Principles of AI Module1

Topics: Introduction: What is AI? Foundations and History of AI, Intelligent Agents: Agents and environment, Concept of Rationality, The nature of environment, The structure of agents.

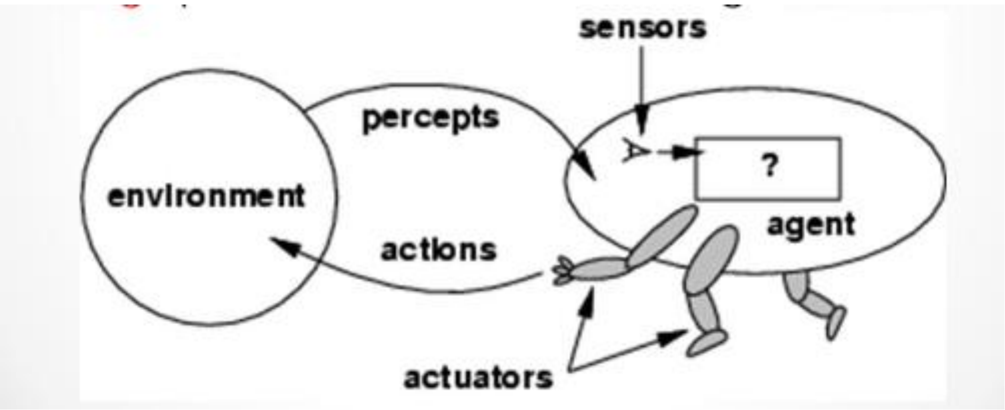
1.1 What is AI?

Artificial Intelligence (AI) is a field of computer science dedicated to develop systems capable of performing tasks that would typically require human intelligence. These tasks include **learning, reasoning, problem-solving, perception, understanding natural language**, and even interacting with the environment. AI aims to create machines or software that can mimic or simulate human cognitive functions.

According to Russell and Norvig, AI can be defined as follows:

" *AI (Artificial Intelligence) is the study of agents that perceive their environment, reason about it, and take actions to achieve goals. Each such agent implements a function that maps percept sequences to actions, and the study of these functions is the subject of AI.*"

Agents: In AI, an agent is any entity that can perceive its environment and take actions to achieve its goals. These agents can be physical robots, software programs, or any system capable of interacting with the world. The agents sense the environment through sensors and act on their environment through actuators. An AI agent can have mental properties such as knowledge, belief, intention, etc.



Source: The Intelligent Agent depicted on books

Percepts: Percepts are the inputs or information an agent receives from the environment. This could include data from sensors, cameras, microphones, or any other sources of information.

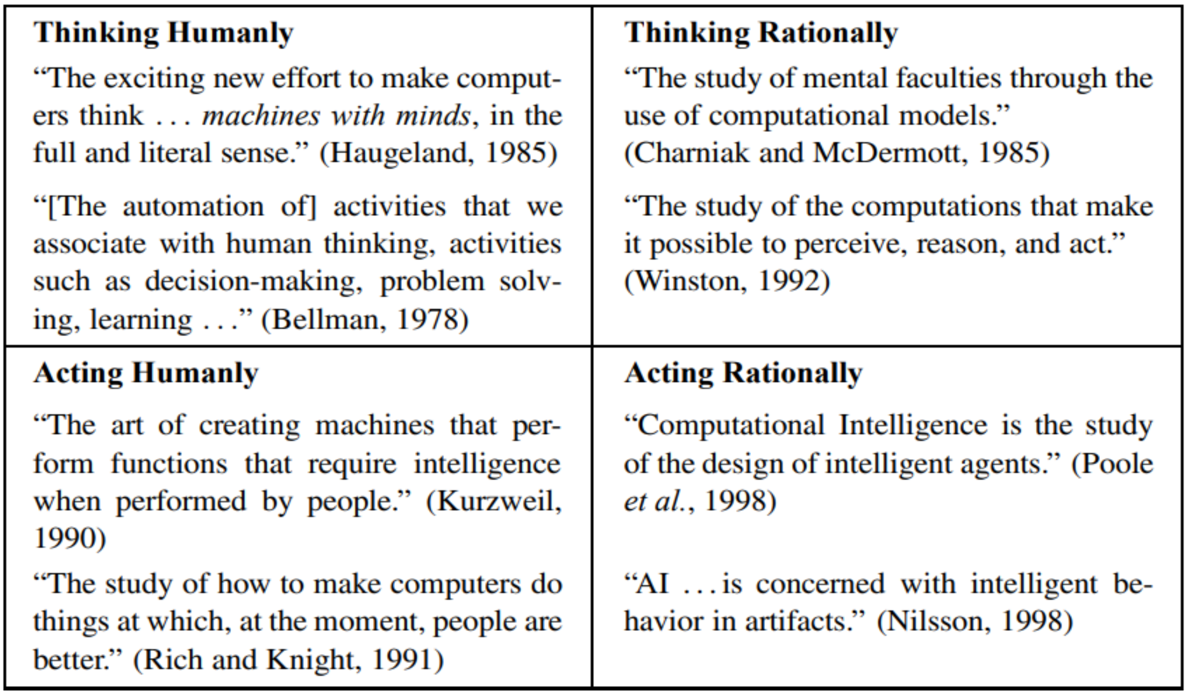
Actions: Actions are the responses or behaviours the agent can exhibit to achieve its goals or objectives. These can be physical movements, data processing, decision-making, or any other form of output.

Function: The central idea of AI is to design a function (or algorithm) that maps percept sequences to actions, allowing the agent to make intelligent decisions based on its observations.

Intelligent Agent: This term refers to the function that enables the agent to act in a way that is considered intelligent, i.e., making decisions that are adaptive and can lead to goal achievement.

Russell and Norvig's definition underscore the goal of AI, which is to create systems that can perceive their environment, reason about it, and take appropriate actions to achieve specific objectives. The field of AI encompasses a wide range of techniques and approaches, from symbolic reasoning to machine learning and deep learning, all aimed at building intelligent agents that can perform tasks typically associated with human intelligence

In **Figure,** there are **eight explanations** of **AI shown in two groups**. The top ones talk about thinking and reasoning, while the bottom ones focus on behavior. The left-side definitions judge based on how close to **human-like performance** they are, while the right-side definitions assess based on an ideal measure called **rationality**. A system is considered **rational** if it makes the best decisions based on the information it has.



Thinking Humanly	**Thinking Rationally**
"The exciting new effort to make computers think . . . *machines with minds*, in the full and literal sense." (Haugeland, 1985) "[The automation of] activities that we associate with human thinking, activities such as decision-making, problem solving, learning . . ." (Bellman, 1978)	"The study of mental faculties through the use of computational models." (Charniak and McDermott, 1985) "The study of the computations that make it possible to perceive, reason, and act." (Winston, 1992)
Acting Humanly	**Acting Rationally**
"The art of creating machines that perform functions that require intelligence when performed by people." (Kurzweil, 1990) "The study of how to make computers do things at which, at the moment, people are better." (Rich and Knight, 1991)	"Computational Intelligence is the study of the design of intelligent agents." (Poole *et al.*, 1998) "AI . . . is concerned with intelligent behavior in artifacts." (Nilsson, 1998)

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

Let us look at the four approaches in more detail.

1. Acting humanly: The Turing Test approach

Turing Test : The Turing test is a method proposed by Alan Turing to evaluate a machine's capability to exhibit intelligent behavior that is indistinguishable from that of a human. In this test, a human evaluator engages in a conversation with both a human and a machine through text, without knowing which is which. If the evaluator cannot reliably differentiate between the machine's responses and the human's, the machine is said to have passed the Turing test, demonstrating human-like intelligence.

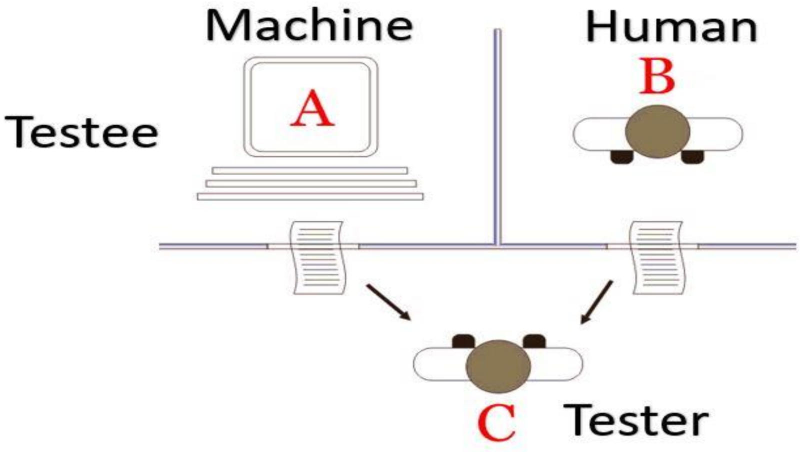


Image Source: https://www.how2shout.com/what-is/what-is-turing-test-and-it-used-for.html

In order to pass **Total Turing Test** the computer must have the following features:

1. **Proficiency in natural language processing** for effective communication in English.
2. **Capability for knowledge representation** to store acquired information.
3. **Automated reasoning to utilize stored data** for answering questions and deriving new conclusions.
4. **Machine learning to adapt to novel situations**, detect patterns, and extrapolate information.
5. **Computer vision** capabilities for object perception.
6. **Robotics proficiency** for manipulating objects and navigating its surroundings.

These **six** disciplines encompass the majority of artificial intelligence (AI), and Turing deserves recognition for designing a test that maintains its relevance six decades later.

2. Thinking humanly: The cognitive modeling approach

To ascertain whether a program exhibits human-like thinking, understanding human thought processes is essential. **This involves exploring the inner workings of the human mind through introspection, psychological experiments, and brain imaging**. Once a precise theory of the mind is established, it can be translated into a computer program. If the program's input-output behavior aligns with human behavior, it suggests shared mechanisms.

3. Thinking rationally: The "laws of thought" approach

Aristotle, an early philosopher, sought to formalize "right thinking" through syllogisms, offering patterns for sound reasoning. These principles, foundational to logic, inspired logicians in the 19th century to develop precise notation for various objects and their relations. By 1965, programs capable of theoretically solving any problem in logical notation emerged, paving the way for the logicist tradition in artificial intelligence.

4. Acting rationally: The rational agent approach

Computer agents are expected to autonomously operate, perceive their environment, persist over time, adapt to change, and pursue goals. A rational agent seeks the best outcome by emphasizing correct inferences, aligning with the Turing Test skills. While correct inference is a facet of rationality, it's not exhaustive, as some situations lack provably correct actions. The rational-agent approach, incorporating knowledge representation, reasoning, and learning, offers generality and scientific advantages. Despite challenges acknowledged in the book, it adopts the hypothesis that perfect rationality is a useful starting point, simplifying complex environments for foundational discussions in the field.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

1.2 The Foundations of Artificial Intelligence

This section offers a concise history of the disciplines influencing AI, focusing on key figures, events, and ideas.

