www.cs.brown.edu/courses/cs141/lectures.html

CS141: Introduction to AI Spring 2001

Professor Amy Greenwald Lecture #1

Intelligent Agents[1]

1 What is AI?

The primary goal of (weak) artificial intelligence 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 Norvig put forth four possible definitions, the first and second of which might be categorized as strong AI, and the third and fourth of which are more along the lines of weak AI:

1. systems that act like humans
2. systems that think like humans
3. systems that think rationally, according to logical laws
4. systems that act rationally, or *do the right thing*

1.1 Acting Like Humans

In 1950, Turing proposed an operational definition 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.

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.`

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

1.2 Thinking Like Humans

In 1963, Newell and 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 specific domain knowledge. GPS used means-end analysis in its search for solutions, computing the difference between the goal and the current state, and then attempting to minimize that difference. 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.

1.3 Thinking Rationally

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 Peano, Boole, Frege, Tarski, Gödel, and others. Perhaps inspired by early progress, 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. Gödel's Incompleteness Theorem (1931), which states that there are unprovable truths, and therefore implies that the search for proofs need not terminate, served to discredit 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 are *inputs*; to reason is to *compute*; and to act is to *output* the result of computation. Typically, an agent is equipped with *goals*, and a *rational* agent is one that acts so as to achieve its goals.

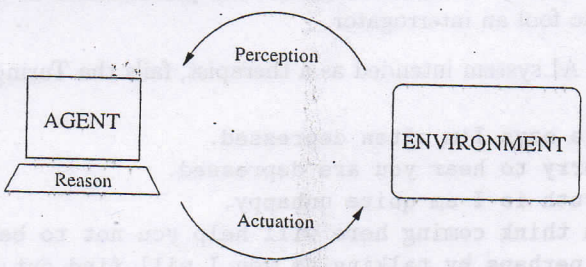

Figure 1: Intelligent Agents = Perception + Reason + Actuation.

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 changing conditions—and *proactive*—their behavior is goal-directed, rather than simply environmental responses. Lastly, agents may have social abilities that guide their interactions with other agents (and humans).

2 Subfields of AI

The subfields of artificial intelligence can be classified in terms of their role in either perception, reasoning, or actuation.

- Perception
  - computer vision
  - speech recognition
  - natural language processing
- Reasoning (*i.e.*, problem solving): mapping from percepts to actuators
  - knowledge representation
  - search and optimization
  - machine learning
  - decision theory
  - game theory
  - planning
- Actuation
  - robotics

2.1 Examples of 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
  - 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

3 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 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. The "construction of computations" in this definition encompasses the computer science aspect of AI; the psychological aspects of AI are perception, reason, and action ([3]).

4 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.