T-622-ARTI
Introduction to AI

Stefán Ólafsson

stefanola@ru.is

Assistant Professor
Department of Computer Science
Reykjavík University

Academic Background

- English and Chinese Studies
  - Linguistics
- Language Technologies
- Personal Health Informatics

Research Areas

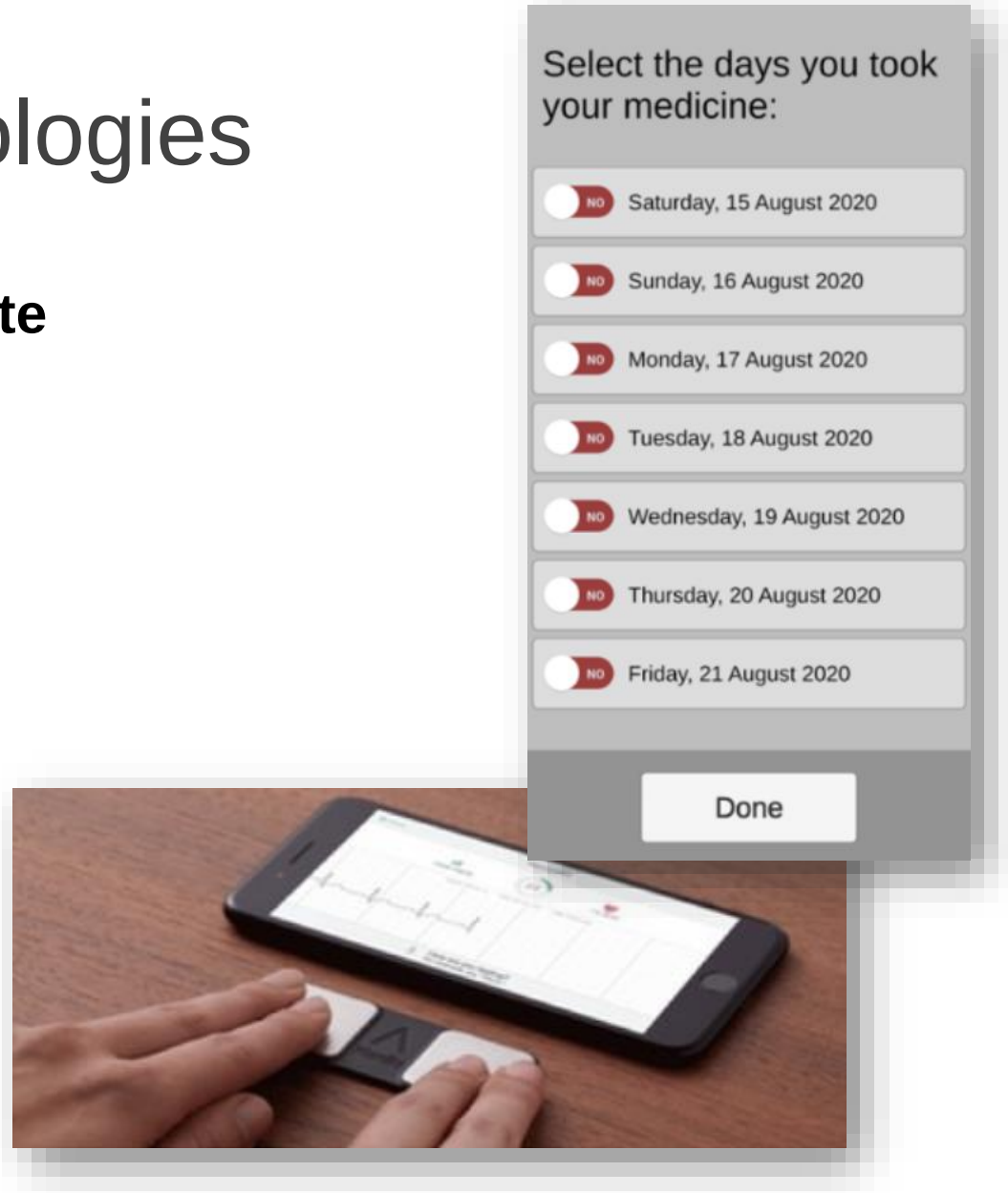
- Behavioral Health Technologies
- Virtual Agents
- Natural Language Processing

RELATIONAL AGENTS GROUP

Northeastern University

Behavioral Health Technologies

- **Using digital technologies to promote healthy behaviors:**
  - mHealth – eHealth
- **Combines:**
  - Theories of health behavior change
  - Psychology
  - Human-computer interaction
- **RQs:**
  - Effectiveness?
  - Most important features?



