

VNU-HUS MAT1206E/3508: Introduction to AI

Introduction

Hoàng Anh Đức

Bộ môn Tin học, Khoa Toán-Cơ-Tin học
Đại học KHTN, ĐHQG Hà Nội
hoanganhduc@hus.edu.vn



Textbook Basis



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

AI and Society

Agents

Knowledge-Based Systems

References

These lecture slides were constructed based primarily on the contents of:

- **Wolfgang Ertel (2025).** *Introduction to Artificial Intelligence*. 3rd. Springer. DOI: 10.1007/978-3-658-43102-0

Contents



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial
Intelligence (AI)?

The History of AI

AI and Society

Agents

Knowledge-Based
Systems

References

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

AI and Society

Agents

Knowledge-Based Systems

Additional Materials



Introduction

Hoàng Anh Đức

3 Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

AI and Society

Agents

Knowledge-Based Systems

References

Prof. Ertel's Lectures at Ravensburg-Weingarten University in 2011

- https://youtu.be/katly95_mxo (Introduction)

A recent conversation (April 2025) between Johann Schumann and Prof. Ertel (retired) about “AI at the Crossroads: Symbolic Logic, Impact on Society, and the Future of Work”

- <https://youtu.be/LDWu3oXnezc>

Michael Wooldridge (Oxford AI professor; multi-agent systems pioneer): “This is not the AI we were promised”

- <https://youtu.be/CyyL0yDhr7I> (Faraday lecture, 2026)

Demis Hassabis (DeepMind co-founder; 2024 Chemistry Nobel laureate): “Accelerating scientific discovery with AI”

- <https://youtu.be/YtPaZsasmNA> (Nobel lecture, 2024)

What is Artificial Intelligence (AI)?



Introduction

Hoàng Anh Đức

Additional Materials

4 What is Artificial Intelligence (AI)?

Brain Science and Problem Solving

The Turing Test and Chatterbots

The History of AI

AI and Society

Agents

Knowledge-Based Systems

References

■ “*Intelligence*”

- What is intelligence?
- How can intelligence be measured?
- How does our brain work?
- Intelligent machine (that behaves like a person, showing intelligent behavior)?

■ “*Artificial*”

- Science fiction? Fears of intelligent cyborgs?
- Rebuild human mind?
- Philosophy, e.g. mind-body dualism?

What is Artificial Intelligence (AI)?

John McCarthy (1955)

The goal of AI is to develop machines that behave as though they were intelligent.

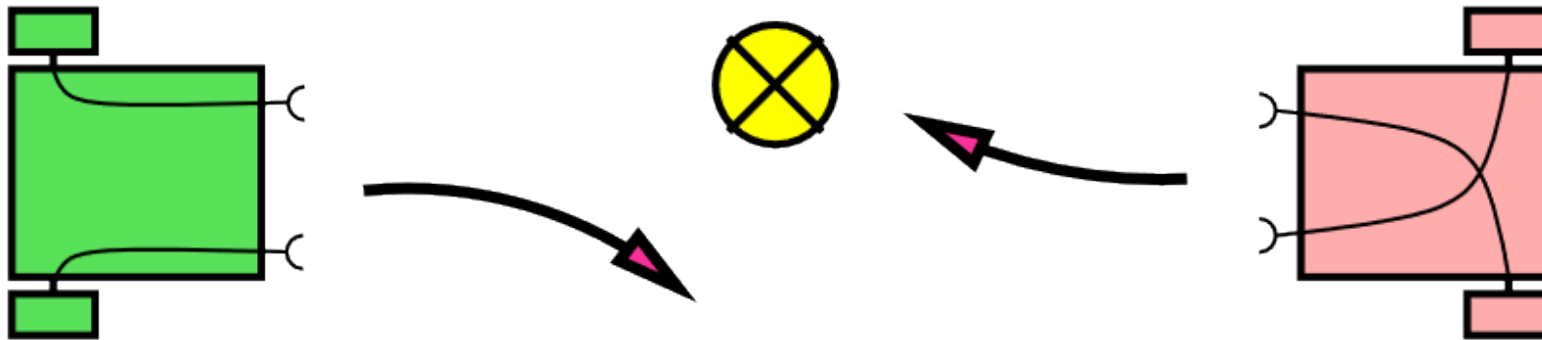


Figure: Two very simple Braitenberg vehicles and their reactions to a light source.

- Braitenberg vehicles: two wheels, each of which is driven by an independent electric motor; light sensor on the front
- The more light that hits the sensor, the faster the motor runs.
- Vehicle 1 (left) moves away from the light source, while Vehicle 2 (right) moves toward the light source.



Introduction

Hoàng Anh Đức

Additional Materials

6 What is Artificial Intelligence (AI)?

Brain Science and Problem Solving

The Turing Test and Chatterbots

The History of AI

AI and Society

Agents

Knowledge-Based Systems

References

What is Artificial Intelligence (AI)?

With some modifications of the vehicles, we can create *different “behavior patterns”*.

- Fifteen or so small robotic vehicles are moving on an enclosed four by four meter square surface.
- Some vehicles *form small groups* with relatively little movement
- Others move peacefully through the space and gracefully *avoid any collision*.
- Still others appear to *follow a leader*.

According to McCarthy’s definition, these robots can be *described as intelligent*. Yet Braitenberg showed that this *seemingly complex behavior* is produced by *very simple electrical circuits*, with *no real problem-solving ability*.

⇒ *McCarthy’s definition is insufficient* because AI has the goal of *solving difficult practical problems* which are surely too demanding for the Braitenberg vehicle.

What is Artificial Intelligence (AI)?



Introduction

Hoàng Anh Đức

Additional Materials

7

What is Artificial Intelligence (AI)?

Brain Science and Problem Solving

The Turing Test and Chatterbots

The History of AI

AI and Society

Agents

Knowledge-Based Systems

References

Encyclopedia Britannica (1991):

AI is the ability of digital computers or computer controlled robots to solve problems that are normally associated with the higher intellectual processing capabilities of humans ...

According to this definition, every computer is an AI-system.

What is Artificial Intelligence (AI)?



Introduction

Hoàng Anh Đức

Additional Materials

8

What is Artificial Intelligence (AI)?

Brain Science and Problem Solving

The Turing Test and Chatterbots

The History of AI

AI and Society

Agents

Knowledge-Based Systems

References

Elaine Rich [Rich 1983]:

Artificial Intelligence is the study of how to make computers do things at which, at the moment, people are better.

- Still up-to-date in the year 2050!
- Humans are still better in many fields (e.g. understanding pictures, learning)!
- Computers are already better in many fields (e.g. playing chess)!

What is Artificial Intelligence (AI)?