Philosophy: In the philosophical exploration of AI, following questions arises:

- Can formal rules be used to draw valid conclusions?
- How does the mind arise from a physical brain?
- Where does knowledge come from?
- How does knowledge lead to action

Aristotle, formulated laws for the rational mind, developed syllogisms for mechanical reasoning. Historical figures like Ramon Lull, Thomas Hobbes, and Leonardo da Vinci contributed to the idea that mechanical artifacts could perform useful reasoning. Descartes introduced the mind-matter distinction, raising debates on free will and advocating rationalism and dualism. Materialism emerged as an alternative to dualism, positing that the brain's operation constitutes the mind. The empiricism movement, led by Bacon and Locke, emphasized sensory origins of understanding. Logical positivism, developed by the Vienna Circle, combined rationalism and empiricism, shaping the computational theory of mind presented by Carnap and Hempel in analysing knowledge acquisition from experience.

Mathematics: In the mathematical exploration of AI, following questions arises:

- What are the formal rules to draw valid conclusions?
- What can be computed?
- How do we reason with uncertain information?

The mathematical formalization of logic, computation, and probability played crucial roles. George Boole developed Boolean logic, extended later by Frege and Tarski. Kurt Godel's incompleteness theorem revealed the limits of logical deduction. Alan Turing addressed computability, introducing the Church-Turing thesis. The concept of tractability, differentiating polynomial and exponential complexity, emerged in the mid-1960s. NP-completeness theory, by Cook and Karp, identified intractable problems. Probability theory, pioneered by Cardano and advanced by Pascal, Bernoulli, Laplace, and Bayes, became essential in handling uncertain information.

Economics: In the Economics of AI, various attempts have been done to address the following questions:

- How should we make decisions so as to maximize payoff?
- How should we do this when others may not go along?
- How should we do this when the payoff may be far in the future

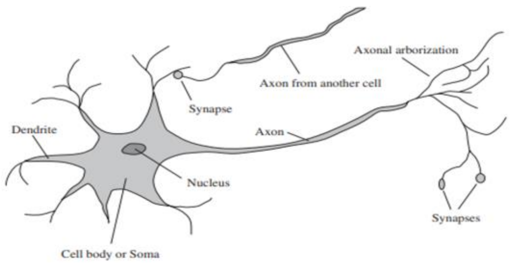
The science of economics began in 1776 with Adam Smith's "An Inquiry into the Nature and Causes of the Wealth of Nations," treating economics as a science focused on individual agents maximizing their economic well-being. Léon Walras formalized the mathematical treatment of utility, later improved by Frank Ramsey, John von Neumann, and Oskar Morgenstern in "The

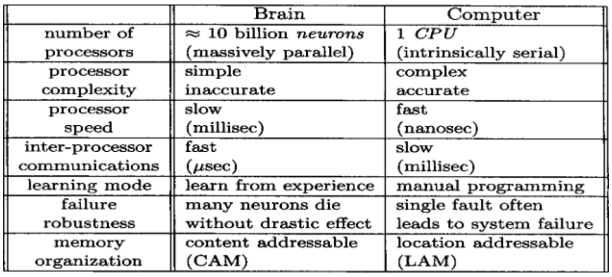
Theory of Games and Economic Behavior" (1944). Decision theory, combining probability and utility theory, provides a framework for decisions under uncertainty. Game theory, developed by Von Neumann and Morgenstern, includes the surprising result that rational agents may adopt randomized policies. Operations research, initiated in World War II, addressed sequential decision problems, formalized by Richard Bellman. Herbert Simon's work on satisficing, making "good enough" decisions, earned him a Nobel Prize in economics in 1978. Recent interest in decision-theoretic techniques for agent systems has emerged since the 1990s.

Neuroscience (How do brains process information?)

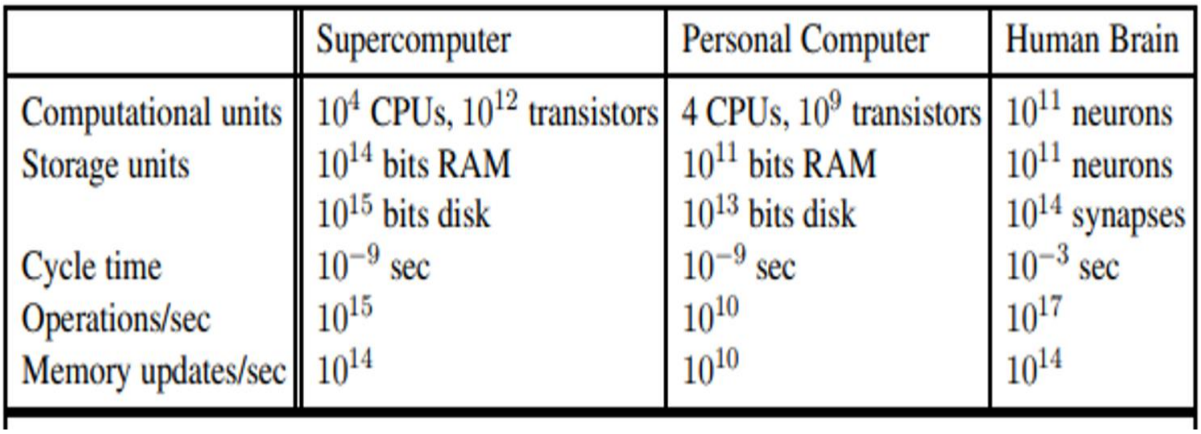
Neuroscience, the study of the nervous system, particularly the brain, aims to understand the mechanisms enabling thought. While the brain's role in cognition has been recognized for millennia, it wasn't until the 18th century that the brain was widely acknowledged as the seat of consciousness. Paul Broca's study of aphasia in 1861 highlighted localized brain areas for specific functions, and Camillo Golgi's staining technique (1873) allowed the observation of individual neurons. Santiago Ramon y Cajal applied mathematical models to the nervous system, while Nicolas Rashevsky pioneered mathematical modeling in neuroscience. Hans Berger's electroencephalograph (EEG) in 1929 and recent developments in functional magnetic resonance imaging (fMRI) provide detailed brain activity images. Despite advances, understanding cognitive processes remains a challenge. Brains and computers differ in properties, with futurists speculating on a singularity when computers surpass human intelligence.

Structure of Neuron and comparison of Human brain with Computers:





	Brain	Computer
number of processors	≈ 10 billion *neurons* (massively parallel)	1 *CPU* (intrinsically serial)
processor complexity	simple inaccurate	complex accurate
processor speed	slow (millisec)	fast (nanosec)
inter-processor communications	fast (μsec)	slow (millisec)
learning mode	learn from experience	manual programming
failure robustness	many neurons die without drastic effect	single fault often leads to system failure
memory organization	content addressable (CAM)	location addressable (LAM)



	Supercomputer	Personal Computer	Human Brain
Computational units	10^4 CPUs, 10^{12} transistors	4 CPUs, 10^9 transistors	10^{11} neurons
Storage units	10^{14} bits RAM	10^{11} bits RAM	10^{11} neurons
	10^{15} bits disk	10^{13} bits disk	10^{14} synapses
Cycle time	10^{-9} sec	10^{-9} sec	10^{-3} sec
Operations/sec	10^{15}	10^{10}	10^{17}
Memory updates/sec	10^{14}	10^{10}	10^{14}

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

Psychology: How do humans and animals think and act?

Scientific psychology traces its origins to Hermann von Helmholtz and Wilhelm Wundt. Helmholtz applied the scientific method to human vision, and Wundt established the first experimental psychology laboratory in 1879. Wundt emphasized introspective experiments, but behaviorism, led by John Watson, emerged, rejecting mental processes. Cognitive psychology, rooted in the information-processing view of the brain, finds early roots in William James's works. Frederic Bartlett and Kenneth Craik, at Cambridge's Applied Psychology Unit, advocated for cognitive modeling. Craik outlined a knowledge-based agent's three key steps: stimulus translation, internal representation manipulation, and retranslation into action. Donald Broadbent continued Craik's work, modeling psychological phenomena as information processing. The field of cognitive science emerged in the U.S., notably at MIT's 1956 workshop where Miller, Chomsky, Newell, and Simon presented influential papers, demonstrating how computer models could address memory, language, and logical thinking in psychology.

Computer Engineering (How can we build an efficient computer?)

For artificial intelligence (AI) to succeed, two key elements are essential: **intelligence and an artifact**, with the **computer** being the chosen artifact. The modern digital electronic computer, crucial for AI, was independently invented in three countries during World War II. Early machines like **Heath Robinson** and **Colossus** paved the way, and the **Z-3**, developed by Konrad Zuse in 1941, marked the first programmable computer. The **ABC**, built by John Atanasoff and Clifford Berry in the early 1940s, was the first electronic computer, but the **ENIAC**, developed by **John Mauchly** and John Eckert, emerged as a highly influential precursor to modern computers.

Each generation of computer hardware since then has seen increased speed, capacity, and reduced costs. While early calculating devices existed, programmable machines like Joseph Marie Jacquard's loom in 1805 and Charles Babbage's designs in the mid-19th century, especially the **Analytical Engine**, were significant precursors to modern computers. Ada Lovelace, Babbage's colleague, is regarded as the world's **first programmer**. AI also acknowledges the debt to computer science's software side, contributing to operating systems, programming languages, and tools, with reciprocal innovation between AI and mainstream computer science.

Control Theory and Cybernetics (How can artifacts operate under their own control?)

Ktesibios of Alexandria, around 250 B.C., constructed the first self-controlling machine—an innovative **water clock with a regulator maintaining** a constant flow rate. This invention marked a shift in the capabilities of artifacts, enabling them to modify behavior in response to environmental changes. Control theory, rooted in stable feedback systems, saw significant developments in the 19th century. James Watt's steam engine governor and Cornelis Drebbel's thermostat exemplify self-regulating feedback control systems.

The central figure in the establishment of control theory, especially cybernetics, was Norbert Wiener (1894–1964). Wiener, along with Arturo Rosenblueth and Julian Bigelow, challenged behaviorist orthodoxy, viewing purposive behavior as arising from regulatory mechanisms minimizing "error" between current and goal states. Wiener's book "Cybernetics" (1948) became influential in shaping the public perception of artificially intelligent machines. W. Ross Ashby in Britain also pioneered similar ideas, emphasizing the creation of intelligence through homeostatic devices.

Modern control theory, particularly stochastic optimal control, aims to design systems maximizing an objective function over time, aligning with AI's goal of creating optimally behaving systems. Despite shared objectives, AI and control theory emerged as distinct fields due to the different mathematical techniques and problem sets each addressed. While control theory relied on calculus and matrix algebra for systems with continuous variables, AI, leveraging logical inference and computation, ventured into addressing problems like language, vision, and planning beyond the scope of control theory.

Linguistics (How does language relate to thought?):

In 1957, B. F. Skinner published "Verbal Behavior," a comprehensive account of the behaviorist approach to language learning. However, the book faced significant criticism in a review by linguist Noam Chomsky, who had recently published his own book, "Syntactic Structures." Chomsky argued that behaviorism failed to explain the creativity inherent in language acquisition, particularly how children understand and generate novel sentences. Chomsky's theory, rooted in syntactic models dating back to the ancient linguist Panini, offered a formal framework programmable in principle.