Behavioral Health Technologies

- Digital intervention for patients with substance use disorders in outpatient treatment
  - Collaboration with SÁÁ
  - Mobile app:
    - Treatment content, day planner, self-reflection
- Interfaces for clinical diagnosis and triage using artificial intelligence
  - Language and Voice Lab
  - Heilsugæsla höfuðborgarsvæðisins
- HuguR: Application for promoting student mental health
  - Psychology and computer science departments at RU

SAMTÖK ÁHUGAFÓLKS UM ÁFENGIS- OG VÍMUEFNAVANDANN

HEILSUGÆSLA HÖFUÐBORGARSVÆÐISINS

Virtual Agents

- **Virtual humans capable of carrying out conversations with people:**
  - Show nonverbal behavior along with speech
- **Application areas:**
  - Education
  - Health care
  - Counseling
- **RQs:**
  - Effectiveness?
  - Naturalness, trust, relationship, etc.?

Natural Language Processing

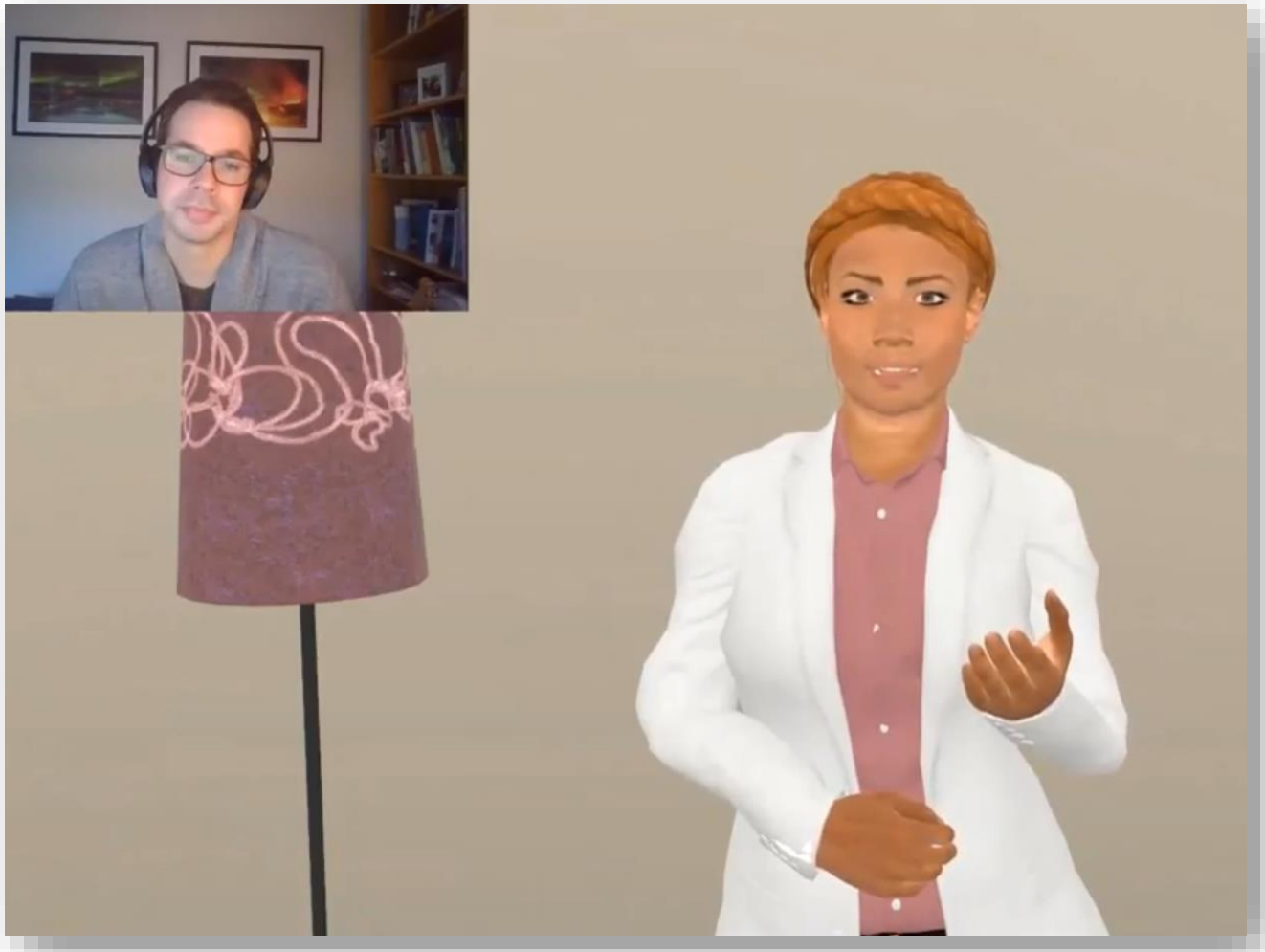
- Icelandic NLP
- Dialog systems
- Conversational agents
- **Projects:**
  - Automatically generating nonverbal behavior for Icelandic virtual humans
  - Language generation for Icelandic using large language models (GPT)
  - Conversational agents for healthcare

