

Lecture 1: Intelligent Agents

10:30 AM, Jan 22, 2009

Contents

1	What is AI?	1
1.1	Acting Like Humans	2
1.2	Thinking Like Humans	2
1.3	Thinking Logically	3
1.4	Acting Rationally	3
2	Subfields of AI	4
3	AI Systems	5
4	Other Definitions of AI	6
5	What if we succeed?	6

1 What is AI?

For centuries, philosophers have tackled such profound questions as: “What is thought?” “What is understanding?” “How do we represent our knowledge of the world?” and “How do we reason about that knowledge?” Only this past century, researchers in the burgeoning field of **artificial intelligence** have attempted to go further, seeking not just to explain thought, but to mechanically simulate understanding and reasoning, as well as perception and action.

The term artificial intelligence was born in 1956 at a two-month summer workshop held at Dartmouth College. Attendees included John McCarthy (inventor of LiSP), Marvin Minsky (pioneer in the area of neural networks), Arthur Samuel (designer of the first computer Checkers player, a program that learned during play and eventually beat its designer), Allen Newell (Turing award winner for his work on cognitive architecture), and Herb Simon (economist and Nobel laureate for his theory of satisficing, a decision-making heuristic).

Artificial intelligence is abbreviated AI. The primary goal of **weak** AI is to *build* intelligent entities. A related (but not a necessary) goal is to *understand* intelligent entities, and perhaps even to understand and engineer human intelligence (**strong** AI). But what are *intelligent entities*? Russell and Norvig¹ put forth four possible definitions, the second and third of which might be categorized as strong AI, and the first and fourth of which are more along the lines of weak AI:

1. systems that act like humans

¹These lecture notes are primarily based on Chapters 1 and 2 of Russell and Norvig [2].

2. systems that think like humans
3. systems that think logically
4. systems that act rationally: i.e., do the right thing, given their knowledge and/or beliefs

If we adopt either of the first two definitions, the human-centered approaches, then to verify that the systems we build are intelligent would require empirical methods. Hypotheses about how humans act and think in various situations would be generated, and then tested in laboratory experiments with human subjects. If we adopt either of the last two definitions, the logical and rational approaches, then to verify that the systems we build are intelligent would require techniques of mathematics and engineering.

1.1 Acting Like Humans

In 1950, Alan Turing proposed an operational definition (rather than a list of sufficient and necessary conditions) of intelligence, later named the Turing Test. An entity is deemed intelligent, according to Turing, if it demonstrates an ability to achieve human-level performance in all cognitive tasks, sufficient to fool an interrogator. Turing predicted that by the year 2000, computers would be able to pass the Turing test.²

ELIZA, an early AI system intended as a therapist, fails the Turing Test.

```
PERSON: My wife says I'm often depressed.
ELIZA: I'm sorry to hear you are depressed.
PERSON: The truth is I am quite unhappy.
ELIZA: Do you think coming here will help you not to be unhappy?
PERSON: Well, perhaps by talking to you I will find out why I
        hate my brother.
ELIZA: Tell me more about your family.
```

Like the word `brother` in this example, a sentence such as `I admire Mother Teresa` triggers the response `Tell me more about your family`.

Designing a system that can simulate human behavior (i.e., pass the Turing test) is not the holy grail of AI research, just as simulating the flight of birds was not the overarching goal driving the work of the Wright brothers.

1.2 Thinking Like Humans

In 1963, Allen Newell and Herb Simon designed the General Problem Solver (GPS), which was intended to be a program that simulated human thought. The name GPS derived from the program's architecture, which distinguished between general knowledge about reasoning and knowledge about specific problem domains. GPS used means-end analysis in its search for solutions.

Quoting Aristotle,

²More specifically, he predicted that the average interrogator would not be able to distinguish a computer from a human more than 70 per cent of the time, after a five minute conversation.