The intersection of modern linguistics and AI occurred around the same time, giving rise to computational linguistics or natural language processing. The complexity of language understanding became evident, extending beyond sentence structure to encompass subject matter and context, a realization that gained prominence in the 1960s. Early work in knowledge representation, aimed at rendering knowledge in a computationally usable form, was closely tied to language and influenced by linguistic research, itself connected to philosophical analyses of language over decades.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

Foundations of AI (Summarized contributions)

Philosophy

Name	Year	Suggestion
Aristotle	384–322 B.C.	Formulated laws for the rational mind and developed syllogisms for mechanical reasoning.
Ramon Lull	Died 1315	Envisioned useful reasoning by mechanical artifacts.
Thomas Hobbes	1588–1679	Likened reasoning to numerical computation.
Leonardo da Vinci	1452–1519	Designed a functional mechanical calculator around 1500.
Wilhelm Schickard	1592–1635	Constructed the first known calculating machine in 1623.
Blaise Pascal	1623–1662	Built the famous Pascaline calculator in 1642.
Gottfried Wilhelm Leibniz	1646–1716	Speculated on machines thinking and acting independently.
Thomas Hobbes (Leviathan)	1588–1679	Suggested the concept of an "artificial animal."
René Descartes	1596–1650	Discussed the mind-matter distinction and advocated rationalism and dualism.
Vienna Circle (Rudolf Carnap)	1891–1970	Developed logical positivism, combining rationalism and empiricism.
David Hume	1711–1776	Proposed the principle of induction in the 18th century.
Francis Bacon	1561–1626	Initiated the empiricism movement.
John Locke	1632–1704	Emphasized the sensory origins of understanding.
Ludwig Wittgenstein	1889–1951	Contributed to the work of the Vienna Circle.
Bertrand Russell	1872–1970	Contributed to the work of the Vienna Circle.
Carl Hempel	1905–1997	Introduced the confirmation theory to analyze knowledge acquisition.
Antoine Arnauld	1612–1694	Correctly described a quantitative formula for deciding what action to take in cases like this.
John Stuart Mill	1806-1873	John Stuart Mill's (1806–1873) book Utilitarianism (Mill, 1863) promoted the idea of rational decision criteria in all spheres of human activity

Mathematics

George Boole	1815–1864	Developed propositional (Boolean) logic.
Gottlob Frege	1848–1925	Extended Boole's logic to include objects and relations, creating first-order logic.
Alfred Tarski	1902–1983	Introduced a theory of reference, linking logic to the real world.
Kurt Godel	1906–1978	Formulated incompleteness theorems, highlighting limits of logical deduction.
Alan Turing	1912–1954	Investigated computability and introduced the Church-Turing thesis.
Steven Cook	1971	Pioneered NP-completeness theory, identifying intractable problems.
Richard Karp	1972	Contributed to NP-completeness theory, identifying intractable problems.
Gerolamo Cardano	1501–1576	Framed the idea of probability in the context of gambling events.
Blaise Pascal	1623–1662	Applied probability to predict outcomes in gambling games.
James Bernoulli	1654–1705	Advanced probability theory.
Pierre Laplace	1749–1827	Contributed to probability theory and introduced statistical methods.
Thomas Bayes	1702–1761	Proposed Bayes' rule for updating probabilities in the light of new evidence.

Economics

Name	Year	Contribution
Adam Smith	1723–1790	Published "An Inquiry into the Nature and Causes of the Wealth of Nations."
Léon Walras	1834–1910	Formalized the mathematical treatment of utility.
Frank Ramsey	1931	Contributed to the formalization of utility theory.
John von Neumann	1903–1957	Developed game theory in "The Theory of Games and Economic Behavior" (1944).
Oskar Morgenstern	1902–1977	Collaborated with von Neumann in developing game theory.
Richard Bellman	1920–1984	Formalized Markov decision processes in operations research.
Herbert Simon	1916–2001	Pioneered satisficing and received the Nobel Prize in economics in 1978.

Neuroscience

Name	Year	Contribution
Aristotle	335 B.C.	Acknowledged human brains as proportionally larger than other animals.
Paul Broca	1824–1880	Demonstrated localized brain areas for specific functions, particularly speech.
Camillo Golgi	1843–1926	Developed staining technique allowing observation of individual neurons.
Santiago Ramon y Cajal	1852–1934	Conducted pioneering studies on the brain's neuronal structures.
Nicolas Rashevsky	1936, 1938	Applied mathematical models to the study of the nervous system.
Hans Berger	1929	Invented the electroencephalograph (EEG) for measuring intact brain activity.
John Searle	1992	Coined the phrase "brains cause minds," emphasizing the connection between brains and consciousness.

Psychology

Name	Year	Contribution
Hermann von Helmholtz	1821–1894	Applied the scientific method to the study of human vision, contributing to physiological optics.
Wilhelm Wundt	1832–1920	Established the first laboratory of experimental psychology in 1879, emphasizing introspective experiments.
John Watson	1878–1958	Led the behaviorism movement, rejecting theories involving mental processes and advocating for the study of objective measures.
William James	1842–1910	Contributed to cognitive psychology, viewing the brain as an information-processing device.
Frederic Bartlett	1886–1969	Directed Cambridge's Applied Psychology Unit, fostering cognitive modeling.
Kenneth Craik	1943	Developed a knowledge-based agent model, emphasizing stimulus translation, internal representation manipulation, and retranslation into action.
Donald Broadbent	1926–1993	Continued Craik's work, modeling psychological phenomena as information processing.
George Miller	1956	Presented influential work on the psychology of memory at the 1956 MIT workshop.
Noam Chomsky	1956	Presented influential work on language models at the 1956 MIT workshop.
Allen Newell and Herbert Simon	1956	Presented influential work on logical thinking models at the 1956 MIT workshop.

Computer Engineering

Name	Year	Contribution
Alan Turing	1940	Led the team that built the electromechanical Heath Robinson computer for deciphering German messages.
Konrad Zuse	1941	Invented the Z-3, the first operational programmable computer in Germany.
John Atanasoff and Clifford Berry	1940–1942	Assembled the ABC, the first electronic computer at Iowa State University.
John Mauchly and John Eckert	1940s	Developed the ENIAC, a highly influential precursor to modern computers.
Joseph Marie Jacquard	1805	Invented a programmable loom that used punched cards for storing weaving instructions.
Charles Babbage	1792–1871	Designed the Difference Engine and Analytical Engine, significant precursors to modern computers.
Ada Lovelace	1815–1852	Collaborated with Charles Babbage, wrote programs for the Analytical Engine, and is considered the world's first programmer.

Control Theory and Cybernetics

Name	Year	Contribution
Ktesibios of Alexandria	c. 250 B.C.	Built the first self-controlling machine—a water clock with a regulator that maintained a constant flow rate.
James Watt	1736–1819	Created the steam engine governor, a self-regulating feedback control system.

Cornelis Drebbel	1572–1633	Invented the thermostat, another example of a self-regulating feedback control system.
Norbert Wiener	1894–1964	Pioneered control theory and cybernetics, exploring the connection between biological and mechanical control systems.
W. Ross Ashby	1903–1972	Elaborated on the idea that intelligence could be created by using homeostatic devices containing appropriate feedback loops.

Linguistics

Name	Year	Contribution
B. F. Skinner	1957	Published "Verbal Behavior," a comprehensive account of behaviorist language learning, sparking subsequent debate.
Noam Chomsky	1957	Criticized behaviorist theory in a review of Skinner's book, highlighting the inadequacy in explaining language creativity. Published "Syntactic Structures" with a formal, programmable theory of language.
Panini	c. 350 B.C.	Ancient linguist whose syntactic models influenced Chomsky's theory and contributed to the formal understanding of language.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

1.3 History of Artificial Intelligence

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

The history of artificial intelligence (AI) is a fascinating journey spanning several decades. Here's a brief timeline highlighting key milestones in the development of AI, merged with a structured overview:

1.3.1 The Gestation of Artificial Intelligence (1943–1955)

1943: Warren McCulloch and Walter Pitts develop the first mathematical model of a neural network.
1949: Hebbian learning, formulated by Donald Hebb, becomes a lasting influence on neural network development.
1950: Alan Turing introduces the Turing Test and key AI principles.
1951: UNIVAC I, the first commercially produced computer, is used for statistical analysis, laying the groundwork for data processing.
1951: McCarthy earns his PhD and later plays a pivotal role in establishing AI at Dartmouth College.

1.3.2 The Birth of Artificial Intelligence (1956)

1956: John McCarthy organizes a landmark AI workshop at Dartmouth College, marking the official birth of artificial intelligence.
1956: The term "artificial intelligence" is coined at the Dartmouth Conference.

1.3.3 Early Enthusiasm, Great Expectations (1952–1969)

Late 1950s: IBM produces early AI programs challenging predefined task limitations.
1958: McCarthy defines the Lisp language and introduces time-sharing.
1963: Marvin Minsky establishes Stanford's AI lab, emphasizing practical functionality.
1965: Joseph Weizenbaum creates ELIZA, an early natural language processing program.
1966–1973: Setbacks occur as early successes fail to scale, leading to reduced support for AI research.
1969: The Stanford Research Institute develops Shakey, the first mobile robot with reasoning abilities.

1.3.4 A Dose of Reality (1966–1973)

1969: Despite setbacks, the discovery of back-propagation learning algorithms for neural networks leads to a resurgence of interest.
1970s: Initial enthusiasm for AI fades, leading to the first "AI winter" as progress stalls

1.3.5 Knowledge-Based Systems: The Key to Power? (1969–1979)

1969: DENDRAL exemplifies a shift towards domain-specific knowledge.
Late 1970s: The Heuristic Programming Project explores expert systems, emphasizing domain-specific knowledge.
1979: Marvin Minsky and Seymour Papert publish "Perceptrons," a book critical of certain AI approaches.

1.3.6 AI Becomes an Industry (1980–Present)

Early 1980s: The first successful commercial expert system, R1, is implemented at Digital Equipment Corporation.
1981: Japan's "Fifth Generation" project and the U.S.'s Microelectronics and Computer Technology Corporation respond to AI's growing influence.
1980s: Rapid growth followed by the "AI Winter," a period of decline in the AI industry.
1985: Expert systems, software that emulates decision-making of a human expert, gain popularity.

1.3.7 The Return of Neural Networks (1986–Present)

Mid-1980s: Rediscovery of the back-propagation learning algorithm leads to the emergence of connectionist models.
Late 1980s: The AI industry experiences a decline known as the AI Winter.

1.3.8 AI Adopts the Scientific Method (1987–Present)

Late 1980s: AI shifts towards a more scientific and application-focused approach, experiencing a revival in the late 1990s.
1990-2005: Neural Networks Resurgence and Practical Applications
1997: IBM's Deep Blue defeats chess champion Garry Kasparov.
1999: Rodney Brooks introduces the concept of "embodied intelligence" with Cog, a humanoid robot.

1.3.9 The Emergence of Intelligent Agents (1995–Present)

Late 1980s: The SOAR architecture addresses the "whole agent" problem.
Late 1990s–2000s: AI technologies underlie Internet tools, contributing to search engines, recommender systems, and website aggregators.