Automated Virtual Counseling

- **Combines:**
  - Behavior health
  - Natural language processing
  - Virtual agents
- **Applications in:**
  - Substance use, mental health, exercise, nutrition …
- **Project:**
  - *Treatment companion*
  - Agent + behavioral health intervention

Virtual Counselor

Stefan Olafsson, Byron Wallace, Timothy Bickmore, **Towards a Computational Framework for Automating Substance Use Counseling with Virtual Agents**. AAMAS – Autonomous Agents and Multiagent Systems, 2020



T-622-ARTI Introduction to AI

- **Teacher:** Stefán Ólafsson (stefanolafs@ru.is)
- **TA:** Magnús Freyr Morthens (magnusfm19@ru.is)
- **Classes:**
  - Tuesdays 12:40 – 14:15
  - Fridays 08:30 – 10:05
  - Wednesday / Thursday 10:10 – 11:45 (Labs)

Topics

- Agents and Architecture (chapter 2)
- Problem Solving / Search (chapters 3-6)
- Logic and Reasoning (chapters 7-9)
- Planning (chapter 10-11)
- Uncertainty / Bayesian Networks (chapter 13-14)
- Learning (chapters 18-21)
- Robotics (chapter 25)

Approach

- Lectures
  - Introduce theory
- Labs (Wednesdays/Thursdays)
  - Hands-on practice and problem solving (bring your laptops!); Get help with the assignments!
- Homework assignments / Quizzes (individual)
  - Test understanding of content of the lectures
- Lab assignments (groups of up to 3 students)
  - Practical applications of knowledge
- Projects
  - Build somewhat complete AI systems
  - Work in teams of 2-4 students

Composition of Final Grade

- Homework assignments / Labs / Quizzes 20%
- Projects 2 x 20%
- Final Written Exam 40%

- Grades are reduced by 1 for each day handed in late (no more than 3 days late).

- Some homework assignments + labs will have bonus points. Bonus points will count towards points missed in other assignments.

If you have questions ...

- **Piazza:** https://piazza.com/class/lam9uc07us22hi
- **Discord:** https://discord.gg/pAGmAKPr
- **In class!**
- **Email:** stefanola@ru.is
  magnusfm19@ru.is

Introduction AI

Russell and Norvig:
Chapter 1

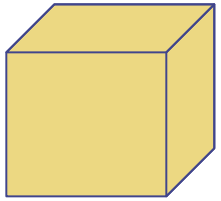
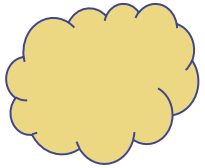
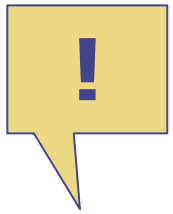
What is AI?

Chapter 1.1

What is AI?