Introduction

Hoàng Anh Đức

Additional Materials

9 What is Artificial Intelligence (AI)?

Brain Science and Problem Solving

The Turing Test and Chatterbots

The History of AI

AI and Society

Agents

Knowledge-Based Systems

References

- It would be *dangerous*, however, *to conclude from Rich's definition that AI is only concerned with the pragmatic implementation of intelligent processes*.
- *Intelligent systems*, in the sense of Rich's definition, *cannot be built without a deep understanding of human reasoning and intelligent action in general*, because of which neuroscience is of great importance to AI.
- This also shows that the other cited definitions reflect important aspects of AI.
- A particular strength of human intelligence is *adaptivity*.
- We are capable of adjusting to various environmental conditions and change our behavior accordingly through *learning*.
- Precisely because our learning ability is so vastly superior to that of computers, *machine learning* is, according to Rich's definition, a central subfield of AI.

What is Artificial Intelligence (AI)?

Brain Science and Problem Solving



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

10 Brain Science and Problem Solving

The Turing Test and Chatterbots

The History of AI

AI and Society

Agents

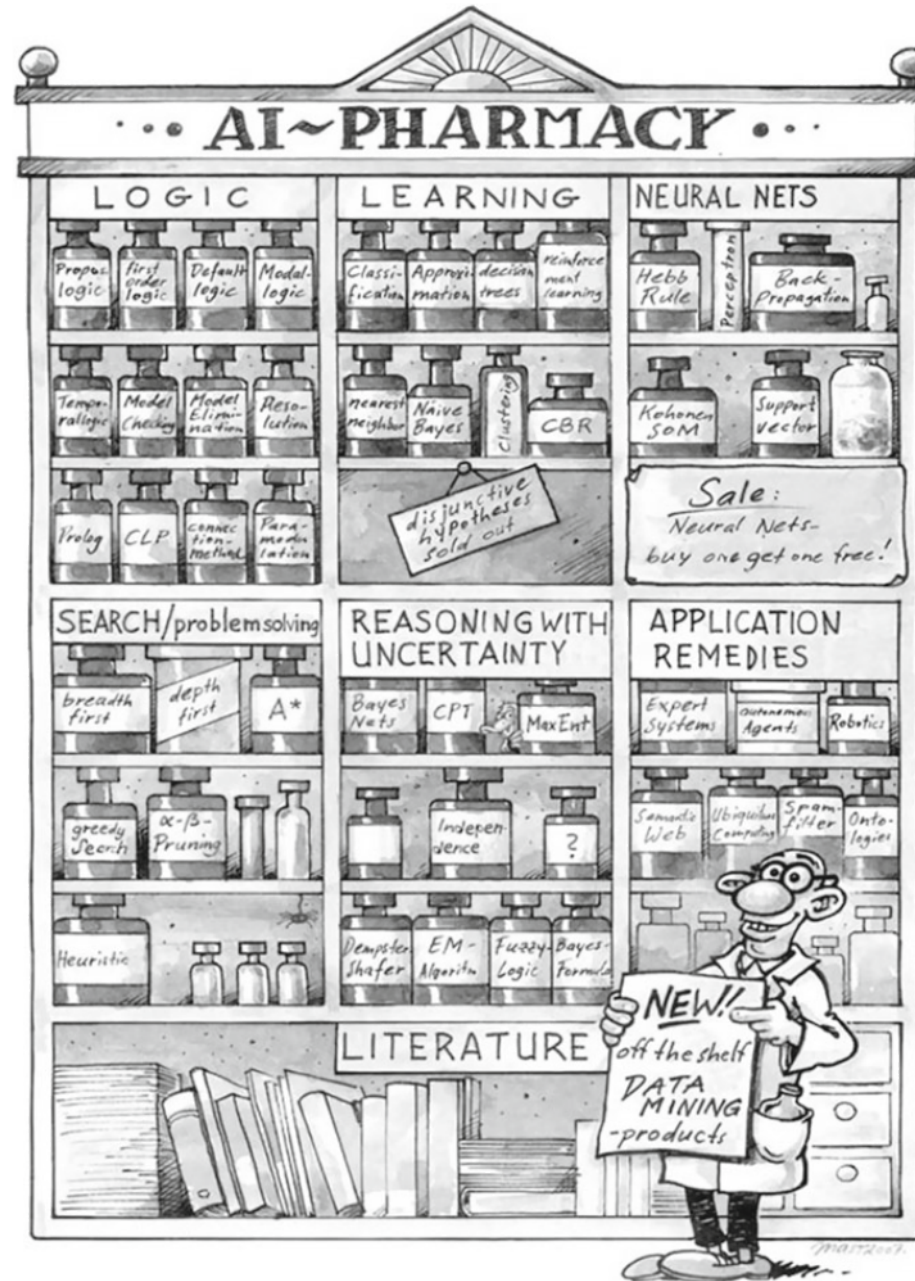
Knowledge-Based Systems

References

Different approaches:

- How does the human brain work?
- Problem oriented: building intelligent agents!
- A “general store” of many tools (no single universal method).

As in medicine, there is *no universal method for all AI applications*, but many solutions for diverse everyday problems, big and small.



What is Artificial Intelligence (AI)?

