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

Contents

Additional Materials

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

- https://youtu.be/katiy95_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

What is Artificial Intelligence (AI)?

- "*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.

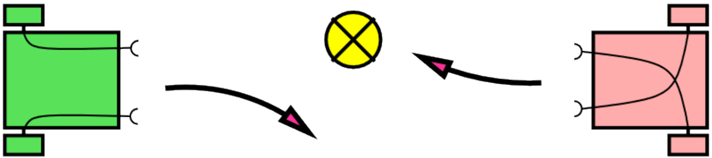


Figure: Two very simple Braitenberg vehicles and their reactions to a light source.

- 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.

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.
$\Rightarrow$ 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)?

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)?

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)?

- 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.

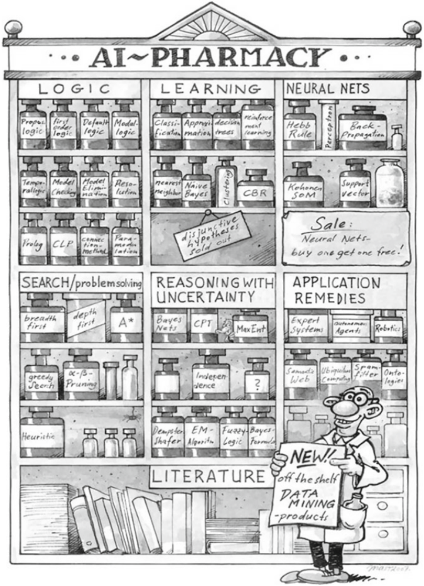
What is Artificial Intelligence (AI)?

Brain Science and Problem Solving

Different approaches:

- How does the human brain work?
- Problem oriented: building intelligent agents!
- General store!

Just as in medicine, there is *no universal method for all application areas of AI*, rather a great number of possible solutions for the great number of various everyday problems, big and small.



What is Artificial Intelligence (AI)?

The Turing Test and Chatterbots

Alan Turing defined intelligence machine—a machine which passes the following test.

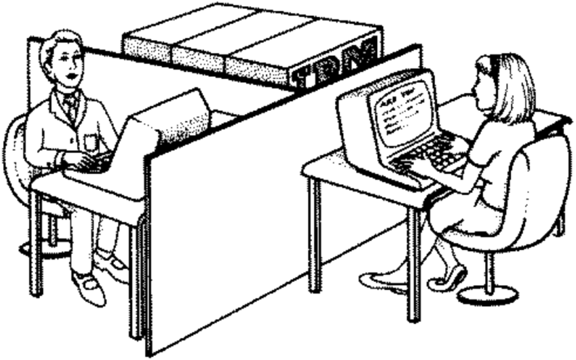


Figure: Turing test: The machine passes the test, if it can mislead Alice in 30% of the cases.

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

- Joseph Weizenbaum (computer critic): the program Eliza talks to his secretary (1966). 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

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

Test the Large Language Models (LLMs) ChatGPT (https://chat.openai.com) and Bard (https://bard.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

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

- In the 1930s Kurt Gödel, Alonso Church, and Alan Turing laid important foundations for logic and theoretical computer science.
- Of particular interest for AI are Gödel's theorems
  - Every true statement that can be formulated in predicate logic is provable using the rules of a formal calculus.
  - In higher-order logics[1], 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.

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?

- In the 1950s. Newell and Simon introduced *Logic Theorist*, *the first automatic theorem prover*, 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: first introducing the name “Artificial Intelligence”.
- In 1965, Robinson invents the *resolution calculus* for predicate logic.

The History of AI
Logic Solves (Almost) All Problems

- In the 1970s, *PROLOG* was introduced. 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*.
- 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.

The History of AI

The New Connectionism

- 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: *Neural networks were 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

- 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.

- 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 brings no new techniques to AI, rather it introduces the requirement of using large databases to gain explicit knowledge.

- *Distributed artificial intelligence:* 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 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.

- *Deep learning networks* can for example learn to classify images with very high accuracy.
- Since image classification is of crucial importance for all types of smart robots, this initiated the *AI revolution* which in turn leads to smart self-driving cars and service robots.

1931 The Austrian Kurt Gödel shows that in first-order *predicate logic* all true statements 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 genetic algorithms.

1951 Marvin Minsky develops a neural network machine. With 3000 vacuum tubes he simulates 40 neurons.

1955 Arthur Samuel (IBM) builds a learning checkers program that plays better than its developer.

1956 McCarthy organizes a conference in Dartmouth College. Here the name *Artificial Intelligence* was first introduced.
Newell and Simon of Carnegie Mellon University (CMU) present the *Logic Theorist*, the first symbol-processing computer program.

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 imitates human thought.

1963 McCarthy founds the AI Lab at Stanford University.

1965 Robinson invents the *resolution calculus* for predicate logic.

The History of AI
Summary

1966 Weizenbaum's program Eliza carries out dialog with people in natural language.

1969 Minsky and Papert show in their book *Perceptrons* that the perceptron, a very simple neural network, can only represent linear functions.

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.

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 Nettalk learns to read texts aloud.

1990 Pearl, Cheeseman, Whittaker, Spiegelhalter bring probability theory into AI with *Bayesian networks*. Multi-agent systems become popular.

1992 Tesauros 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.

The History of AI
Summary

1997 IBM's chess computer Deep Blue defeats the chess world champion Gary 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 begin to improve their behavior through learning.

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.

2015 Daimler premiers 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.
Paintings in the style of the Old Masters can be automatically generated with deep learning. AI becomes creative!