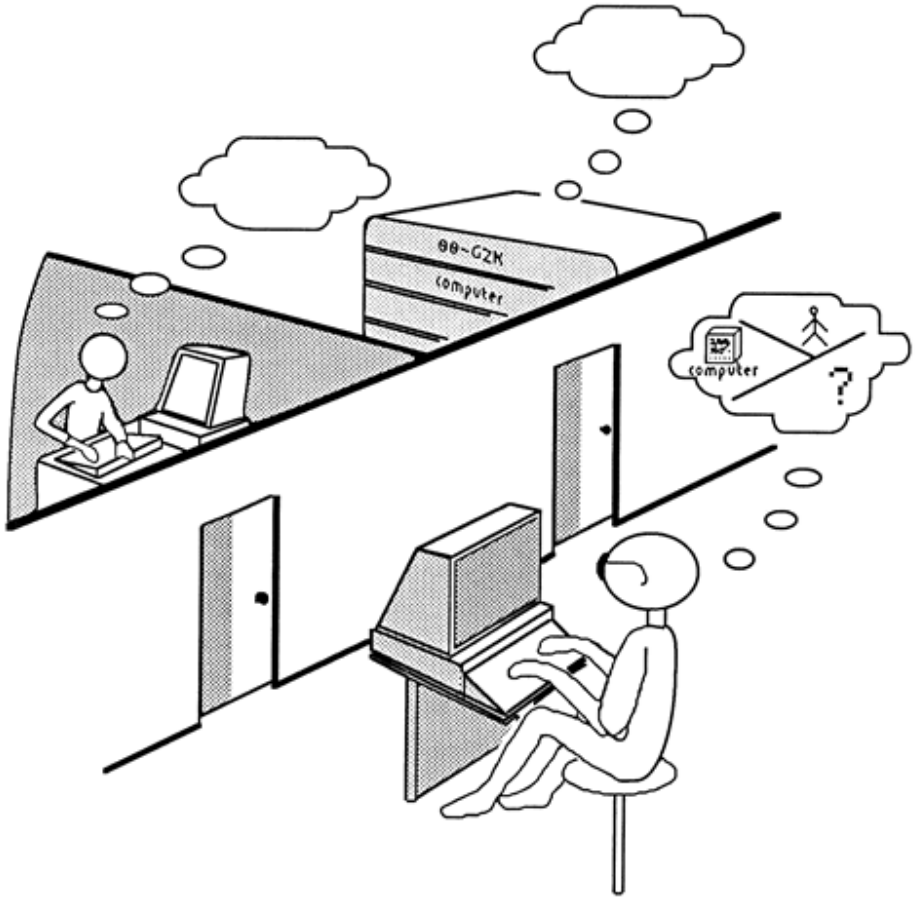
Understanding and building intelligent entities – machines that can compute how to act effectively and safely in a wide variety of novel situations.

What is AI?

Empirical Science	Math / Engineering
	
 THINK like HUMANS	THINK RATIONALLY
 ACT like HUMANS	ACT RATIONALLY

Acting Humanly

- The Turing Test
  - Proposed by Alan Turing (1950)
  - Establishes human action as the benchmark
  - AI passes test if written interrogation by human does not unveil it as a computer
  - Provides plenty to work on!
- Natural Language Processing
- Knowledge Representation
- Automated Reasoning
- Machine Learning



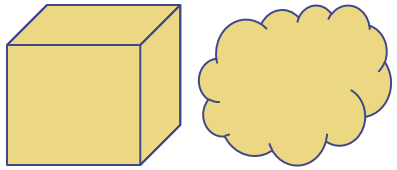
Acting Humanly

- The Turing Test (cont.)
  - The “Physical” test has also been proposed
  - Involes even more fields including
    - Computer Vision
    - Robotics
- Seems to cover most of AI!
- BUT! Does it help us to build intelligence?
- Human flight came with study of aerodynamics, not by imitating birds.

Thinking Humanly

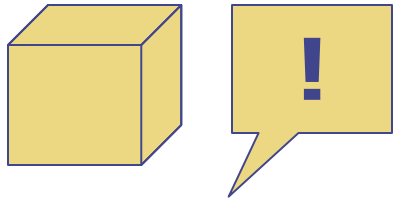
- Understanding the inner working of the human mind through psychological experiments leading to
  - Precise and testable theories
  - Computational models
- This is the field of Cognitive Science
- Cognitive Science and AI fertilize each other, e.g.,
- computer vision is partly based on human vision

Thinking Rationally



- What is “right thinking”?
- The greeks tried to answer this with laws of thought
  - Initiated the field of logic
- Logicist AI tries to describe all kinds of things and problems with a precise logical notation and use that to find “right solutions”
- Problems: (A) Incomplete information
  (B) Complexity of the algorithms

Acting Rationally



- Rational Agents try to achieve the best (expected) outcome
- May use logic inference, but ALSO other approaches to rational behavior
  - E.g. reflexes can produce rational reaction
- Here we choose the Rational Agent perspective because
  - More general than pure logic inference
  - Better defined than human rationality

Foundations of AI

Chapter 1.2

Philosophy

- Aristotle (384-322 BC)
  - Generating conclusions mechanically given a premise
- Hobbes (1588-1679)
  - Reasoning like numerical computation
- Pascal (1623-1662)
  - Numerical calculating machine – “like thought!”
- Leibniz (1646-1716)
  - Machine operating on concepts, not numbers