The Turing Test and Chatterbots



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

Brain Science and Problem Solving

11

The Turing Test and Chatterbots

The History of AI

AI and Society

Agents

Knowledge-Based Systems

References

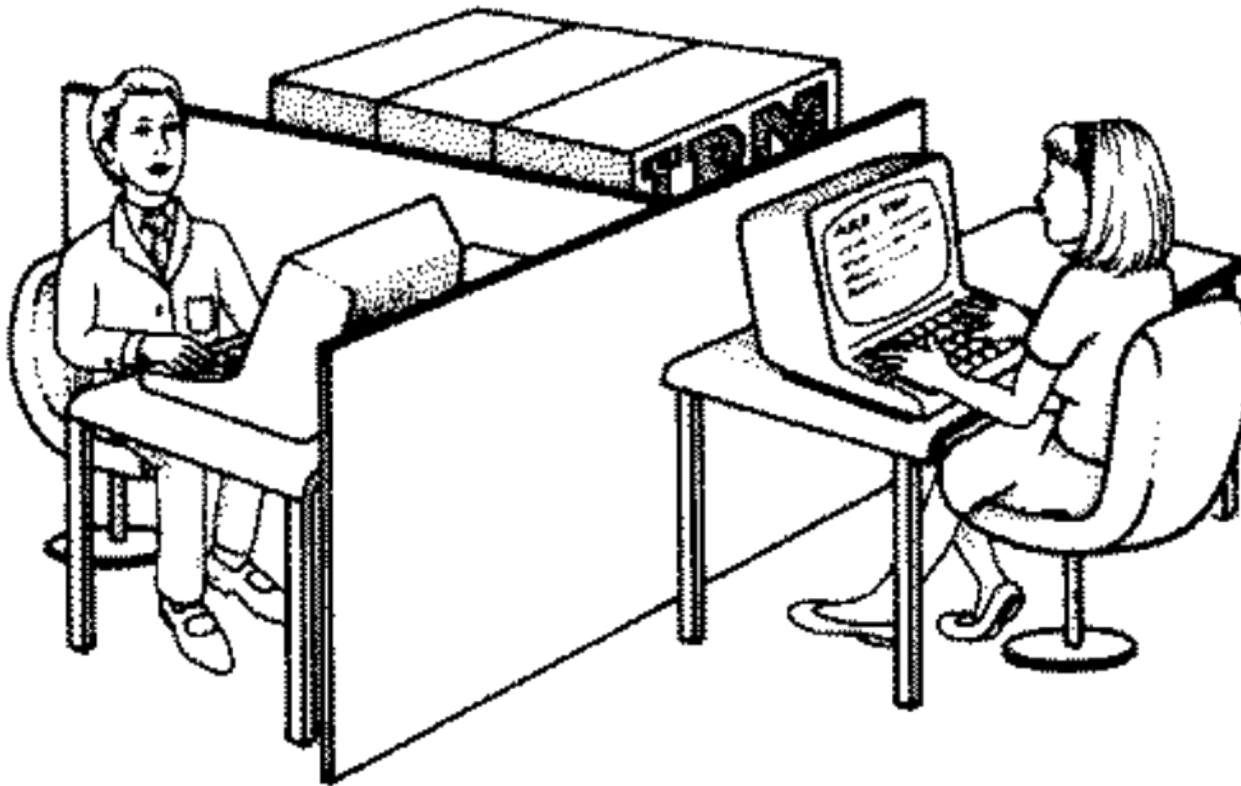


Figure: Turing test: The machine passes the test if it can mislead Alice in *at least* 30% of the cases.

While the test is very interesting philosophically, for practical AI, which deals with problem solving, it is not a very relevant test.

What is Artificial Intelligence (AI)?

The Turing Test and Chatterbots



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

Brain Science and Problem Solving

12

The Turing Test and Chatterbots

The History of AI

AI and Society

Agents

Knowledge-Based Systems

References

- Joseph Weizenbaum (*social critic*) developed *Eliza* (1966), a program meant to answer questions like a human psychologist; reportedly his secretary often had long discussions with it. He was in fact able to demonstrate success in many cases.
- Today in the internet there are many so-called *chatterbots*, some of whose initial responses are quite impressive. After a certain amount of time, however, their artificial nature becomes apparent.
- Some of these programs are actually capable of learning, while others possess extraordinary knowledge of various subjects, for example geography or software development.
- There are already commercial applications for chatterbots in online customer support and there may be others in the field of e-learning.

What is Artificial Intelligence (AI)?

The Turing Test and Chatterbots



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

Brain Science and Problem Solving

13

The Turing Test and Chatterbots

The History of AI

AI and Society

Agents

Knowledge-Based Systems

References

Exercise 1 ([Ertel 2025], Exercise 1.1, p. 23)

Test the Large Language Models (LLMs) ChatGPT

(<https://chatgpt.com>) and Gemini

(<https://gemini.google.com>) for their intelligence. Note down a starting question and measure the time you need for the different programs until you can say for sure that it is not a human. Try to assess whether the systems pass the Turing test. Look for studies that may prove that LLMs pass the Turing test.

Exercise 2 ([Ertel 2025], Exercise 1.2, p. 23)

Give reasons for the unsuitability of the Turing test as a definition of “artificial intelligence” in practical AI.

The History of AI



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

14 The History of AI

The first beginnings
Logic Solves (Almost) All Problems
The New Connectionism
Reasoning Under Uncertainty
Distributed, Autonomous and Learning Agents
AI Grows Up
The Deep Learning Breakthrough
The AI Revolution
Summary

AI and Society

Agents

Knowledge-Based Systems

References

Phases of AI history

- The first beginnings
- Logic solves (almost) all problems
- The new connectionism
- Reasoning with uncertainty
- Distributed, autonomous and learning agents
- AI has grown up
- The deep learning breakthrough
- The AI revolution

The History of AI

The first beginnings



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

15

The first beginnings

Logic Solves (Almost) All Problems

The New Connectionism

Reasoning Under Uncertainty

Distributed, Autonomous and Learning Agents

AI Grows Up

The Deep Learning Breakthrough

The AI Revolution

Summary

AI and Society

Agents

Knowledge-Based Systems

References

- In the 1930s Kurt Gödel, Alonzo Church, and Alan Turing laid important foundations for logic and TCS.
- Of particular interest for AI are Gödel's theorems
 - Every logically valid statement that can be formulated in first-order predicate logic is provable using the rules of a formal calculus.
 - In higher-order logics¹, there exist true statements that are unprovable.
- Alan Turing proved the *undecidability of the halting problem*.
 - There is no program that can decide whether a given arbitrary program (and its respective input) will run in an infinite loop.
- In the 1940s, based on results from neuroscience, McCulloch, Pitts and Hebb designed the *first mathematical models of neural networks*. However, computers at that time lacked sufficient power to simulate simple brains.

¹ Higher-order logics are extensions of predicate logic, in which not only variables, but also function symbols or predicates can appear as terms in a quantification.

The History of AI

The first beginnings



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

16

The first beginnings

Logic Solves (Almost) All Problems

The New Connectionism

Reasoning Under Uncertainty

Distributed, Autonomous and Learning Agents

AI Grows Up

The Deep Learning Breakthrough

The AI Revolution

Summary

AI and Society

Agents

Knowledge-Based Systems

References

Exercise 3 ([Ertel 2025], Exercise 1.3, p. 23)

Many well-known inference processes, learning processes, etc. are NP-complete or even undecidable. What does this mean for AI?

The History of AI

Logic Solves (Almost) All Problems



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings

17

Logic Solves (Almost) All Problems

The New Connectionism

Reasoning Under Uncertainty

Distributed, Autonomous and Learning Agents

AI Grows Up

The Deep Learning Breakthrough

The AI Revolution

Summary

AI and Society

Agents

Knowledge-Based Systems

References

- In the 1950s. Newell and Simon introduced *Logic Theorist*, *one of the first automatic theorem provers*, and thus also showed that with computers, which actually only work with numbers, one can also process symbols.
- McCarthy introduced, with the language *LISP*, *a programming language specially created for the processing of symbolic structures*.
- Dartmouth Conference: the name “Artificial Intelligence” appeared in its 1955 proposal.
- In 1965, Robinson invents the *resolution calculus* for predicate logic.