We deliberate not about ends, but about means. For a doctor does not deliberate whether he shall heal, nor an orator whether he shall persuade . . . They assume the end and consider how and by what means it is attained, and if it seems easily and best produced thereby; while if it is achieved by one means only they consider *how* it will be achieved by this and by what means *this* will be achieved, till they come to the first cause, . . . and what is last in the order of analysis is first in the order of becoming.

In today's parlance, means-end analysis assumes an initial state and a goal state, computes the difference between them, and if that difference is not zero, searches for ways to minimize it by replacing the goal with a sufficient set of subgoals, and then recurring.

By comparing GPS traces with those of human subjects, Newell and Simon discovered that the behavior of GPS was largely a subset of human behavior. Today, the study of human cognition characterizes the field of cognitive science, rather than AI.

1.3 Thinking Logically

The *Laws of Thought* approach to AI relies on patterns for argument structure rooted in Aristotle's syllogisms (*e.g.*, All men are mortal; Socrates is a man; therefore, Socrates is mortal). In the late 1800's and early 1900's, the formal logic movement was advanced by Guiseppe Peano, George Boole, Gottlob Frege, Alfred Tarski, Kurt Gödel, and others. Perhaps inspired by early progress, David Hilbert became a proponent of a school of thought known as *logicism*, or *formalism*. The goal of this program was to devise a logic, or formal system, capable of deriving all mathematical theorems, thereby uncovering all possible mathematical intuitions. Ultimately, Gödel's Incompleteness Theorem (1931), which states that in sufficiently powerful languages there are unprovable truths, served to dismantle the logicist/formalist program.

1.4 Acting Rationally

Modern AI can be characterized as the engineering of *rational agents*. An *agent* is an entity that (i) perceives, (ii) reasons, and (iii) acts. In computational terms, that which is perceived is an *input*; to reason is to *compute*; to act is to *output* the result of computation. Typically, an agent is equipped with objectives. A *rational* agent is one that acts optimally with respect to its objectives.

Agents are often distinguished from typical computational processes by their *autonomy*. They operate without direct human intervention. In addition, agents are *reactive*—they perceive their environments, and attempt to respond in a timely manner to possibly changing conditions—and *proactive*—their behavior is goal-directed, rather than simply response-driven.

Agent	Sensors	Actuators
Human	Senses	Arms, Legs, etc.
Robotic	Cameras	Motors, Wheels, etc.
Software	Bit Strings	Bit Strings

Table 1: Examples of Agents.

Examples of agents include human agents, robotic agents, and software agents. (See Table 1.)

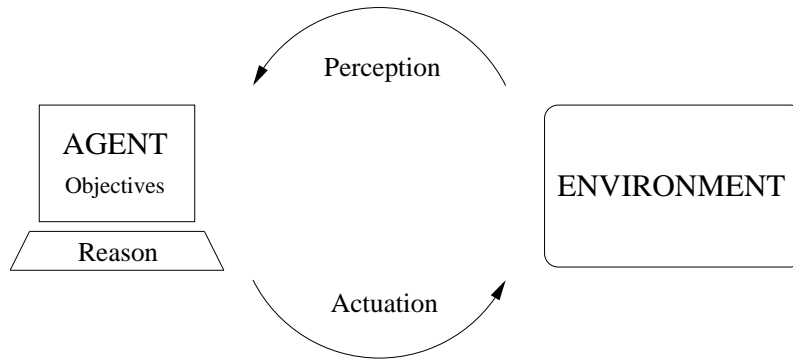


Figure 1: Intelligent Agents = Perception + Reason + Actuation.

Autonomous agents may be *rule-based*, *goal-based*, or *utility-based*. Rule-based agents operate according to hard-coded sets of rules, like ELIZA. A goal-based agent acts so as to achieve its goals, by planning a path from its current state to a goal state, like GPS or theorem provers. Utility-based agents distinguish between multiple goals based on their respective utilities.