1.3.10 The Availability of Very Large Data Sets (2001–Present)

Late 1990s: A revival in AI with a shift towards a more scientific approach.
2000s: Emphasis on the importance of large datasets in AI research, leading to significant advancements.

2001: The DARPA Grand Challenge initiates research in autonomous vehicles.
2005: Stanford's Stanley wins the DARPA Grand Challenge, showcasing advances in self-driving technology.
2006: Geoffrey Hinton and colleagues publish a paper on deep learning, reigniting interest in neural networks.
2011: IBM's Watson wins Jeopardy!, demonstrating the power of natural language processing.
2012: AlexNet, a deep convolutional neural network, achieves a breakthrough in image recognition at the ImageNet competition.
2012-Present: AI in the Mainstream and Ethical Concerns
2014: Facebook's AI lab introduces DeepFace for facial recognition, reaching human-level accuracy.
2016: AlphaGo, an AI developed by DeepMind, defeats world champion Go player Lee Sedol.
2018: OpenAI releases GPT-2, a large-scale language model.
2020s: AI applications become integral in various industries, raising concerns about ethics, bias, and job displacement.

Present: AI applications are deeply embedded in various industries, marking a new spring for the field.
The history of AI is marked by cycles of optimism, followed by periods of stagnation, but recent years have seen unprecedented progress and integration of AI technologies into everyday life. Ongoing research and ethical considerations will continue to shape the future of artificial intelligence.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

1.4 Intelligent Agents

Topics: Agents and environment, Intelligent Agents, Concept of Rationality, The nature of environment, The structure of agents.

1.4.1 Agents and Environments:

An agent is defined as anything capable of perceiving its environment through sensors and acting upon that environment through actuators. This basic concept is depicted in Figure 2.1.

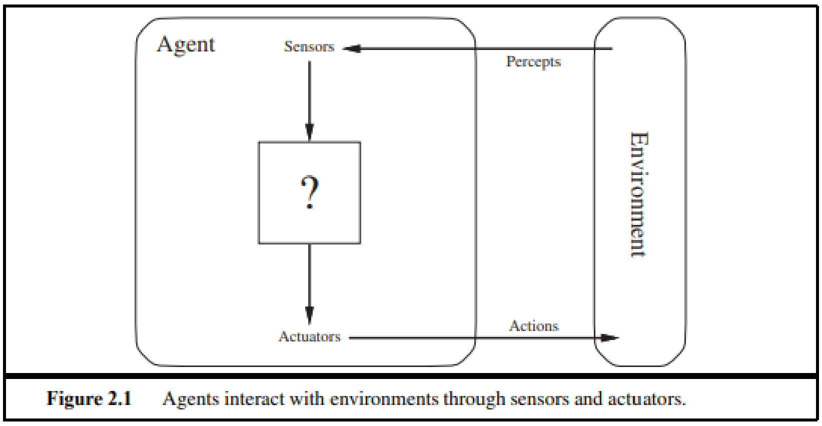


Figure 2.1 Agents interact with environments through sensors and actuators.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

For instance, a **human agent** employs eyes, ears, and other sensory organs while utilizing hands, legs, vocal tract, etc., as actuators.

In contrast, a **robotic agent** might use cameras and infrared range finders as sensors and various motors as actuators.

A **software agent** takes in sensory inputs such as keystrokes, file contents, and network packets and responds by displaying on the screen, writing files, or sending network packets.

Percept: The term **"percept"** refers to an agent's perceptual inputs at any given moment, and a percept sequence encompasses the complete history of everything the agent has perceived. Mathematically, an agent's behavior is described by the agent function, which maps a given percept sequence to an action.

Table: A table serves as an external representation of an agent's behaviour, specifically presented in tabular form that outlines the agent's actions for every possible percept sequences. The table contains entries corresponding to different percept sequences and the corresponding actions the agent would take in response to those sequences

Agent Program: An agent program represents the internal implementation of an agent's behavior. It is a concrete realization of the abstract agent function, running within a physical system. The agent program executes the specified actions determined by the agent function in response to the sensory inputs received by the agent. It involves the use of computational processes, algorithms, or code that dictates how the agent translates perceptual inputs into actions within its environment.

Example: Consider a simple example—the vacuum-cleaner world as depicted in **Figure 2.2**.

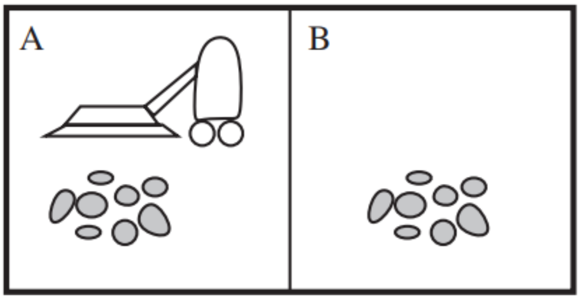
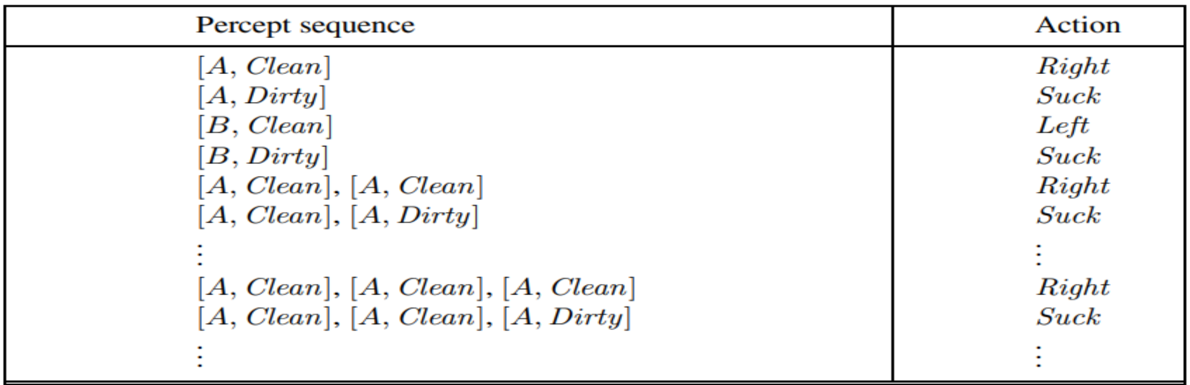


Figure 2.2 A vacuum-cleaner world with just two locations.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

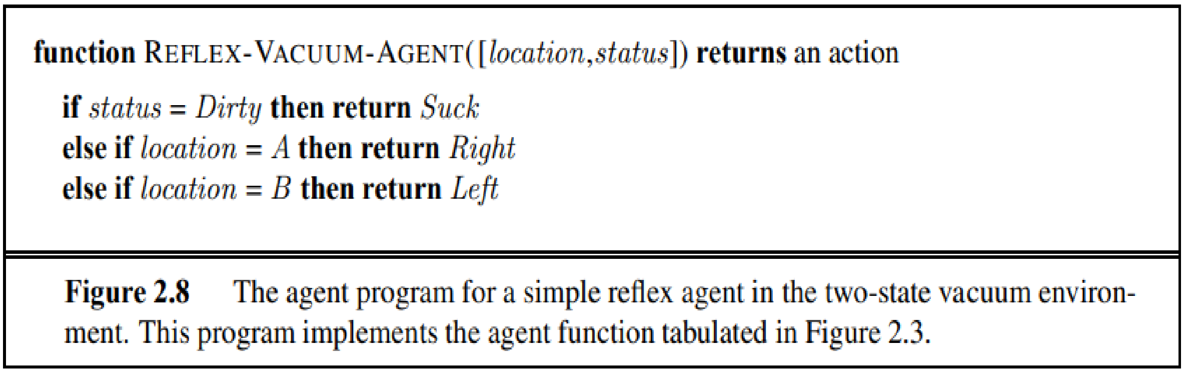
This world consists of **two locations**, squares A and B, where the vacuum agent perceives its location and the presence of dirt. The agent can choose to move left, move right, suck up dirt, or do nothing. An example of a simple agent function is presented: if the current square is dirty, then suck; otherwise, move to the other square. **Figure 2.3** provides a partial tabulation of this agent function and **Figure 2.8** provides agent program.



Percept sequence	Action
$[A, Clean]$	*Right*
$[A, Dirty]$	*Suck*
$[B, Clean]$	*Left*
$[B, Dirty]$	*Suck*
$[A, Clean], [A, Clean]$	*Right*
$[A, Clean], [A, Dirty]$	*Suck*
⋮	⋮
$[A, Clean], [A, Clean], [A, Clean]$	*Right*
$[A, Clean], [A, Clean], [A, Dirty]$	*Suck*
⋮	⋮

Figure 2.3 Partial tabulation of a simple agent function for the vacuum-cleaner world shown in Figure 2.2.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015



```
function REFLEX-VACUUM-AGENT([location,status]) returns an action
  if status = Dirty then return Suck
  else if location = A then return Right
  else if location = B then return Left
```

Figure 2.8 The agent program for a simple reflex agent in the two-state vacuum environment. This program implements the agent function tabulated in Figure 2.3.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

1.4. 2: Intelligent Agents

In artificial intelligence (AI), intelligent agents are entities that perceive their environment and take actions to achieve specific goals. These agents are designed to operate autonomously, making decisions based on their perception of the world and their programming or learning mechanisms. The concept of intelligent agents is fundamental in AI as it provides a framework for building systems that can exhibit intelligent behaviour.

Key characteristics of intelligent agents include:

Perception: Agents have the ability to perceive or sense their environment. This involves gathering information from the surroundings through sensors or other means.
Action: Intelligent agents can take actions in response to their perceptions. These actions are chosen to influence the state of the environment or achieve specific goals.
Autonomy: Agents operate autonomously, making decisions and taking actions without direct human intervention. Autonomy allows them to adapt to changing conditions in their environment.
Goal-Directed Behaviour: Intelligent agents are typically designed to achieve specific goals. These goals can be predefined by a programmer or learned by the agent through experience.
Learning and Adaptation: Many intelligent agents have the capability to learn from experience and adapt their behaviour over time. This learning process may involve acquiring new knowledge or improving decision-making strategies.
Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

Communication: In some cases, intelligent agents can communicate with other agents or systems. Communication enables collaboration and coordination between multiple agents to achieve common goals.

1.4.3 The Concept of Rationality (Good behaviour)

A rational agent is defined as one that consistently does right thing and makes decisions leading to favourable outcomes. This is reflected in the **correct completion of every entry in the agent function table**. However, determining what constitutes the "**right thing**" involves a detailed consideration of consequences.

- **Assessing Consequences:** To evaluate the rationality of an agent's actions, one must examine the consequences of its behaviour. The agent's sequence of actions influences the environment, causing it to transition through different states. Desirability of this sequence is gauged by a performance measure that evaluates resultant environment states.
- **Performance Measure**: Performance measures play a pivotal role in evaluating an agent's behavior and defining success. These measures are instrumental in shaping the assessment of successful behavior, focusing specifically on the resulting environment states rather than the agent's subjective perception of its performance.