The History of AI

Logic Solves (Almost) All Problems



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings

18

Logic Solves (Almost) All Problems

The New Connectionism

Reasoning Under Uncertainty

Distributed, Autonomous and Learning Agents

AI Grows Up

The Deep Learning Breakthrough

The AI Revolution

Summary

AI and Society

Agents

Knowledge-Based Systems

References

- In the 1970s, *PROLOG* was introduced as the *European counterpart to LISP*. PROLOG offers the advantage of allowing direct programming using Horn clauses, a subset of predicate logic. Like LISP, PROLOG has data types for convenient processing of lists.
- Until well into the 1980s, the string of impressive achievements in symbol processing leads to *heavy investment into the construction of intelligent computers* (e.g., Japan's *Fifth Generation Project* and Europe's *ESPRIT* program).
- For small problems, automatic provers and other symbol-processing systems sometimes worked very well. The combinatorial explosion of the search space, however, defined a very narrow window for these successes.
- Because the economic success of AI fell short of expectations, *funding for logic-based AI in the US fell dramatically during the 1980s*.

The History of AI

The New Connectionism



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings

Logic Solves (Almost) All Problems

19 The New Connectionism

Reasoning Under Uncertainty

Distributed, Autonomous and Learning Agents

AI Grows Up

The Deep Learning Breakthrough

The AI Revolution

Summary

AI and Society

Agents

Knowledge-Based Systems

References

- Computer scientists, physicists, and Cognitive scientists were able to show using computers which were now sufficiently powerful, that *mathematically modeled neural networks are capable of learning using training examples, to perform tasks which previously required costly programming.*
- Such systems improve *pattern recognition*. (E.g., Facial recognition in photos and handwriting recognition.)
- In 1986: the system *Nettalk* was able to learn speech from example texts.
- Limitation: *It was usually not possible to capture the learned concept in simple formulas or logical rules.* Attempts to combine neural nets with logical rules or the knowledge of human experts met with great difficulties.

The History of AI

Reasoning Under Uncertainty



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings

Logic Solves (Almost) All Problems

The New Connectionism

20

Reasoning Under Uncertainty

Distributed, Autonomous and Learning Agents

AI Grows Up

The Deep Learning Breakthrough

The AI Revolution

Summary

AI and Society

Agents

Knowledge-Based Systems

References

- Many diagnostic and expert systems have been built for problems of everyday reasoning using *Bayesian networks*.
 - The success of Bayesian networks stems from their intuitive comprehensibility, the clean semantics of conditional probability, and from the centuries-old, mathematically grounded probability theory.
- The weaknesses of logic, which can only work with two truth values, can be solved by *fuzzy logic*, which pragmatically introduces infinitely many values between zero and one.
- A much different path led to the *successful synthesis of logic and neural networks* under the name *hybrid systems*.
 - For example, neural networks were employed to learn heuristics for reduction of the huge combinatorial search space in proof discovery.

The History of AI

Reasoning Under Uncertainty



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings
Logic Solves (Almost) All Problems
The New Connectionism
Reasoning Under Uncertainty
Distributed, Autonomous and Learning Agents
AI Grows Up
The Deep Learning Breakthrough
The AI Revolution
Summary

AI and Society

Agents

Knowledge-Based Systems

References

21

- Methods of *decision tree learning from data* also work with probabilities. Systems like CART, ID3 and C4.5 can quickly and automatically build very accurate decision trees which can represent propositional logic concepts and then be used as expert systems.
- Since about 1990, *data mining* has developed as a sub-discipline of AI in the area of statistical data analysis for extraction of knowledge from large databases.
 - Data mining combines techniques from databases, statistics, and machine learning, and develops scalable algorithms for extracting patterns from large datasets.

51

The History of AI

Distributed, Autonomous and Learning Agents



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings

Logic Solves (Almost) All Problems

The New Connectionism

Reasoning Under Uncertainty

22 Distributed, Autonomous and Learning Agents

AI Grows Up

The Deep Learning Breakthrough

The AI Revolution

Summary

AI and Society

Agents

Knowledge-Based Systems

References

- One goal of *distributed artificial intelligence* is to use parallel computers to increase the efficiency of problem solvers.
- A very different conceptual approach results from the *development of autonomous software agents and robots that are meant to cooperate like human teams*.
 - There are many cases in which an individual agent is not capable of solving a problem, even with unlimited resources.
 - Only the cooperation of many agents leads to the intelligent behavior or to the solution of a problem.
- *Active skill acquisition by robots* is an exciting area of current research. There are robots today, for example, that independently learn to walk or to perform various motorskills related to soccer.

The History of AI

AI Grows Up



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings
Logic Solves (Almost) All Problems
The New Connectionism
Reasoning Under Uncertainty
Distributed, Autonomous and Learning Agents
AI Grows Up
The Deep Learning Breakthrough
The AI Revolution
Summary

AI and Society

Agents

Knowledge-Based Systems

References

23

51

- The above systems offered by AI today are *not a universal recipe*, but a *workshop* with a manageable number of *tools for very different tasks*.
- Most of these tools are well-developed and are available as finished software libraries, often with convenient user interfaces.
- The selection of the right tool and its sensible use in each individual case is left to the AI developer or knowledge engineer.
- More than nearly any other science, AI is *interdisciplinary*, for it draws upon interesting discoveries from such diverse fields as logic, operations research, statistics, control engineering, image processing, linguistics, philosophy, psychology, and neurobiology.

The History of AI

The Deep Learning Breakthrough



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings

Logic Solves (Almost) All Problems

The New Connectionism

Reasoning Under Uncertainty

Distributed, Autonomous and Learning Agents

AI Grows Up

24

The Deep Learning Breakthrough

The AI Revolution

Summary

AI and Society

Agents

Knowledge-Based Systems

References

- Research on neural networks dates back to the very beginning of AI, but the great breakthrough was long delayed: *object recognition in images* could not be solved for many years.
- This changed rapidly between *2006 and 2016*: researchers found a handful of tricks that allowed them to successfully *train backpropagation networks with more than three layers* of neurons.
- Advances in object recognition, together with progress in localization, planning, and control, helped advance autonomous systems such as service robots and self-driving vehicles.
- Deep learning also shows strong results in *medicine*: for some image-based diagnostic tasks (e.g., MRI and CT scans), carefully validated systems can match or sometimes exceed clinicians' performance.

The History of AI

The AI Revolution



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings
Logic Solves (Almost) All Problems
The New Connectionism
Reasoning Under Uncertainty
Distributed, Autonomous and Learning Agents
AI Grows Up
The Deep Learning Breakthrough

25

The AI Revolution
Summary

AI and Society

Agents

Knowledge-Based Systems

References

51

- *Deep learning networks* can for example learn to classify images with very high accuracy.
- Progress in image classification was one important driver of the *AI revolution* and contributed to advances in self-driving cars and service robots.