Philosophy

- So the mind is a machine?
- What about free will?
  - Rocks governed by physics don‘t „decide“ to fall!
- Explained in terms of “the non-physical side”
  - Dualism: part of the human mind is is excempt from physical laws (= soul/spirit)
- Explained in terms of a natural choice process
  - Materialism: free will is just the way in which we perceive our brain's operation

Philosophy

- The mind manipulates knowledge
- Where does the knowledge come from?
- It all starts at the senses, so perception is key!
- Finally, the mind controls actions
- Aristotle proposed a planning algorithm based on goals and the knowledge of action outcomes

Mathematics

- Logic: Boolean logic (Boole, 1847)
- Logic: First-order logic (Frege, 1879)
- Computation: Computability/Decidability
- Computation: Intractability (1960s)
  - Computation time grows exponentially with instance size
- Computation: NP-completeness (Cook, 1971)
  - We can identify the really hard problems
- Probability: (Cardano, 1501-1576)
  - Incorporate new evidence (Bayes, 1702-1761)

Economics

- Rationality leading to preferred outcomes or utility (Walras, 1834-1910)
- Decision Theory
  - Combines Probability Theory and Utility Theory
  - How to make decisions under uncertainty
- Game Theory
  - Decision Theory with other rational agents in the environment
- Operations Research
  - Sequence of decisions without immediate payoffs
  - Used for complex management decisions

Neuroscience

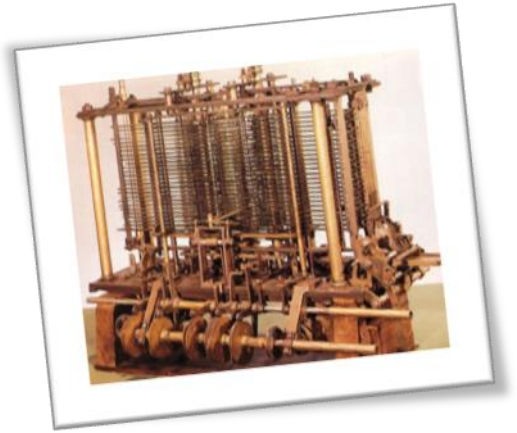
- The brain seems to “cause minds”!
- Collection of simple cells leads to thought, action and consciousness – exactly how is still mystery
- Areas of the brain seem to map to cognitive functions or body parts, yet this can change
- There are 10^{11} neurons in the brain,
  CPUs/GPUs have ~10^{10} gates
- But in the brain, all units are active simultaneously!

Psychology
How do humans think and act?

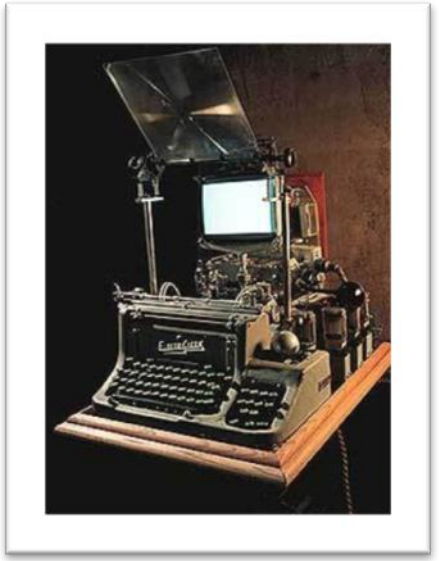
- Introspection (Wundt, 1832-1920)
- Behaviorism (Watson, 1878-1958)
  - no introspection: We can only study the stimulus and response.
  - beliefs and goals are “folk psychology”
- Cognitive Psychology (James, 1842-1910)
  - The brain as an information-processing device
  - Beliefs and goals just as scientific as pressure of a gas (Craik, 1943)
- Cognitive Science (MIT Workshop, 1956)
  - Computer models addressing psychology

Computer Engineering

- Punchcard Loom (Jacquard, 1805)
  - Programmable machine
- Difference Engine (Babbage, 1792-1871)
  - Math tables for engineering (not built, but works)
- Analytical Engine (also Babbage)
  - Universal computation
    - memory, programs, jumps
  - Ada Lovelace wrote programs for it
  - Never built: What if?