Some dimensions of agent *environments* include:

- Deterministic vs. Nondeterministic: is the next state predictable (*e.g.*, chess), or is there uncertainty about state transitions (*e.g.*, backgammon)?
- Discrete vs. Continuous: can the environment be described in discrete terms (*e.g.*, chess), or is the environment continuous (*e.g.*, driving)?
- Static vs. Dynamic: is the environment static (*e.g.*, chess), or can it change while the agent is reasoning about its plan of action (*e.g.*, driving)?
- Sequential vs. One-shot: does the agent need to reason about the future impact of its immediate actions (*e.g.*, chess), or can it treat each action independently (*e.g.*, *Rochambeau*)?
- Single agent vs. Multiagent: can we assume the agent operating alone in its environment, or need it explicitly reason about the actions of other agents (*e.g.*, chess, backgammon, *Rochambeau*, driving)?

2 Subfields of AI

What is required of a machine if it is to pass the Turing test? At a minimum, problem-solving skills such as automated reasoning, knowledge representation, and machine learning. To pass the total Turing test, perception (machine vision and/or natural language processing) and actuation (robotics) are also required. The subfields of artificial intelligence can be classified in terms of their role in either perception, problem solving, or actuation.

- Perception
 - machine vision
 - natural language processing

- Problem solving: mapping from percepts to actuators
 - automated reasoning
 - knowledge representation
 - decision/game theory
 - machine learning
- Actuation
 - robotics
 - softbotics

3 AI Systems

Some important examples of AI systems include the following, described in terms of their mechanisms for perception, reason, and actuation.

- Xavier, the mail delivery robot, developed at CMU
 - Perception: vision, sonar, web interface
 - Reason: A* search, Bayes classification, hidden Markov models
 - Actuation: wheeled robotic actuation
- Pathfinder, the medical diagnosis system, developed by Heckerman and other Microsoft researchers
 - Perception: input symptoms and test results
 - Reason: Bayesian networks, Monte-Carlo simulations
 - Actuation: output diagnoses and further test suggestions
- TDGammon, the world champion backgammon player, built by Gerry Tesauro of IBM Research
 - Perception: keyboard input
 - Reason: reinforcement learning, neural networks
 - Actuation: graphical output shows dice and movement of pieces
- ALVINN, the automated driver, developed by Pomerleau at CMU
 - Perception: video camera
 - Reason: neural networks and hand-engineered solutions
 - Actuation: land vehicle controller to turn the steering wheel
- PROVERB, a world class crossword puzzle solver, developed by Littman and his students at Duke University

- Perception: grid, clues, background databases
- Reason: belief net inference and “turbo decoding”
- Actuation: filling in the grid puzzle

4 Other Definitions of AI

AI is the business of getting computers to do things they cannot already do, or things they can only do in movies and science fiction stories.

AI is the enterprise of constructing an intelligent artifact ([?]).

AI is the design of flexible programs that respond productively in situations that were not specifically anticipated by the designer ([1]).

AI is the construction of computations that perceive, reason, and act effectively in uncertain environments. In this definition, the psychological aspects of AI are perception, reason, and action, and the “construction of computations” encompasses the computer science aspect of AI ([3]).

The goal of AI is to build autonomous machines (i.e., machines that function without human intervention), but at the same time replicate human faculties (e.g., see, hear, speak) in complex, dynamic environments. [Adapted from Russell and Norvig’s discussion of the history of AI.]

5 What if we succeed?

Here’s what Woody Allen has to say: “My father lost his job because his plant bought a machine that is capable of doing everything my father could do . . . it wasn’t so bad, until my mother went out and bought one as well.”

References

- [1] T. Dean, J. Allen, and Y. Aloimonos. *Artificial Intelligence: Theory and Practice*. Addison-Wesley, 1995.
- [2] S. Russell and P. Norvig. *Artificial Intelligence: A Modern Approach*. Prentice-Hall, 1995.
- [3] P. Winston. *Artificial Intelligence*. Addison-Wesley, 1992.