The History of AI

Summary



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings
Logic Solves (Almost) All Problems
The New Connectionism
Reasoning Under Uncertainty
Distributed, Autonomous and Learning Agents
AI Grows Up
The Deep Learning Breakthrough
The AI Revolution
Summary

AI and Society

Agents

Knowledge-Based Systems

References

- 1930–1931** The Austrian Kurt Gödel shows that all logically valid statements in first-order *predicate logic* are derivable. In higher-order logics, on the other hand, there are true statements that are unprovable. (Gödel showed that predicate logic extended with the axioms of arithmetic is incomplete.)
- 1937** Alan Turing points out the limits of intelligent machines with the halting problem.
- 1943** McCulloch and Pitts model *neural networks* and make the connection to propositional logic.
- 1950** Alan Turing defines machine intelligence with the *Turing test* and writes about learning machines and evolution-inspired methods for improving them.
- 1951** Marvin Minsky and Dean Edmonds develop *SNARC*, an early neural-network learning machine.
- 1955** Arthur Samuel (IBM) builds a learning checkers program that plays better than its developer.

26

51

The History of AI

Summary



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings
Logic Solves (Almost) All Problems
The New Connectionism
Reasoning Under Uncertainty
Distributed, Autonomous and Learning Agents
AI Grows Up
The Deep Learning Breakthrough
The AI Revolution

27

Summary

AI and Society

Agents

Knowledge-Based Systems

References

- 1956 McCarthy organizes a conference at Dartmouth College. The name *Artificial Intelligence* appeared in its 1955 proposal. Newell and Simon of Carnegie Mellon University (CMU) present the *Logic Theorist*, one of the first symbol-processing computer programs.
- 1958 McCarthy invents at MIT (Massachusetts Institute of Technology) the high-level language *LISP*. He writes programs that are capable of modifying themselves.
- 1959 Gelernter (IBM) builds the Geometry Theorem Prover.
- 1961 The General Problem Solver (GPS) by Newell and Simon simulates aspects of human problem solving.
- 1963 McCarthy founds the AI Lab at Stanford University.
- 1965 Robinson invents the *resolution calculus* for predicate logic.

51

The History of AI

Summary



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings

Logic Solves (Almost) All Problems

The New Connectionism

Reasoning Under Uncertainty

Distributed, Autonomous and Learning Agents

AI Grows Up

The Deep Learning Breakthrough

The AI Revolution

28

Summary

AI and Society

Agents

Knowledge-Based Systems

References

51

- 1966 Weizenbaum's program Eliza carries out dialog with people in natural language.
- 1969 Minsky and Papert show in their book *Perceptrons* that a single-layer perceptron is limited to classifications with a linear decision boundary.
- 1972 French scientist Alain Colmerauer invents the logic programming language *PROLOG*. British physician de Dombal develops an *expert system* for diagnosis of acute abdominal pain. It goes unnoticed in the mainstream AI community of the time.
- 1976 Shortliffe and Buchanan develop MYCIN, an expert system for diagnosis of infectious diseases, which is capable of dealing with uncertainty.
- 1981 Japan begins, at great expense, the "Fifth Generation Project" with the goal of building a powerful PROLOG machine.

The History of AI

Summary



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings
Logic Solves (Almost) All Problems
The New Connectionism
Reasoning Under Uncertainty
Distributed, Autonomous and Learning Agents
AI Grows Up
The Deep Learning Breakthrough
The AI Revolution
Summary

AI and Society

Agents

Knowledge-Based Systems

References

- 1982** R1, the expert system for configuring computers, saves Digital Equipment Corporation 40 million dollars per year.
- 1986** Renaissance of neural networks through, among others, Rumelhart, Hinton and Sejnowski. The system Nottalk learns to read texts aloud.
- 1990** Work by Pearl, Cheeseman, Whittaker, and Spiegelhalter helps make probabilistic reasoning, including *Bayesian networks*, influential in AI. Multi-agent systems become popular.
- 1992** Tesauro's TD-gammon program demonstrates the advantages of reinforcement learning.
- 1993** Worldwide *RoboCup* initiative to build soccer-playing autonomous robots.
- 1995** From statistical learning theory, Vapnik develops support vector machines, which are very important today.

29

51

The History of AI

Summary



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings
Logic Solves (Almost) All Problems
The New Connectionism
Reasoning Under Uncertainty
Distributed, Autonomous and Learning Agents
AI Grows Up
The Deep Learning Breakthrough
The AI Revolution

30

Summary

AI and Society

Agents

Knowledge-Based Systems

References

- 1997** IBM's chess computer Deep Blue defeats the chess world champion Garry Kasparov. First international RoboCup competition in Japan.
- 2003** The robots in RoboCup demonstrate impressively what AI and robotics are capable of achieving.
- 2006** Service robotics becomes a major AI research area.
- 2009** First Google self-driving car drives on the California freeway.
- 2010** Autonomous robots use learning to improve their behavior.
- 2011** IBM's "Watson" beats two human champions on the television game show "Jeopardy!". Watson understands natural language and can answer difficult questions very quickly.

51

The History of AI

Summary



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings
Logic Solves (Almost) All Problems
The New Connectionism
Reasoning Under Uncertainty
Distributed, Autonomous and Learning Agents
AI Grows Up
The Deep Learning Breakthrough
The AI Revolution
Summary

AI and Society

Agents

Knowledge-Based Systems

References

- 2014** Generative Adversarial Networks (*GANs*) are introduced.
- 2015** Daimler premieres the first autonomous truck on the Autobahn.
Google self-driving cars have driven over one million miles and operate within cities.
Deep learning enables very good image classification.
Neural style transfer can generate images in the styles of famous paintings.
- 2016** GAN-based systems generate images from text descriptions and short videos.
- 2015–16** The Go program AlphaGo by Google DeepMind defeats European champion Fan Hui 5–0 in October 2015 and Lee Sedol, one of the world's best Go players, 4–1 in March 2016. Deep learning techniques applied to pattern recognition, as well as reinforcement learning and Monte Carlo tree search lead to this success.

31

51

The History of AI

Summary



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

The first beginnings
Logic Solves (Almost) All Problems
The New Connectionism
Reasoning Under Uncertainty
Distributed, Autonomous and Learning Agents
AI Grows Up
The Deep Learning Breakthrough
The AI Revolution

32

Summary

AI and Society

Agents

Knowledge-Based Systems

References

- 2017–18** AlphaGo Zero learns Go solely through self-play, without human game data. AlphaZero generalizes the approach to separately trained systems for chess, shogi, and Go.
- 2020** OpenAI's GPT-3, trained on web text, Wikipedia, and many books, can hold impressive academic-style conversations and would probably pass simple, non-adversarial Turing-test settings.