    - **Challenges and Considerations in Designing Performance Measures**: Designing a suitable performance measure is a multifaceted task that involves careful consideration of desired outcomes in the environment. This complex process requires avoiding potential pitfalls, such as the introduction of biased measures influenced by the agent's personal opinion.

        - **Example:** Taking the vacuum-cleaner agent as an example, the design of a performance measure becomes critical. Simply measuring dirt cleaned up in an eight-hour shift may lead to undesired behaviours, such as repetitive cleaning and dumping. A more effective measure would reward the agent for maintaining a clean floor.

  - **Alignment with Environmental Goals**: Emphasizing the importance of aligning performance measures with actual environmental goals is crucial. This underscores the need to base measures on tangible objectives within the environment, steering clear of preconceived notions about how the agent should behave.
- **Philosophical Considerations**: Even with careful performance measure design, philosophical questions arise. Comparisons between agents with different approaches (e.g., a consistent but mediocre performer vs. an energetic but intermittently idle cleaner) prompt deeper reflections on preferences in various contexts.

1.4.4 What is Rationality?

Rationality encompasses the state of being reasonable, sensible, and possessing a sound judgment. **Rational** agent is the one, which is capable of doing expected actions to maximize its performance measure, on the basis of −

- Its percept sequence.
- Its built-in knowledge base.

A rational agent always performs right action, where the right action means the action that causes the agent to be most successful in the given percept sequence. The problem the agent solves is characterized by Performance Measure, Environment, Actuators, and Sensors (PEAS).

Rationality at Any Given Time depends on the following four things:

1. **Performance Measure**: The criterion defining success.
2. **Agent's Prior Knowledge**: Understanding of the environment.
3. **Available Actions**: The actions the agent can perform.
4. **Percept Sequence**: The agent's historical sensory input.

1.4.5 Omniscience, Learning, and Autonomy in Rational Agents

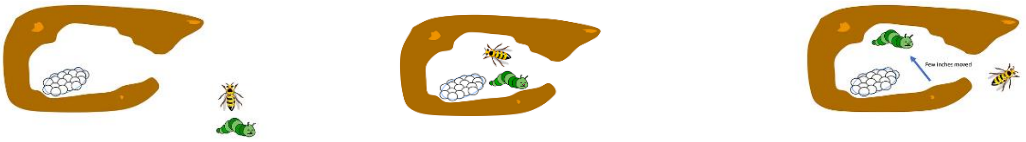
Omniscience and Rationality: Distinguishing between rationality and omniscience is crucial. An omniscient agent knows the actual outcome of its actions and can act accordingly; but omniscience is impossible in reality. While rationality maximizes expected performance.

Information Gathering : Rational agents should perform actions to modify future percepts, known as information gathering. This concept is vital for maximizing expected performance.

Learning in Rational Agents: Rational agents are not only expected to gather information but also to learn from their perceptions. An agent's initial configuration may reflect prior knowledge, but as it gains experience, learning becomes imperative.

*Think about the **dung beetle**, a small insect that carefully constructs its nest, lays eggs, and uses a ball of dung to seal the entrance. If the dung ball is taken away while the beetle is carrying it, the beetle continues its job as if the dung ball is still there, not realizing its missing. Evolution has shaped the beetle's behaviour to expect the dung ball, and any change from this expectation results in unsuccessful actions.*

*The **sphex wasp** is a bit smarter. It digs a hole, stings a caterpillar, drags it into the hole, checks everything is fine, and then lays eggs. The caterpillar becomes food for the hatching eggs. However, if someone moves the caterpillar a little while the wasp is checking, it goes back to dragging the caterpillar, not realizing the change. Even after many attempts to move the caterpillar, the wasp doesn't learn that its plan isn't working and keeps doing it the same way.*



Autonomy in Rational Agents: Autonomy is emphasized as a crucial aspect of rationality. An autonomous agent learns to compensate for **partial or incorrect prior knowledge**, ensuring effective adaptation to the environment. While acknowledging the practical need for some initial knowledge, over time, a rational agent's behavior becomes effectively independent of its initial knowledge through the incorporation of learning.

In essence, the combination of **rationality, information gathering, and learning enables the design of autonomous agents** capable of success across diverse environments.

1.4.4 The Nature of Environments

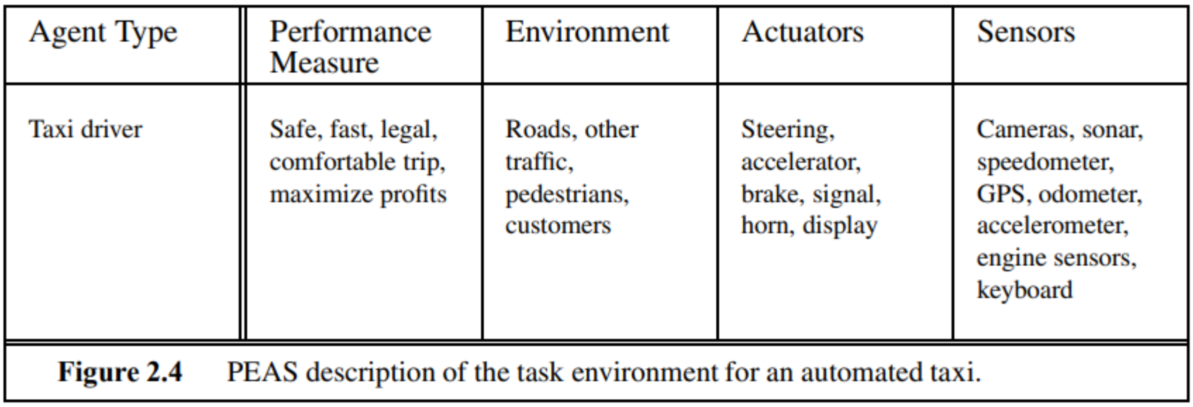
The task environment, in the context of rational agents, refers to the **external system or surroundings** within which an **agent** operates and performs tasks. It encompasses the conditions, challenges, and dynamics that influence the agent's behavior and decision-making processes. The task environment plays a pivotal role in shaping the **rationality, learning, and autonomy of an agent**.

Specifying the Task Environment

In the design of any agent, the initial crucial step is to comprehensively specify the "**task environment**," referred to as **PEAS** (Performance, Environment, Actuators, Sensors).

To illustrate the concept, consider the example of a more intricate problem: an **automated taxi driver**.

The PEAS description for the taxi's task environment is summarized in Figure 2.4, encompassing aspects such as performance measures, **driving environment, actuators, and sensors**.



Agent Type	Performance Measure	Environment	Actuators	Sensors
Taxi driver	Safe, fast, legal, comfortable trip, maximize profits	Roads, other traffic, pedestrians, customers	Steering, accelerator, brake, signal, horn, display	Cameras, sonar, speedometer, GPS, odometer, accelerometer, engine sensors, keyboard

Figure 2.4 PEAS description of the task environment for an automated taxi.

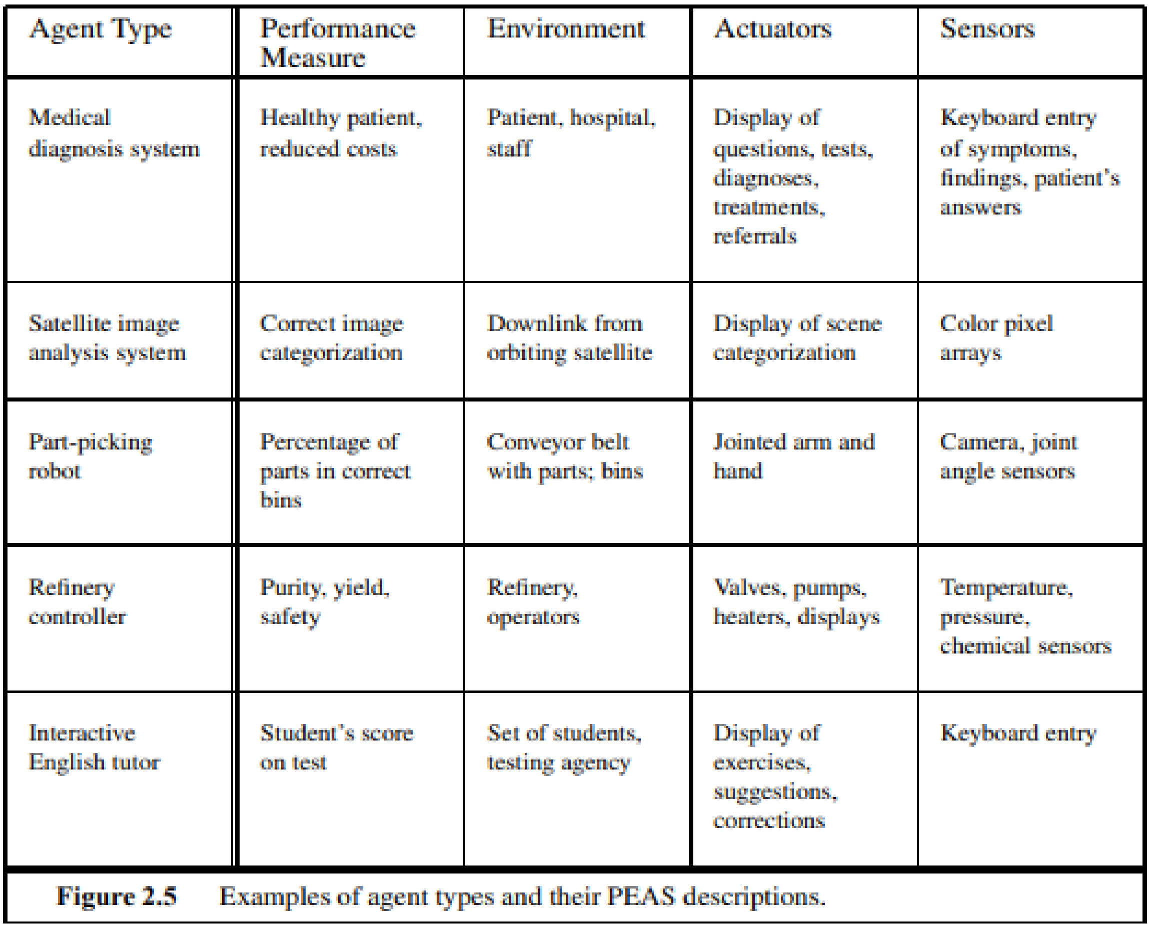
Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

Performance Measure: Determining the desired qualities for the automated driver, such as reaching the correct destination, minimizing fuel consumption, trip time, or cost, adhering to traffic laws, and maximizing safety and passenger comfort, involves addressing conflicting goals and necessitates trade-offs.

Driving Environment: The taxi must navigate a diverse range of roads, contend with various traffic conditions, interact with passengers, and make optional choices like operating in different regions with distinct weather conditions or driving on different sides of the road. The complexity of the environment directly impacts the design challenge.

Actuators and Sensors: Actuators for the automated taxi include control over the engine, steering, and braking, along with interfaces for communication with passengers and other vehicles. Sensors comprise controllable video cameras, infrared or sonar sensors, speedometer, accelerometer, and various vehicle state sensors.

