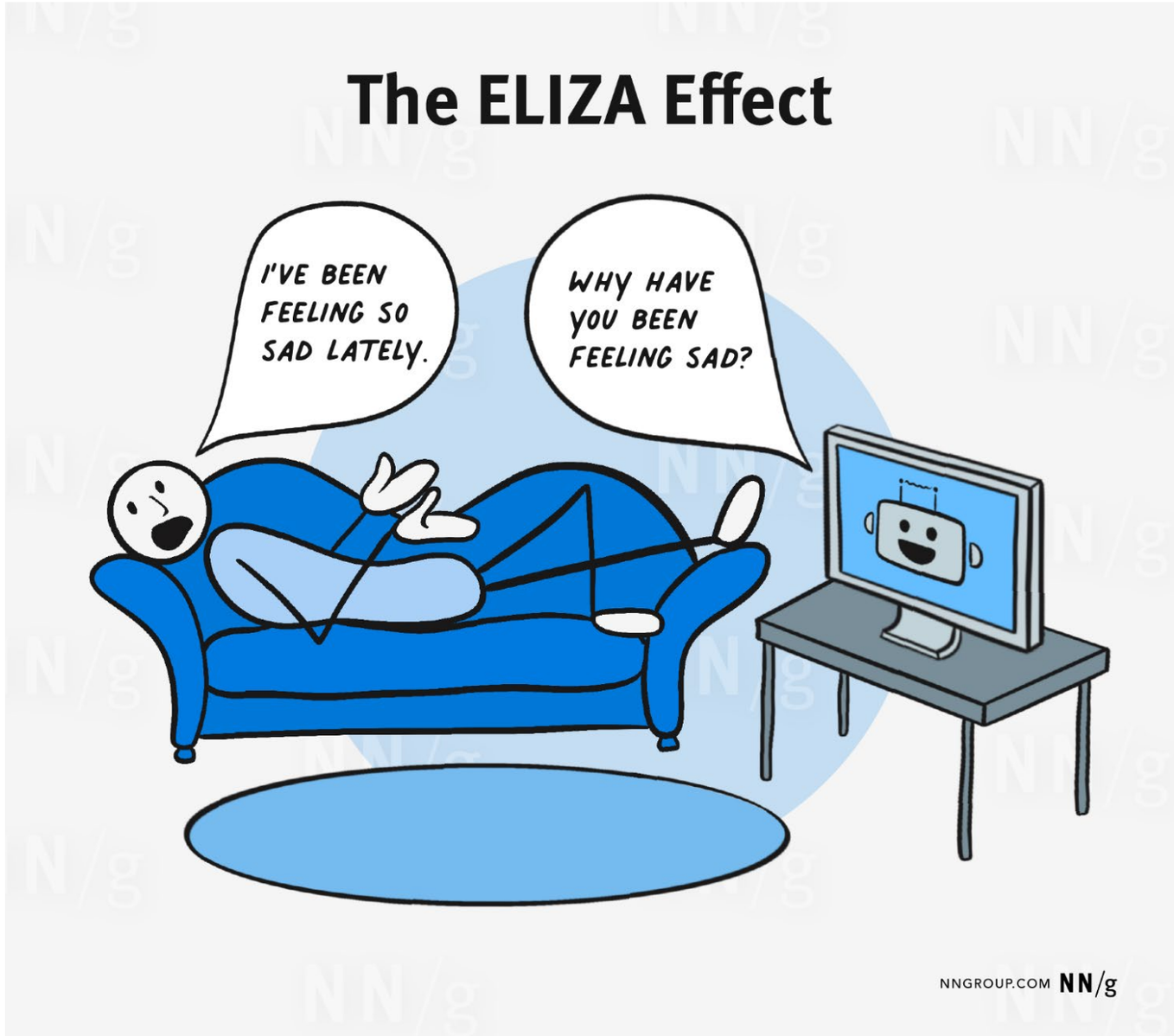
```
ELIZA: Is something troubling you ?
YOU:   Men are all alike.
ELIZA: What is the connection, do you suppose ?
YOU:   They're always bugging us about something or other.
ELIZA: Can you think of a specific example ?
YOU:   Well, my boyfriend made me come here.
ELIZA: Is it important to you that your 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/

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?

?