51

The History of AI

Summary



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

- The first beginnings
- Logic Solves (Almost) All Problems
- The New Connectionism
- Reasoning Under Uncertainty
- Distributed, Autonomous and Learning Agents
- AI Grows Up
- The Deep Learning Breakthrough
- The AI Revolution

33

Summary

AI and Society

Agents

Knowledge-Based Systems

References

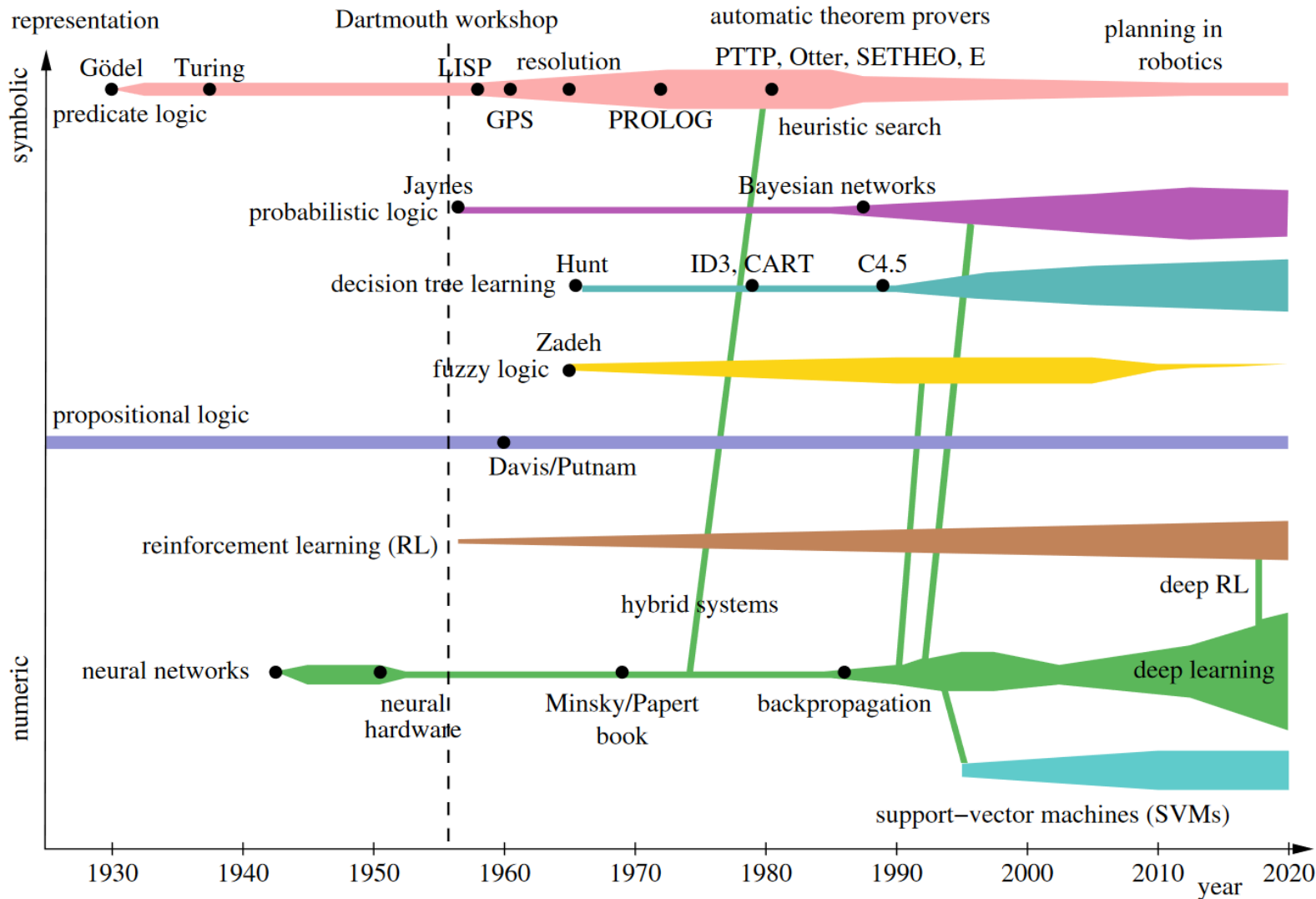


Figure: History of the various AI areas. The width of the bars indicates prevalence of the method's use.

51

AI and Society

Does AI destroy jobs?



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

34 AI and Society

Agents

Knowledge-Based Systems

References

- In January 2016, the *World Economic Forum* projected a net loss of about *5.1 million jobs* across 15 major economies from 2015 to 2020: 7.1 million jobs lost and about 2 million gained, due to the Fourth Industrial Revolution and other socio-economic, geopolitical, and demographic changes.
- AI has been one of the most important drivers of this automation trend since about 2010.

The social impact of AI is therefore a question every future AI developer must engage with.

AI and Society



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

35 AI and Society

Agents

Knowledge-Based Systems

References

The purpose of AI

- AI saves energy
- AI saves time
- AI saves money
- AI increases productivity



The road to paradise?

- Many unpleasant (physically hard, dirty, unhealthy) jobs can be done faster, more precisely, and above all cheaper by machines.
- Thus automation is a complete blessing for humanity, assuming it does not result in negative side effects, such as harm to the environment.
- We have to do less and less unpleasant work and in turn have more time for the good things in life.



Not necessarily!

- Some workers still face long hours, stress, and burnout, and in some OECD countries real labour income has not kept pace with productivity growth.
- One explanation is that competitive pressure can push companies to lower production costs and reduce employment.
- To maintain sales, more products must be manufactured and sold, fueling the need for continuous economic growth.
- In a country in which the population is no longer growing, this means each citizen must consume more, driving the creation of new markets and aggressive marketing.
- The growth/consumption spiral is seen as necessary for prosperity but leads to negative consequences.
 - A strong focus on wealth and possessions is associated with lower well-being; experiments also find that materialistic cues can reduce well-being.
 - Bad effects on our living conditions. (We are over-exploiting nature's nonrenewable resources.)

AI and Society



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

38 AI and Society

Agents

Knowledge-Based Systems

References

Where do the profits go?

- Not to the workers!
- Part of the profits is spent on investment and further growth.
- The rest is taken by the capital owners.

What is missing is a fair and just distribution of profits. How can this be achieved?

- Change the rules and laws of the economy so that all people profit from increased productivity.



AI in transportation: Autonomous vehicles, such as self-driving cars and robotaxis; electric self-driving cars are one example.

■ Pros:

- Comfort and convenience.
- Reducing in the number of cars and places for parking.
- More environmentally friendly than current vehicles.
- Increase safety.

■ Cons:

- Taxi-driving jobs may be displaced or transformed.
- Rebound effect: Shorter driving time + comfortable, cheap driving $\Rightarrow$ Driving more.
- Security. (Hackers can manipulate the vehicles' controls through security holes in their network protocols.)



Service robots are being developed for household and assistance tasks. The assistance robot *Marvin* is one research example; Boston Dynamics' *Atlas* is an industrial humanoid robot.

■ Pros:

- A large gain in comfort (e.g., a robot that cleans and tidies without complaint).
- Assistance robots could support the *elderly* and people with disabilities — increasingly important in the coming demographic shift.

■ Cons:

- Potential trade-offs: service robots may create additional demand for natural resources and energy.
- General-purpose humanoid robots could be misused, including for military purposes.

⇒ We should engage in social discourse before such robots reach the market.