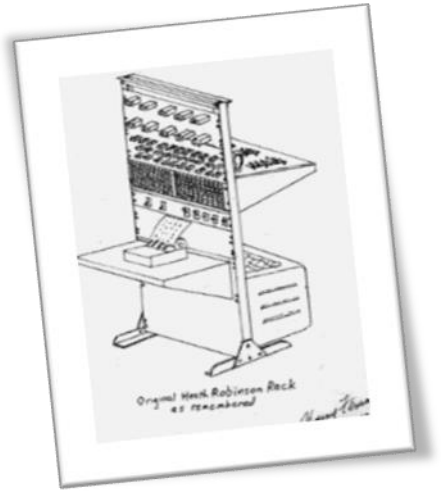


Analytical Engine Part

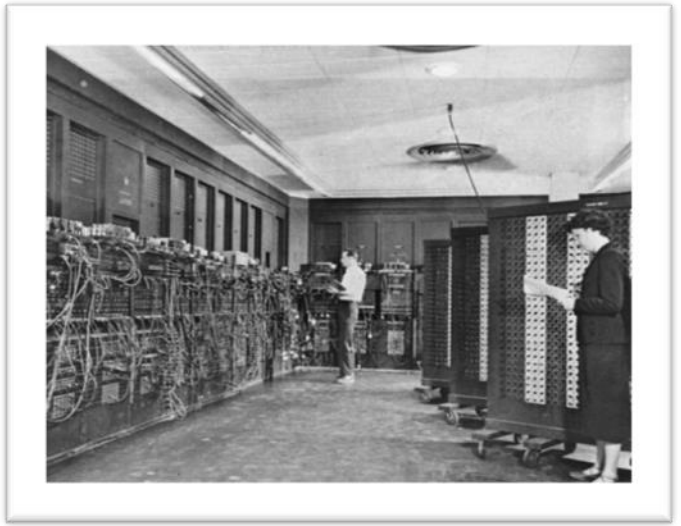


Steampunk Fiction

Computer Engineering



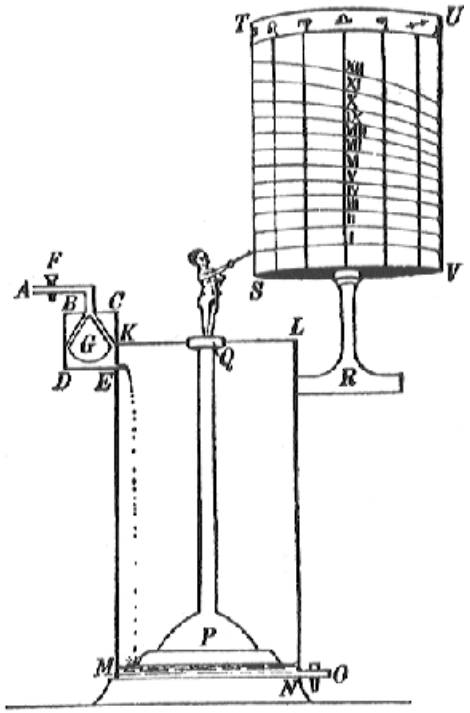
- Heath Robinson (Turing, 1940)
  - Designed to decypher German messages
- Colossus (Turing, 1943)
  - General purpose machine based on vacuum tubes
- Z-3 (Zuse, 1941)
  - Programmable
- ABC (Atanasoff, 1942)
  - First electronic computer
- ENIAC (1946)



Control Theory and Cybernetics

How can artifacts operate under their own control?

- First, only living things could modify behavior in response to changes in environment!
  - Water Clock (Ktesibios, 250 BC)
  - Kept water running at constant pace
- Thermostat (Drebbel, 1572-1633)
- Steam Engine Governor (Watt, 1736-1819)
- Control Theory and Cybernetics
  - Wiener (1894-1964) looking at control and cognition
  - Purposive behavior = mental mechanism trying to minimize the difference between current state and goal state



Control Theory and Cybernetics

- Modern Control Theory, especially stochastic optimal control tries to maximize an objective function over time
- Optimal behavior, like the rational agents

- Why not the same field as AI?
- AI breaks out of the math of control theory and considers “softer” things like language, vision and planning

Linguistics
How does language relate to thought?

- Behaviorist theory does not address creativity in language (e.g., children make new sentences from known words)
- Chomsky (1957) explains this creativity with syntactic structures, going back to Panini (350 BC), formal enough for programming
- Computational Linguistics
  - Understanding language is not just syntactic structure but also understanding the subject
  - Therefore connected with Knowledge Representation

History of AI

Chapter 1.3

Artificial Neuron (1943)