Figure 2.5 outlines the basic PEAS elements for different agent types, showcasing the variety of considerations in specifying task environments.



Agent Type	Performance Measure	Environment	Actuators	Sensors
Medical diagnosis system	Healthy patient, reduced costs	Patient, hospital, staff	Display of questions, tests, diagnoses, treatments, referrals	Keyboard entry of symptoms, findings, patient's answers
Satellite image analysis system	Correct image categorization	Downlink from orbiting satellite	Display of scene categorization	Color pixel arrays
Part-picking robot	Percentage of parts in correct bins	Conveyor belt with parts; bins	Jointed arm and hand	Camera, joint angle sensors
Refinery controller	Purity, yield, safety	Refinery, operators	Valves, pumps, heaters, displays	Temperature, pressure, chemical sensors
Interactive English tutor	Student's score on test	Set of students, testing agency	Display of exercises, suggestions, corrections	Keyboard entry

Figure 2.5 Examples of agent types and their PEAS descriptions.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

1.4.5 Properties of Task Environments

In AI, there are various task environments, and we can classify them based on a few key dimensions as follows:

1. **Fully Observable vs. Partially Observable**:
   Fully Observable: The agent can see the complete environment all the time.
   Partially Observable: The agent's sensors may not capture all relevant aspects due to noise or missing information.

2. **Single Agent vs. Multiagent:**
   Single Agent: An agent operates independently, like solving a crossword puzzle.
   Multiagent: Agents interact with each other, introducing complexities in decision-making.

3. **Deterministic vs. Stochastic**:
   Deterministic: The environment's next state is entirely determined by the current state and the agent's action.
   Stochastic: Some uncertainty exists, making predictions challenging. For example, traffic behavior in taxi driving is stochastic.

4. **Episodic vs. Sequential:**
   Episodic: Each decision is independent, not influenced by past decisions.
   Sequential: Current decisions affect future ones, as seen in chess or taxi driving.

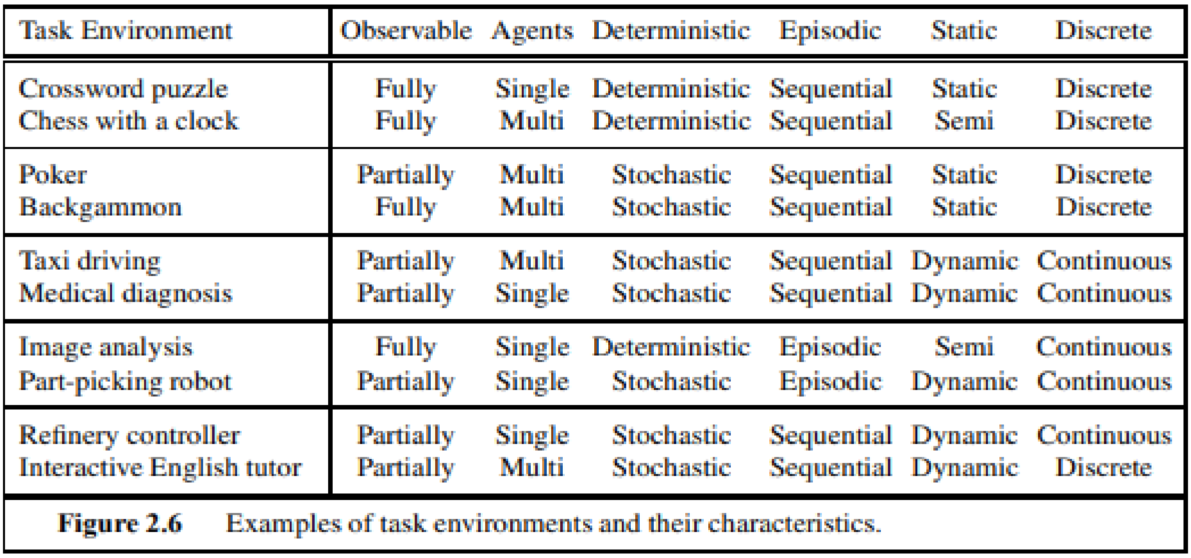
5. **Static vs. Dynamic**:
   Static: The environment remains unchanged during decision-making.
   Dynamic: The environment evolves continuously, demanding constant decision-making.

6. **Discrete vs. Continuous**:
   Discrete: Finite states, actions, or percepts.
   Continuous: Values or actions vary smoothly.

7. **Known vs. Unknown:**
   Known: The agent knows outcomes or probabilities.
   Unknown: The agent needs to learn about the environment.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

Understanding these dimensions helps in determining the complexity of an environment. For instance, *taxi driving is challenging due to being partially observable, multiagent, stochastic, sequential, dynamic, continuous, and unknown in many cases*.



Task Environment	Observable	Agents	Deterministic	Episodic	Static	Discrete
Crossword puzzle	Fully	Single	Deterministic	Sequential	Static	Discrete
Chess with a clock	Fully	Multi	Deterministic	Sequential	Semi	Discrete
Poker	Partially	Multi	Stochastic	Sequential	Static	Discrete
Backgammon	Fully	Multi	Stochastic	Sequential	Static	Discrete
Taxi driving	Partially	Multi	Stochastic	Sequential	Dynamic	Continuous
Medical diagnosis	Partially	Single	Stochastic	Sequential	Dynamic	Continuous
Image analysis	Fully	Single	Deterministic	Episodic	Semi	Continuous
Part-picking robot	Partially	Single	Stochastic	Episodic	Dynamic	Continuous
Refinery controller	Partially	Single	Stochastic	Sequential	Dynamic	Continuous
Interactive English tutor	Partially	Multi	Stochastic	Sequential	Dynamic	Discrete

Figure 2.6 Examples of task environments and their characteristics.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

Figure 2.6 lists the properties of a number of familiar environments.

1.4.6 The structure of agents

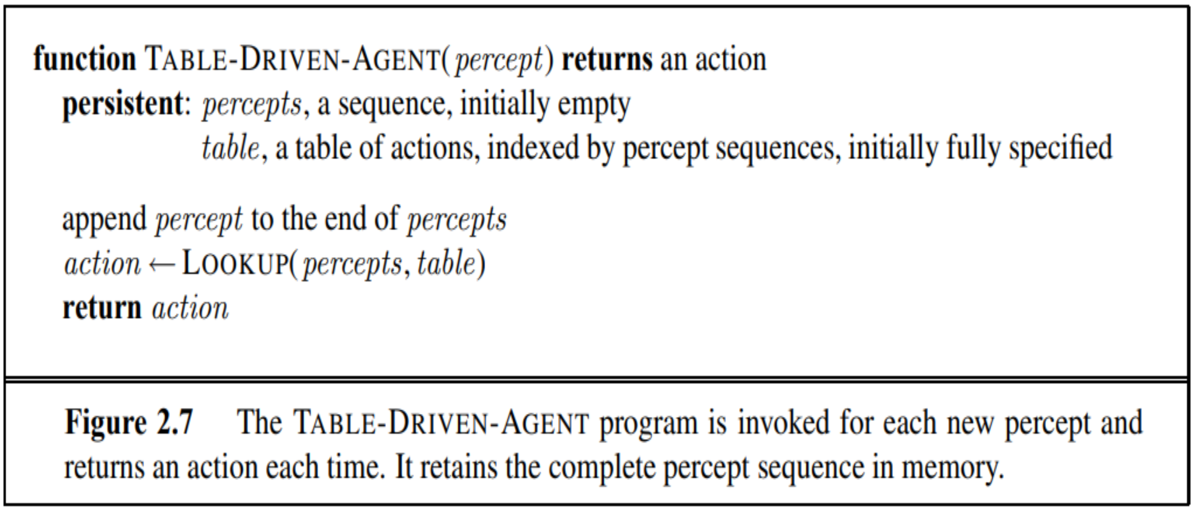
The structure of the agent comprises two essential components: the **architecture** and the **program**. The architecture refers to the combination of a computing device equipped with physical sensors and actuators. The program, created by AI, embodies the **agent function**, determining how the agent translates percepts into actions. The overall structure is denoted by the equation:

agent = architecture + program.

Agent Programs: All agent programs share a common structure: they take the current percept as input from the **sensors** and produce an action for the **actuators**.

It's crucial to note the distinction between the **agent program**, which operates based on the current percept, and the **agent function**, which considers the entire percept history. The agent programs are presented in a simple pseudocode language.

An example in **Fig 2.7** illustrates a basic agent program that tracks the percept sequence and uses it to determine the appropriate action from a table.



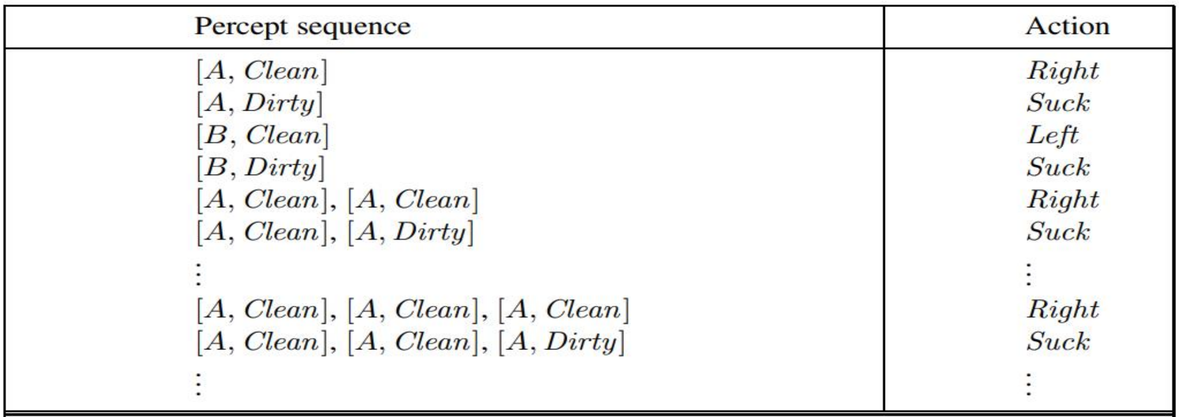
```
function TABLE-DRIVEN-AGENT(percept) returns an action
  persistent: percepts, a sequence, initially empty
              table, a table of actions, indexed by percept sequences, initially fully specified

  append percept to the end of percepts
  action ← LOOKUP(percepts, table)
  return action
```

Figure 2.7 The TABLE-DRIVEN-AGENT program is invoked for each new percept and returns an action each time. It retains the complete percept sequence in memory.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

The table, exemplified for the vacuum world in **Figure 2.3**, explicitly represents the **agent function embodied** by the agent program.