Agents



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

AI and Society

41 Agents

Knowledge-Based Systems

References

Agent denotes rather generally *a system that processes information and produces an output from an input.*

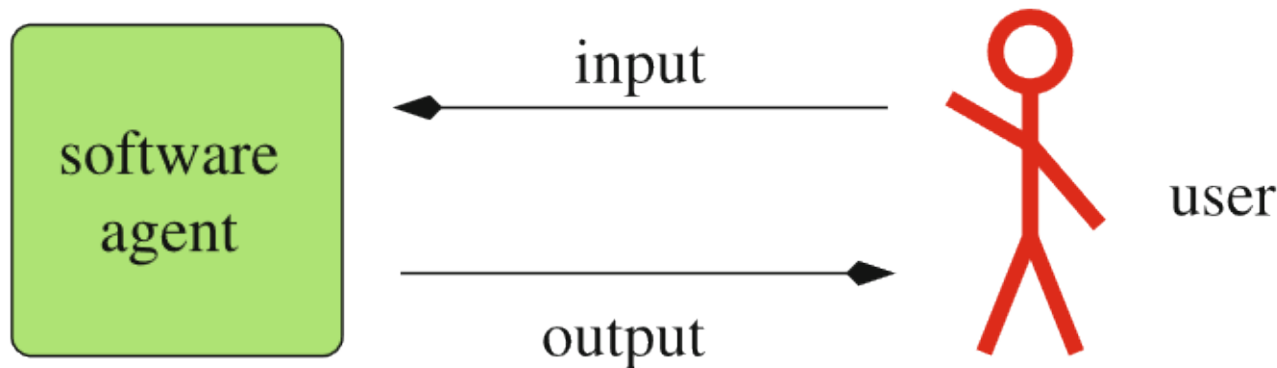


Figure: A software agent with user interaction.

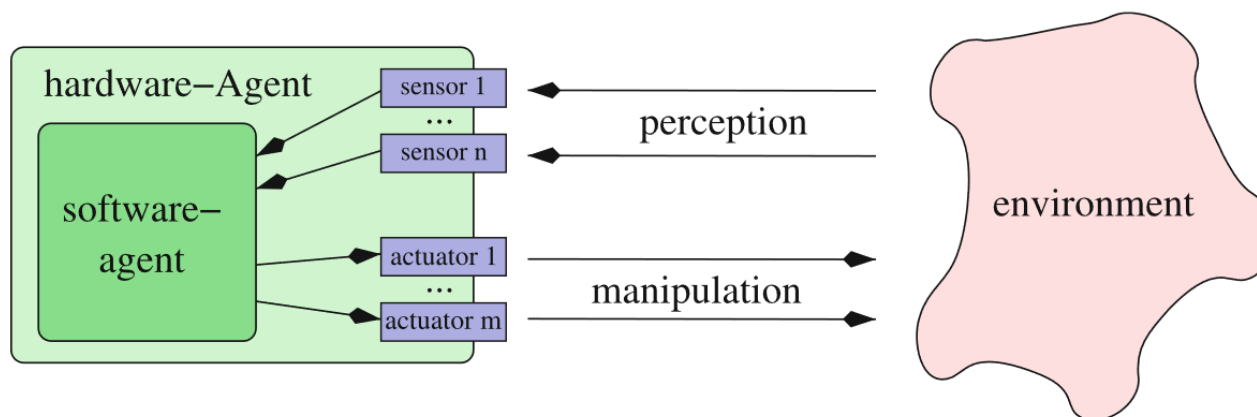


Figure: A hardware agent. (Autonomous robot)

Agents



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

AI and Society

42 Agents

Knowledge-Based Systems

References

Reflex Agents only react to input. (Representing a function from the set of all inputs to the set of all outputs. Useful in problems involving the Markov decision process.)

Agents with memory can also include the past in their decisions. (Not representing a function. The same input may lead to different outputs because of the past memory.)

Goal-based Agents act depending on the goal.

Learning Agents are capable of changing themselves given training examples or through positive or negative feedback, such that the average utility of their actions grows over time.

Distributed Agents the intelligence are not localized in one agent, but rather can only be seen through cooperation of many agents.

Agents



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

AI and Society

43 Agents

Knowledge-Based Systems

References

Example 1 (Spam filter)

A *spam filter* is an agent that puts incoming emails into wanted or unwanted (spam) categories, and deletes any unwanted emails.

Agent 1:

		correct class	
		wanted	spam
spam filter decides	wanted	189	1
	spam	11	799

Agent 2:

		correct class	
		wanted	spam
spam filter decides	wanted	200	38
	spam	0	762

Which agent is better?

- *Agent 1 makes fewer errors than Agent 2*: out of the same set of 1000 emails, Agent 1: 12 errors, Agent 2: 38 errors.
- *But those fewer errors may be severe*: the user loses 11 potentially important emails.
- Because there are in this case *two types of errors of differing severity*, each error should be *weighted with the appropriate cost factor*.



Cost-based Agents

The goal of a *cost-based agent* is to *minimize the long-term cost (i.e. the average cost) caused by wrong decisions*. The sum of all weighted errors results in the total cost.

- Example: The medical diagnosis system LEXMED.

Utility-based Agents

The goal of a *utility-based agent* is to *maximize the utility derived from correct decisions in the long term*, that is, on average. The sum of all decisions weighted by their respective utility factors gives the total utility.

Agents



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

AI and Society

45 Agents

Knowledge-Based Systems

References

Exercise 4 ([Ertel 2025], Exercise 1.6, p. 23)

- (a) Determine for both agents in the spam filter example the costs created by the errors and compare the results. Assume here that having to manually delete a spam email costs one cent and retrieving a deleted email, or the loss of an email, costs one dollar (= 100 cents).
- (b) Determine for both agents the profit created by correct classifications and compare the results. Assume that for every desired email recognized, a profit of one dollar accrues and for every correctly deleted spam email, a profit of one cent.

Agent 1:

		correct class	
		wanted	spam
spam filter decides	wanted	189	1
	spam	11	799

Agent 2:

		correct class	
		wanted	spam
spam filter decides	wanted	200	38
	spam	0	762

Agents



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

AI and Society

46 Agents

Knowledge-Based Systems

References

Environment:

observable the agent always knows the complete state of the world. (E.g., chess computer.)

partially observable the agent does not always know the complete state of the world. (E.g., robot.)

deterministic the current state and the action(s) taken determine exactly one next state. (E.g., 8-puzzle.)

nondeterministic the same current state and action(s) can lead to more than one possible next state. (E.g., a robot action with uncertain outcomes.)

discrete state, time, percept, and action variables take values in finite or countably infinite sets. (E.g., chess.)

continuous at least one state, time, percept, or action variable varies over a continuous range. (E.g., position, speed, or steering angle in driving.)