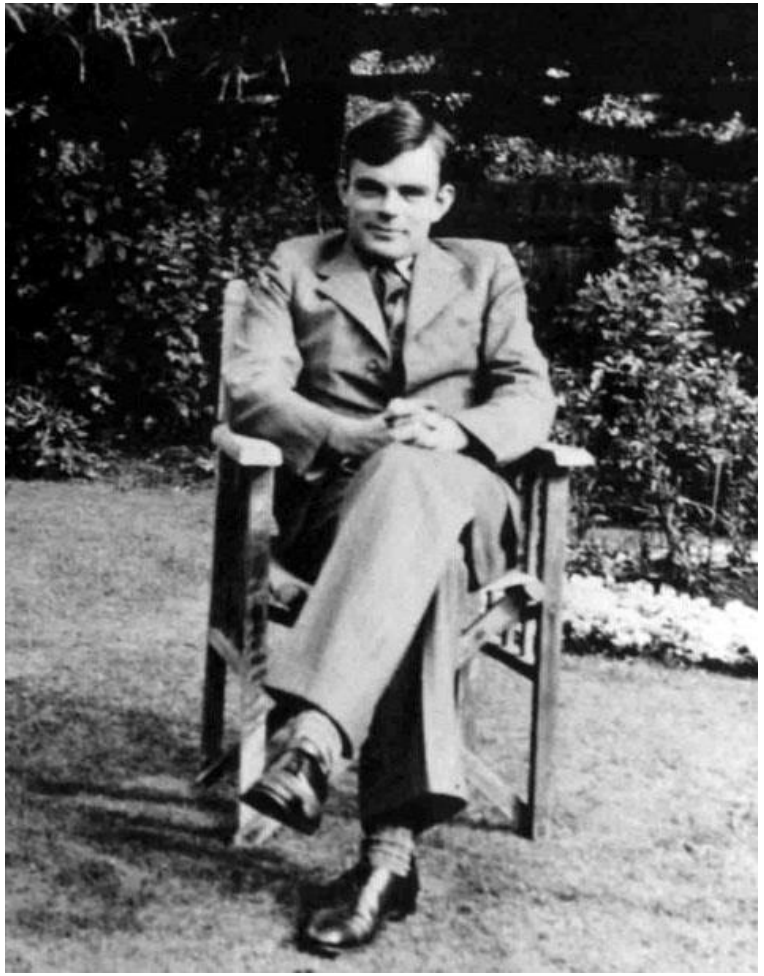
- Warren McCulloch and Walter Pitts
- ON or OFF, depending on enough stimulation by neighboring neurons
- All logical connectives (AND, OR, NOT) could be implemented by simple nets
- Suggested that these could also be made to “learn”

Neural Network Computer (1950)

- Marvin Minsky and Dean Edmonds
- 3000 vacuum tubes simulated 40 neurons

“Complete Vision of AI” (1950)

- Alan Turing articulated this vision in an article called “Computing Machinery and Intelligence”
- That’s where he proposed the Turing Test as well as machine learning and genetic algorithms.



Darthmouth Workshop (1956)

- 10 people brought together, who all had shared interest in automata theory, neural nets and study of intelligence
- Newell and Simon showed the Logic Theorist reasoning program
- McCarthy's term "Artificial Intelligence" was adopted for the field.

AI@50

50s: Exciting Early Years

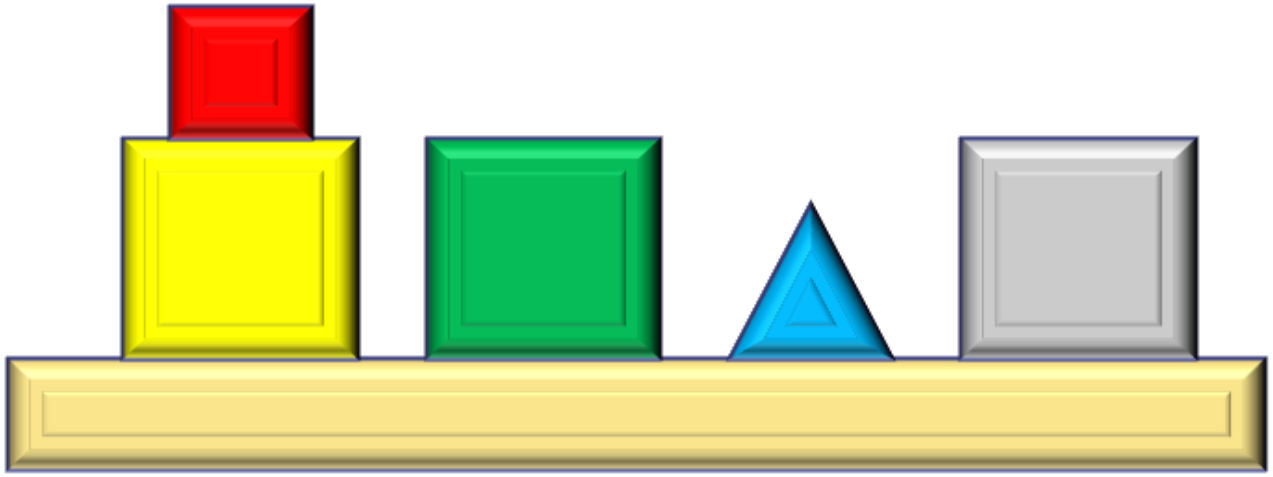
- General Problem Solver (Newell and Simon, 1957)
- Geometry Theorem Prover (Gelernter, 1959)
- Checkers players (Samuel, 1956)