2016 The Go program AlphaGo by Google DeepMind beats the European champion 5:0 in January and Korean Lee Sedol, one of the world's best Go players, 4:1 in March. Deep learning techniques applied to pattern recognition, as well as reinforcement learning and Monte Carlo tree search lead to this success.

The History of AI
Summary

2017 AlphaGoZero learns to play Go on an even higher level by playing against itself using deep learning, reinforcement learning, and Monte Carlo tree search. Humans have no chance against AlphaGoZero. This approach is generalized in AlphaZero, which achieved superhuman chess abilities in 24 hours of training against itself.

2020 The GPT3 language model trained on Wikipedia and many books by the company OpenAI can conduct academic conversations and passes the Turing Test.

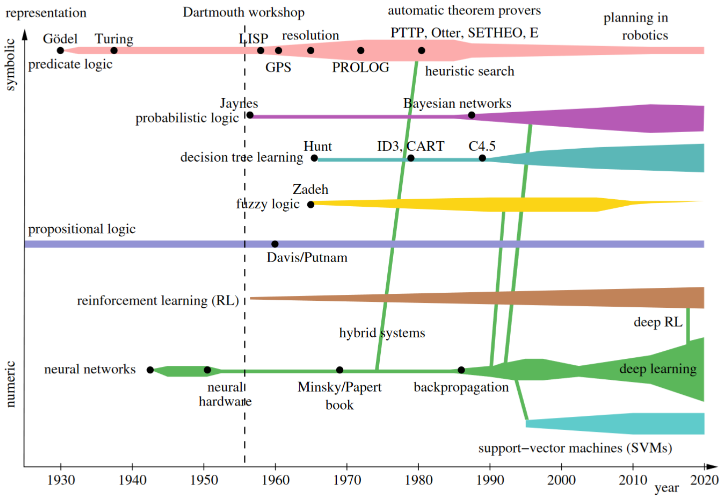


Figure: History of the various AI areas. The width of the bars indicates prevalence of the method's use.

AI and Society

The purpose of AI

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

AI and Society

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.

AI and Society

Not necessarily!

- Long working hours, stress, burnout, and declining real wages have persisted for decades while the productivity is continually increasing.
- Economists attribute this to competitive pressure, forcing companies to lower production costs and lay off workers, leading to unemployment.
- 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.
  - Increased consumption does not result in happiness but contributes to rising mental illness.
  - Bad effects on our living conditions. (We are over-exploiting nature's nonrenewable resources.)

AI and Society

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 and Society

AI in transportation: Autonomous vehicles that can operate independently such as robotics cars (= electric self-driving cars), robotics taxi, etc.

- Pros:
  - Comfort and convenience.
  - Reducing in the number of cars and places for parking.
  - More environmentally friendly than current vehicles.
  - Increase safety.
- Cons:
  - No more “taxi driver” job.
  - 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.)

Agents

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

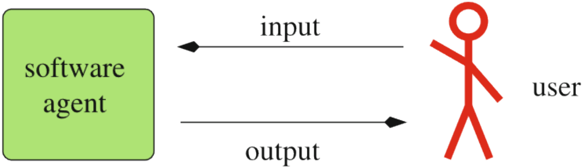


Figure: A software agent with user interaction.

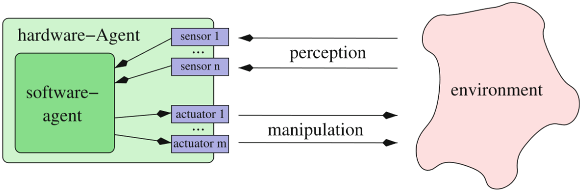


Figure: A hardware agent. (Autonomous robot)

Agents

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 leads 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.

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*.

Agents

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.

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

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 same action of the agent leads to the same result. (E.g., 8-puzzle.)
- **nondeterministic** the same action of the agent may not lead to the same result. (E.g., chess computer, robot.)
- **discrete** finitely many states and actions occur. (E.g., chess computer.)
- **continuous** infinitely many states or actions occur.

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.

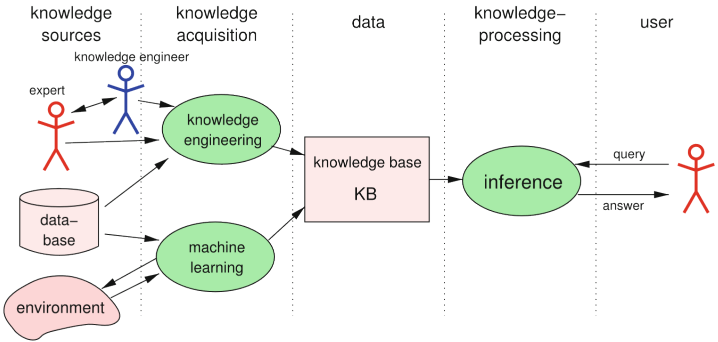


Figure: Structure of a classic knowledge-based system.

Knowledge-Based Systems

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

- knowledge sources: human experts, the knowledge engineer, and databases.
- machine learning: active exploration of the world.

Knowledge-Based Systems

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

- knowledge sources: human experts, the knowledge engineer, and databases.
- machine learning: active exploration of the world.

Separation of knowledge and inference

- Inference is application-independent (e.g. medical expert system).
- 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.)

Knowledge-Based Systems

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

- knowledge sources: human experts, the knowledge engineer, and databases.
- machine learning: active exploration of the world.

Separation of knowledge and inference

- Inference is application-independent (e.g. medical expert system).
- 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.)

Knowledge Representation with formal languages

- Propositional logic.
- First-order logic (shortly: FOL).
- Probabilistic logic.
- Fuzzy logic.
- Decision trees.

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.