Exercise 5 ([Ertel 2025], Exercise 1.4, p. 23)

- (a) Why is a deterministic agent with memory not a function from the set of all inputs to the set of all outputs, in the mathematical sense?
- (b) How can one change the agent with memory, or model it, such that it becomes equivalent to a function but does not lose its memory?

Exercise 6 ([Ertel 2025], Exercise 1.5, p. 23)

Let there be an agent with memory that can move within a plane. From its sensors, it receives at clock ticks of a regular interval Δt its exact position (x, y) in Cartesian coordinates.

- (a) Give a formula with which the agent can calculate its velocity from the current time t and the previous measurement of $t - \Delta t$.
- (b) How must the agent be changed so that it can also calculate its acceleration? Provide a formula here as well.

Knowledge-Based Systems



Strict separation of:

Knowledge

Inference mechanism the system or program, which uses the knowledge to, for example, reach conclusions, answer queries, or come up with a plan.

Knowledge base (KB) where the knowledge is stored.

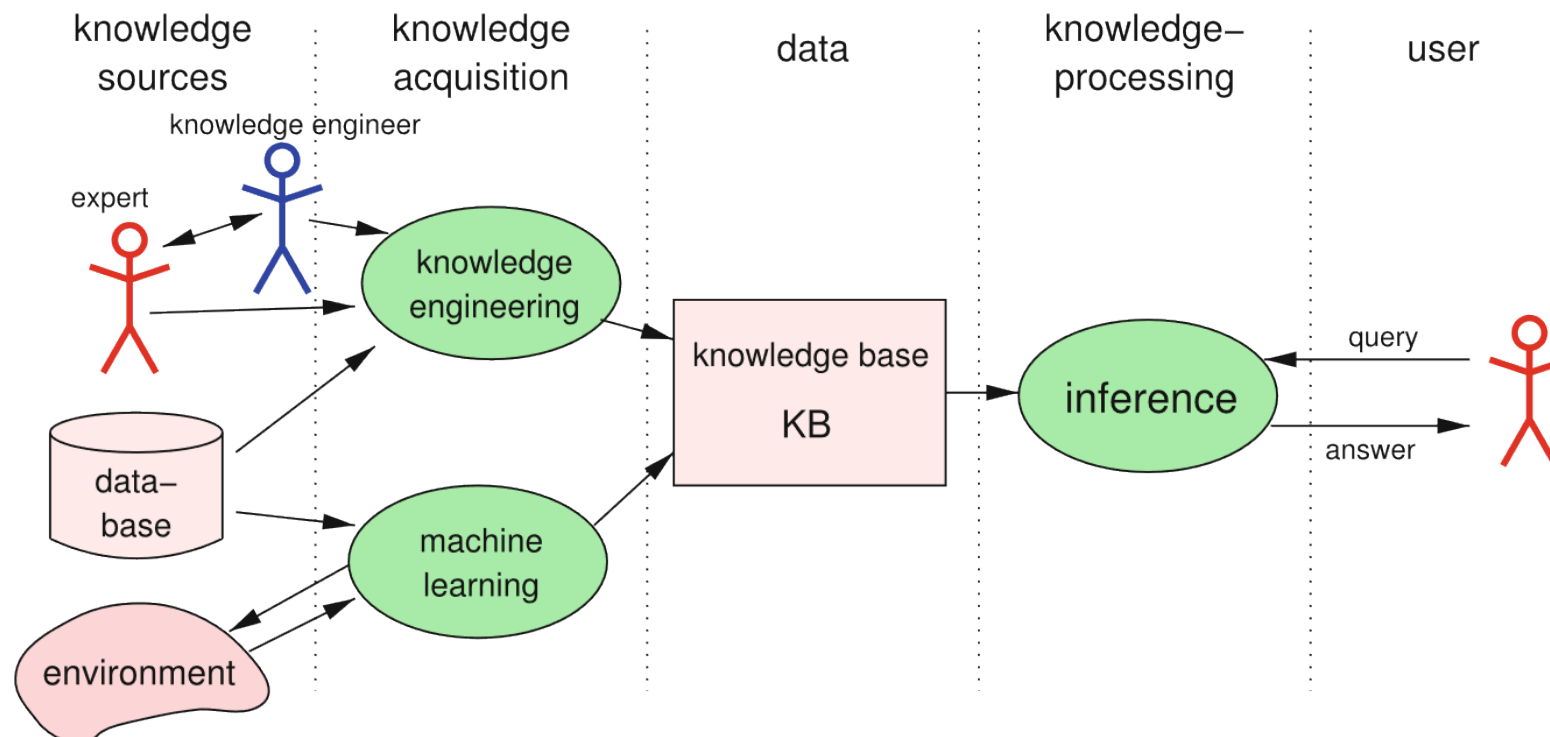


Figure: Structure of a classic knowledge-based system.

Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

AI and Society

Agents

48 Knowledge-Based Systems

References

Knowledge-Based Systems



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

AI and Society

Agents

49 Knowledge-Based Systems

References

Knowledge Engineering: How to get the knowledge into the AI system?

- knowledge sources: human experts, the knowledge engineer, and databases.
- machine learning: learning from data or experience, including examples, observations, feedback, and active exploration.

Separation of knowledge and inference

- The same inference mechanism can often be reused across applications that use the same knowledge-representation framework.
- Knowledge can be stored declaratively. (In the knowledge base there is only a description of the knowledge, which is independent from the inference system in use.)
- *Concrete payoff*: adapting a medical expert system to another disease can be easier: revise or replace the knowledge base and reuse the inference engine when the new knowledge fits the same framework.

Knowledge Representation with formal languages

- Propositional logic.
- First-order predicate logic (shortly: PL1).
- Probabilistic logic.
- Decision trees.

Knowledge-Based Systems

Large-scale example: IBM Watson



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial Intelligence (AI)?

The History of AI

AI and Society

Agents

50 Knowledge-Based Systems

References

- *Watson* (IBM) is a question-answering software agent: it is fed clues in natural language and returns answers.
- Its knowledge base comprises about *four terabytes*, including the full text of Wikipedia. Watson was built in IBM's *DeepQA* project, combining natural language processing, information retrieval, machine learning, knowledge representation and reasoning, and massively parallel computation.
- In February 2011, Watson defeated the human champions Brad Rutter and Ken Jennings on the quiz show “Jeopardy!”.
- Watson’s fast responses relied on a massively parallel *DeepQA* architecture and *90 IBM Power 750 servers* with *2,880 POWER7 processor cores*.

References



Introduction

Hoàng Anh Đức

Additional Materials

What is Artificial
Intelligence (AI)?

The History of AI

AI and Society

Agents

Knowledge-Based
Systems

51

References



Ertel, Wolfgang (2025). *Introduction to Artificial Intelligence*. 3rd. Springer. DOI:
10.1007/978-3-658-43102-0.



Rich, Elaine (1983). *Artificial intelligence*. McGraw-Hill.