Percept sequence	Action
$[A, Clean]$	$Right$
$[A, Dirty]$	$Suck$
$[B, Clean]$	$Left$
$[B, Dirty]$	$Suck$
$[A, Clean], [A, Clean]$	$Right$
$[A, Clean], [A, Dirty]$	$Suck$
⋮	⋮
$[A, Clean], [A, Clean], [A, Clean]$	$Right$
$[A, Clean], [A, Clean], [A, Dirty]$	$Suck$
⋮	⋮

Figure 2.3 Partial tabulation of a simple agent function for the vacuum-cleaner world shown in Figure 2.2.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

Constructing a rational agent using this approach requires designers to create a table containing actions for every possible percept sequence. However, the table-driven approach faces insurmountable challenges. The size of the lookup table grows exponentially with the set of possible percepts and the agent's lifetime, making it impractical for complex environments. For instance, a lookup table for an automated taxi's visual input would be astronomically large. Despite these limitations, TABLE-DRIVEN-AGENT fulfils the objective of implementing the desired agent function.

Types of Agent Programs: There are, five basic types of agent programs as given below:

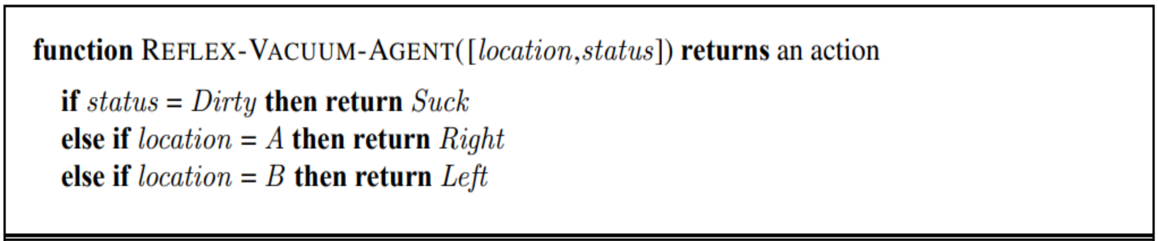
1. Simple reflex agents
2. Model-based reflex agents
3. Goal-based agents
4. Utility-based agents
5. Learning Agents

Each kind of agent program combines particular components in particular ways to generate actions.

1. **Simple Reflex Agents:**

The simplest kind of agent is the simple reflex agent. These agents select actions on the basis of the current percept, ignoring the rest of the percept history.

Example: For example, the vacuum agent whose agent function is tabulated in Figure 2.3 is a simple reflex agent, because its decision is *based only on the current location* and on whether that location contains dirt. An **agent program** for this agent is shown in Figure 2.8.



```
function REFLEX-VACUUM-AGENT([location,status]) returns an action
    if status = Dirty then return Suck
    else if location = A then return Right
    else if location = B then return Left
```

Figure 2.8 The agent program for a simple reflex agent in the two-state vacuum environment. This program implements the agent function tabulated in Figure 2.3.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

The program illustrated in Figure 2.8 is designed for a specific vacuum environment.

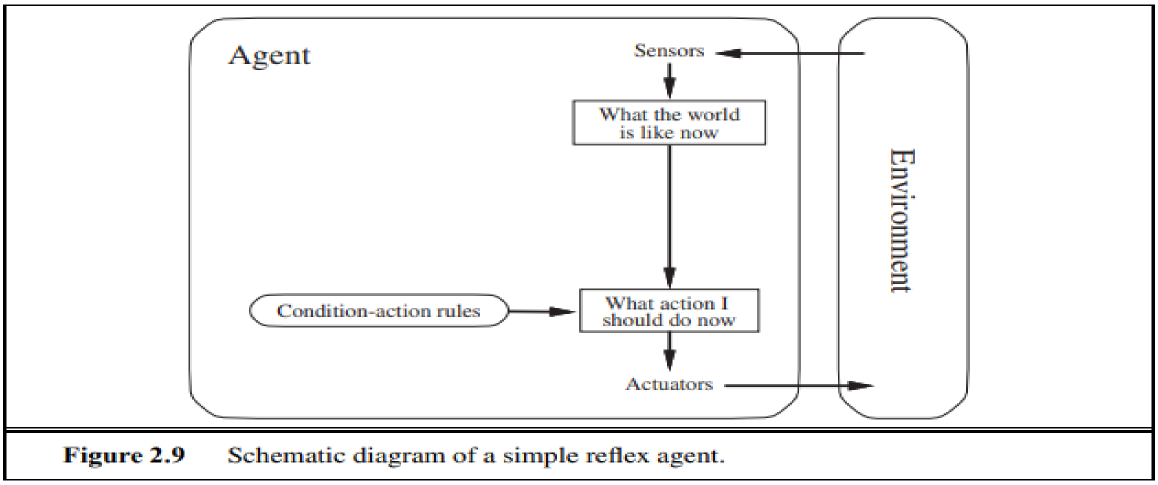
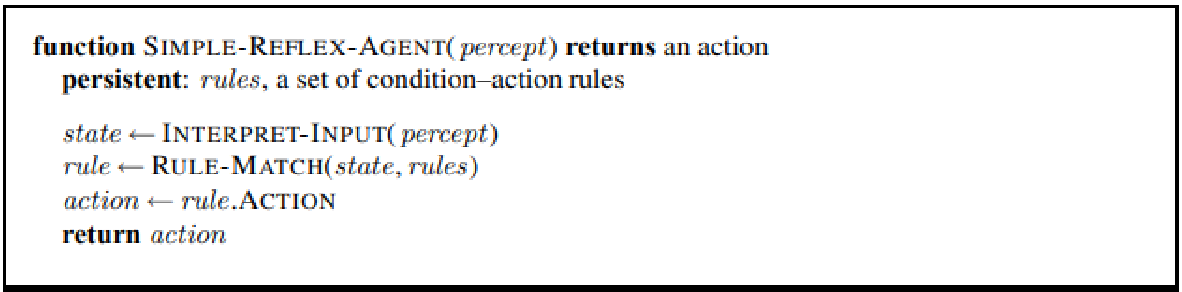


Figure 2.9 Schematic diagram of a simple reflex agent.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

Figure 2.9 provides a schematic representation of this generalized program, demonstrating how condition-action rules enable the agent to establish connections from perception to action. **Rectangles** represent the current internal state of the agent's decision process, while **ovals** depict background information used in the process.



```
function SIMPLE-REFLEX-AGENT(percept) returns an action
  persistent: rules, a set of condition–action rules

  state ← INTERPRET-INPUT(percept)
  rule ← RULE-MATCH(state, rules)
  action ← rule.ACTION
  return action
```

Figure 2.10 A simple reflex agent. It acts according to a rule whose condition matches the current state, as defined by the percept.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

The agent program, depicted in Figure 2.10, is also **straightforward**. The INTERPRET-INPUT function generates an abstracted description of the current state from the percept, and the RULE-MATCH function identifies the first matching rule in the rule set based on the given state description. Note that the conceptual terms **"rules" and "matching**" are used, but actual implementations can be as simple as a collection of logic gates constituting a Boolean circuit.

2. **Model based Reflex Agents:**

In order to address the challenges posed by partial observability, it is most effective for an agent to maintain an internal state that reflects the unobserved aspects of the current state. Updating this internal state requires encoding two types of knowledge in the **agent program**. **First**, knowledge about how the world evolves independently of the agent's actions, and **second**, information about how the agent's actions affect the world. This knowledge is referred to as a **"model of the world,"** and an agent using such a model is termed a model-based agent.

Figure 2.11 illustrates the structure of a model-based reflex agent with an internal state, where the current percept combines with the **old internal state** to generate an updated description of the current state based on the agent's model of the world.

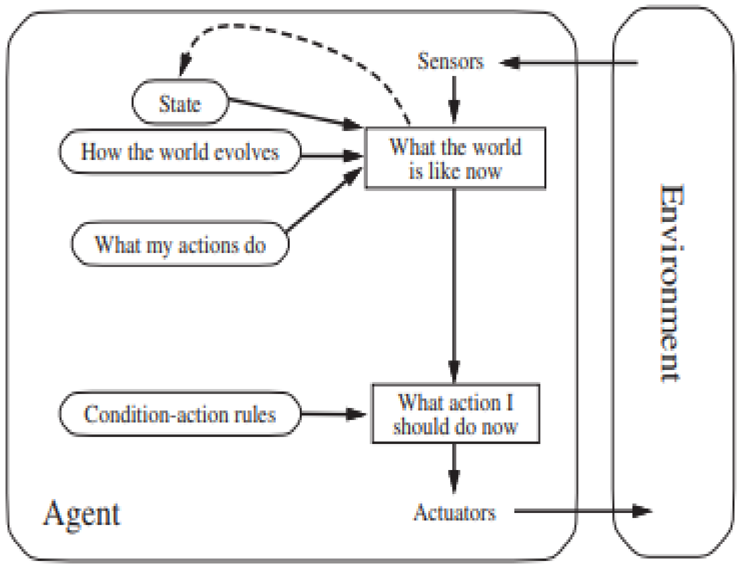
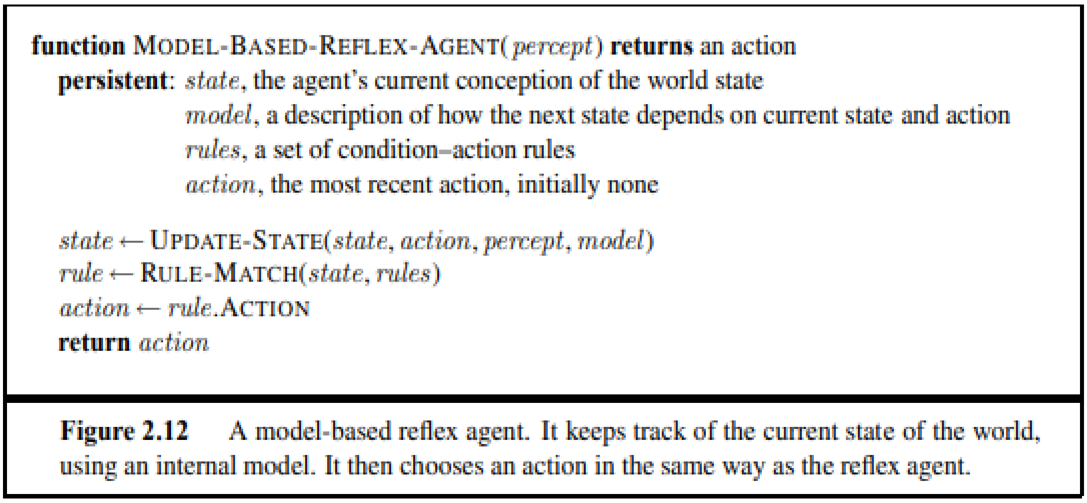


Figure 2.11 A model-based reflex agent.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

The agent program is shown in Figure 2.12. The UPDATE-STATE function is crucial in this process, responsible for creating the new internal state description.



```
function MODEL-BASED-REFLEX-AGENT(percept) returns an action
  persistent: state, the agent's current conception of the world state
              model, a description of how the next state depends on current state and action
              rules, a set of condition–action rules
              action, the most recent action, initially none

  state ← UPDATE-STATE(state, action, percept, model)
  rule ← RULE-MATCH(state, rules)
  action ← rule.ACTION
  return action
```

Figure 2.12 A model-based reflex agent. It keeps track of the current state of the world, using an internal model. It then chooses an action in the same way as the reflex agent.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

3. **Goal-based agents :**

Deciding on actions based solely on the current state of the environment may not always be sufficient. The agent's program combines **the goal information with its model**, similar to the model-based reflex agent, to select actions that align with the goal. Figure 2.13 illustrates the structure of a goal-based agent.