1958: Good Year for McCarthy

- The Lisp programming language
- Time sharing (multiple users on a computer)
- Describes hypothetical “Advise Taker”
  - General knowledge representation and reasoning

60s: Minsky and students flourish

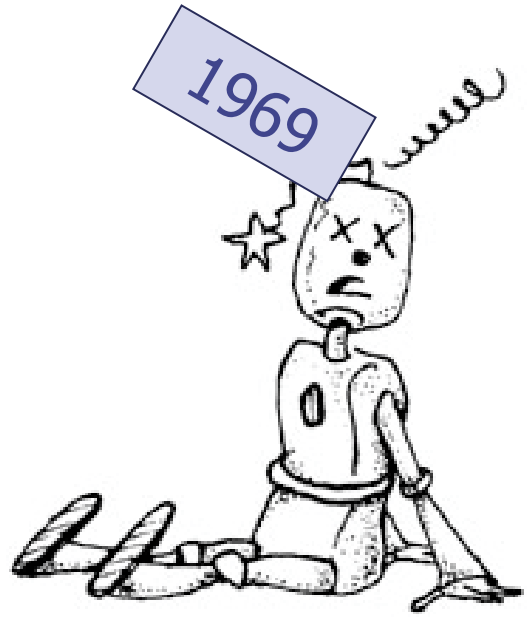
- Chose series of limited problems that appeared to require intelligence to solve: Microworlds
- Most famous is the blocks world



But the REAL world is tough!

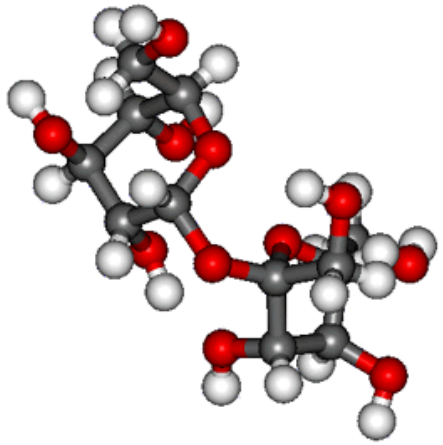
- The AI lacked general knowledge
  - Russian to English translation programs failed
  - Simple syntactic manipulation not enough
- The AI methods didn’t scale up
  - Intractable problems out of reach
  - The world is BIG
- Doubts about capabilities of neural nets
  - Could represent less than first expected

Weak AI



70s: Expert Systems Save the Day

- DENDRAL (Buchanan, 1969) - Chemistry
  - Inferring molecular structure from mass spectrum
  - Intractable to check all structures
  - Instead: Checks structure patterns known by human experts



- MYCIN (Feigenbaum, 70s) – Medical
  - Diagnoses blood infections
  - 450 rules and system better than junior doctors

80s: Industrial AI Boom

- DEC XCON Expert System saved them $40 million per year in 1986!
- By 1988 DEC's AI group had 40 expert systems deployed.
- Nearly every major US corporation established their own AI group and was using or looking into expert systems.
- Extravagant promises, but failure to deliver in the end caused a new "AI Winter".

90s: AI Becomes a Science

- Gone back to existing theories to build a strong foundation – comparing methods
- Example
  - Hidden Markov Models (HMMs) based on math
  - Bayesian Networks based on neural nets
- Resurgence of formalization and specialization has lead to isolation of more “cutting edge” work like vision and robotics

Bringing it all together: Agents

- Agents provide an opportunity to work on a complete AI system, across approaches

2000-present: deep learning, big data, robotics

- Big success for machine learning due to:
  - faster computers
  - access to large amounts of data
  - advanced machine learning techniques
- Examples:
  - Watson (IBM) wins Jeopardy (2011)
  - self-driving cars
  - AlphaGo beets Lee Sedol in Go (2016)
  - ChatGPT (2022)