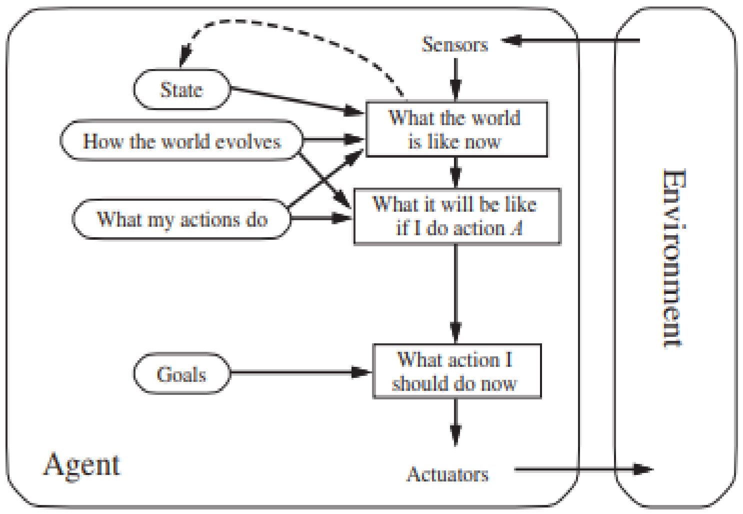

Sensors
State
How the world evolves
What my actions do
What the world is like now
What it will be like if I do action A
Goals
What action I should do now
Agent
Actuators
Environment

Figure 2.13 A model-based, goal-based agent. It keeps track of the world state as well as a set of goals it is trying to achieve, and chooses an action that will (eventually) lead to the achievement of its goals.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

4. **Utility based Agents**:

Achieving high-quality behaviour in various environments requires more than just setting goals. While goals indicate success or failure, they don't distinguish between different ways of achieving them in terms of efficiency, safety, reliability, or cost. A more detailed evaluation is needed, and this is where the concept of utility comes into play.

Utility, a term used by economists and computer scientists, represents a more refined measure of desirability than the binary distinction provided by goals. A utility function internalizes the performance measure, **assigning a score to different sequences of environmental state**s. If the internal utility function aligns with the external performance measure, an agent maximizing utility is considered rational. The utility-based agent structure is presented in Figure 2.14.

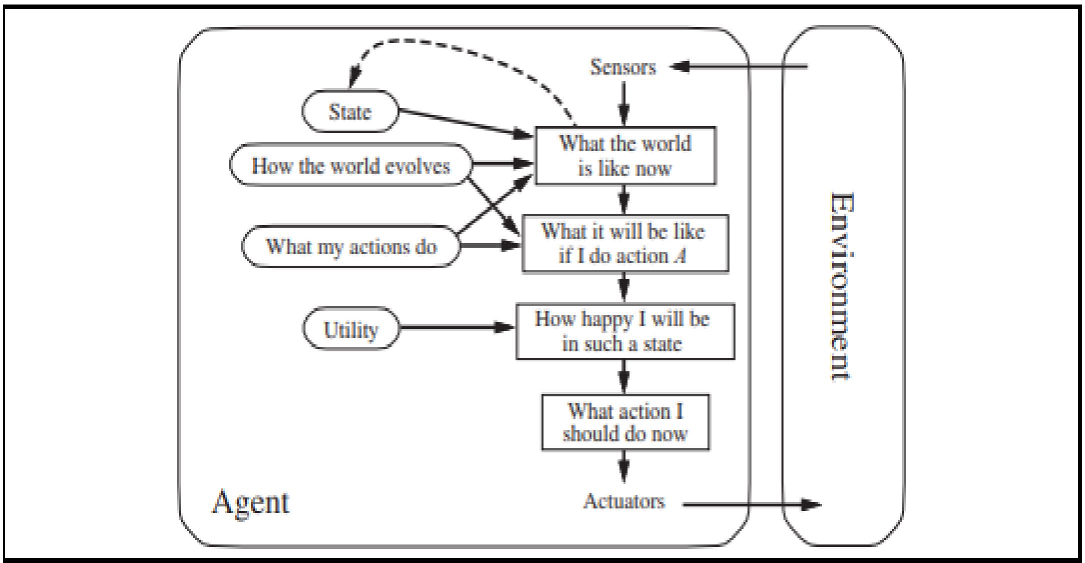

Sensors
State
What the world is like now
How the world evolves
What my actions do
What it will be like if I do action A
Utility
How happy I will be in such a state
What action I should do now
Agent
Actuators
Environment

Figure 2.14 A model-based, utility-based agent. It uses a model of the world, along with a utility function that measures its preferences among states of the world. Then it chooses the action that leads to the best expected utility, where expected utility is computed by averaging over all possible outcome states, weighted by the probability of the outcome.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015

5. Learning Agents

Learning offers the advantage of enabling agents to operate effectively in unfamiliar environments, surpassing their initial knowledge. A learning agent comprises **four conceptual components**, as illustrated in Figure 2.15. The crucial distinction lies between the **learning element**, responsible for improvements, and the **performance element**, tasked with selecting external actions. The learning element utilizes feedback from the **critic** to enhance the **performance element** for future improvement.

The critic informs the learning element of the agent's performance relative to a fixed standard. This standard is crucial, as percepts alone do not indicate the agent's success.

The problem generator, the last component, suggests actions for new and informative experiences. While the performance element tends to favor optimal actions based on current knowledge, the problem generator encourages exploration for potentially superior long-term actions.

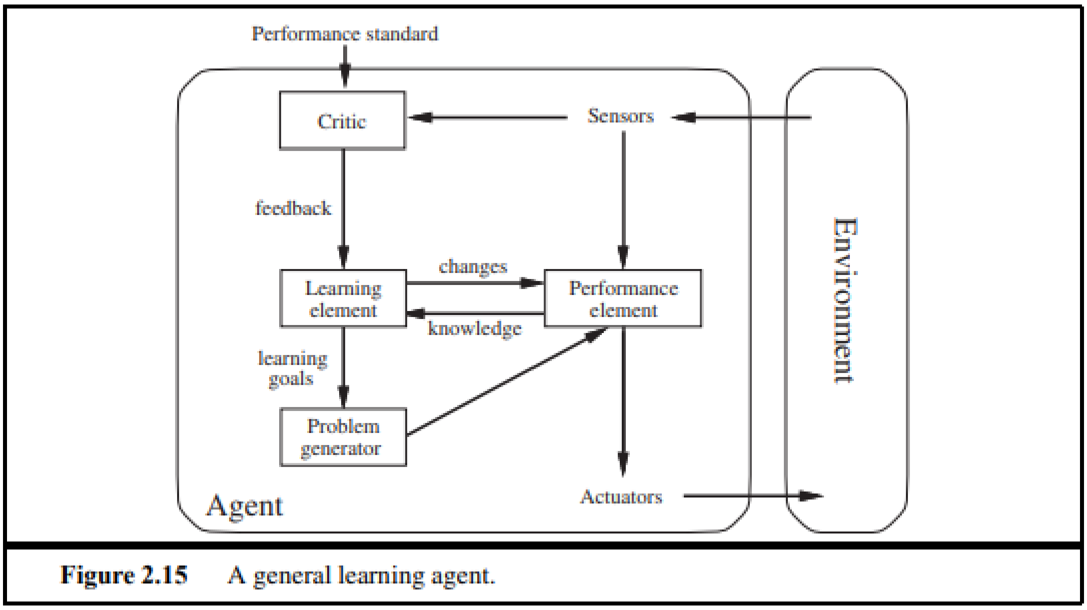


Figure 2.15 A general learning agent.

The automated taxi example: The performance element involves the knowledge and procedures guiding the taxi's driving actions, while the critic observes the world and provides feedback to the learning element. The learning element, in turn, formulates rules based on feedback, modifying the performance element accordingly. The problem generator suggests experiments to improve certain behaviors.

How the components of agent programs work

The representations of components of agents can be categorized along an axis of increasing complexity and expressive power: **atomic, factored, and structured**.

Atomic representations view each state of the world as indivisible, treating them as black boxes with no internal structure. This simplification is exemplified in algorithms like search, game-playing, Hidden Markov models, and Markov decision processes.

Factored representations, on the other hand, break down each state into a fixed set of variables or attributes, allowing shared attributes between different states. This facilitates dealing with uncertainty, as attributes can be left blank. Factored representations underlie various AI areas, including constraint satisfaction algorithms, propositional logic, planning, Bayesian networks, and machine learning algorithms.

Structured representations go beyond variables and values, explicitly describing objects and their relationships. This level of representation is essential for capturing complex scenarios, such as the interaction between a truck and a loose cow blocking its path. Structured representations are foundational to relational databases, first-order logic, first-order probability models, knowledge-based learning, and natural language understanding.

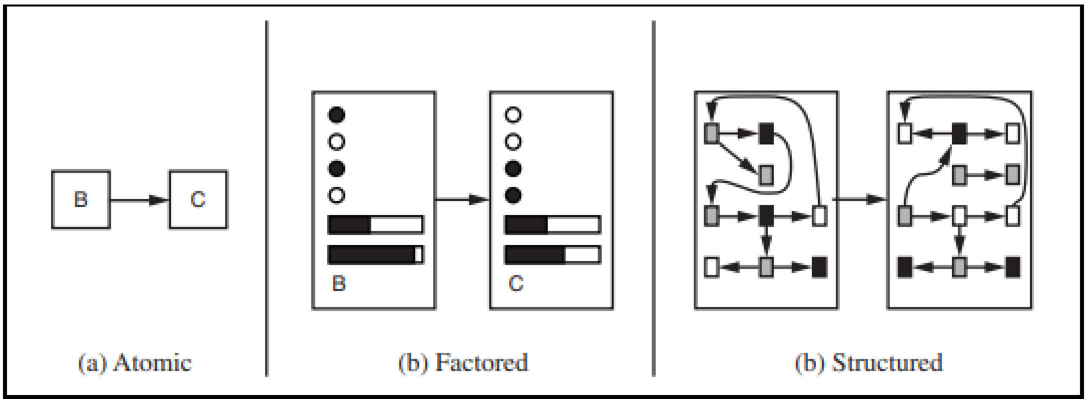


Figure 2.16 Three ways to represent states and the transitions between them. (a) Atomic representation: a state (such as B or C) is a black box with no internal structure; (b) Factored representation: a state consists of a vector of attribute values; values can be Boolean, real-valued, or one of a fixed set of symbols. (c) Structured representation: a state includes objects, each of which may have attributes of its own as well as relationships to other objects.

Expressiveness is the key axis along which these representations lie, with increasing expressiveness allowing more concise representation of information. However, as expressiveness grows, reasoning and learning become more complex. Balancing the benefits and drawbacks of expressive representations, real-world intelligent systems may need to operate at various points along this axis simultaneously to effectively handle diverse scenarios.

Source Book: Stuart J. Russell and Peter Norvig, Artificial Intelligence, 3rd Edition, Pearson,2015