

1 Introduction

The term *artificial intelligence* (AI) is used to describe machines that demonstrate intelligent behavior. Often AI refers to computers or robots that carry out tasks or solve problems in a way that resembles intelligent human behavior. But an AI does not necessarily mimic the human cognitive system — it could be constructed and operate in a completely different manner, as long as it is in some sense *intelligent*.

Even though computers easily outperform humans on many tasks, not all such tasks require *intelligence*. For example, most of us have probably memorized the value of π to at least three decimal points of precision (it is about 3.14 as you may recall), but the world record holder managed to memorize more than 100 000 decimals. Even though this is a impressive and remarkable achievement for a human being, it can be debated how much intelligence it requires. A computer can easily compute and memorize an almost infinite number of decimals, limited only by the size of its storage capacity. Simply outperforming humans on a difficult task is not enough to label the computer program as intelligent.

1.1 Intelligence

To approach the question of what artificial intelligence is, perhaps we should start by defining more precisely what we mean by the word *intelligence* itself. This might not be so easy, since intelligence is a term which can mean different things to different people.

1.1.1 *Some requisites for defining intelligence*

Some of the aspects of intelligence that we might identify could include the ability to interact, learn from experience, and achieve goals:

Interact It is difficult to imagine an intelligent creature without the ability to interact with the physical world. It seems reasonable to say that a requisite for intelligence is to perceive the world and take actions which affect the world.

There is no such thing as a disembodied mind. The mind is implanted in the brain, and the brain is implanted in the body.

—Antonio Damasio

According to the theory of *embodied cognition* human intelligence is strongly influenced by our bodies, and shaped by our abilities to perceive and act.

Learn from experience It is hard to imagine a *static* intelligent creature: The ability to learn and adapt one's behavior to the environment is also a reasonable requisite for defining intelligent behavior. Related to learning are concepts such as forming mental representations of percepts and mental models of the world, as well as remembering and being able to generalize from past events.

Achieve goals Finally, a reasonable requisite of intelligence is to have certain goals and act in a way that helps achieve them. It is difficult to imagine an intelligent creature without purpose, that does not somehow act to obtain some desired outcome. In this sense the creature must be *rational* and take actions that it believes will accomplish its goals.

1.1.2 *A criterion of mind*

In one view, intelligence in both humans and animals has evolved through natural selection, and acts as a mechanism that allows an organism to adapt individually to its environment. George John Romanes defines a *criterion of mind* that indicates intelligent behavior in an organism:

Does the organism learn to make new adjustments, or to modify old ones, in accordance with the results of its own individual experience?

—George John Romanes (Animal Intelligence, 1882)

This should be seen in opposition to *instincts* which also involve mental operations that are adapted to the environment in pursuance of a goal, but which do not rely on individual experience and acquired knowledge. Romanes notes that it is impossible to draw a clear line between instinct and reason, and that different levels of intelligent behavior is better described as a continuum. However, in an attempt to make such a distinction, Romanes writes:

For reason, involving a mental constituent, and besides being concerned in adaptive action, is always subsequent to individual experience, never acts but upon a definite and often laboriously acquired knowledge of the relation between means and ends, and is very far from being always similarly performed under the same appropriate circumstances by all the individuals of the same species.

—George John Romanes (*Animal Intelligence*, 1882)

According to this definition, intelligence (or *reason* as Romanes writes) is to take actions adapted to the environment to achieve goals based on individual learned experience.

1.1.3 A working definition of intelligence

Based on the ideas presented so far, we could formulate the following definition of what we mean by *intelligence*:

Definition 1.1 Intelligence

The ability to learn and apply individual knowledge and skills to achieve goals.

While this definition captures some important aspects of the concept of intelligence, it is clearly also very limited. Within this definition, perhaps even some plants might be considered to be intelligent? Intelligence could also be defined to include other, more advanced aspects, such as logical reasoning, social and emotional aptitude, creativity, as well as self-awareness and consciousness. According to the definition above, an intelligent agent could easily be an uninventive, socially inept, illogical, nonconscious creature—but the simple definition might be a good starting point, if we want to attempt to implement intelligence in a machine.

1.1.4 Artificial general intelligence

We can imagine that one day it might be possible to create a machine that can perform all task that humans can perform. Philosophers refer to such imagined systems as *artificial general intelligence* or *strong AI*. In contrast, we might say that all artificial intelligence systems that have been created so far are *artificial specific intelligence* or *weak AI* because they exhibit intelligent behavior only for a specific problem within a limited set of conditions.

Even though no one have yet created a strong AI system, it might be worth to consider the consequences it would have for humanity. Scientists are approaching the creation of strong AI from two directions: One approach is the continuing effort to build more and more advanced weak AI systems, and expand their capabilities to growing domains of applicaions. Another approach is to replicate a biological brain in the form of a computer system, so that its behavior can be simulated in full scale.

On the one hand, we might be afraid that a strong AI would render humans worthless and take over the world:

The development of full artificial intelligence could spell the end of the human race. [...] It would take off on its own, and redesign itself at an ever-increasing rate. Humans, who are limited by slow biological evolution, couldn't compete and would be superseded.

—Stephen Hawking (2014)

On the other hand, presumably humans would be in control of the AI—and if things go in the wrong direction,

[...] you just have to have somebody close to the power cord. Right when you see it about to happen, you gotta yank that electricity out of the wall, man.

—Barack Obama (2016)

1.2 Learning

It is not easy to define precisely which tasks require intelligence to solve, and which tasks a simple computer program can manage, as the following pithy quotation from John von Neumann points out:

If you will tell me precisely what it is that a machine cannot do, then I can always make a machine which will do just that!

—John von Neumann (1950)

This point illustrates that learning and adapting one's behavior to new and unforeseen circumstances is a key element of intelligence. If we are able to fully describe a problem, with no detail left vague or undefined, it should be possible to instruct a computer program to solve the problem. One of the important differences between humans and computers is that while humans can be quick to learn from their experience and modify their behavior, computers can basically only do exactly what they are programmed to do. So creating an artificial intelligence seems to require us to program a computer to *learn*.

Learning is the process of acquiring knowledge or skills that can later be put to use. The learning process comes in many forms: We can individually study the world and look for recognizable patterns, we can be instructed or shown examples, or we can interact and experiment with the world around us.

Definition 1.2 Learning

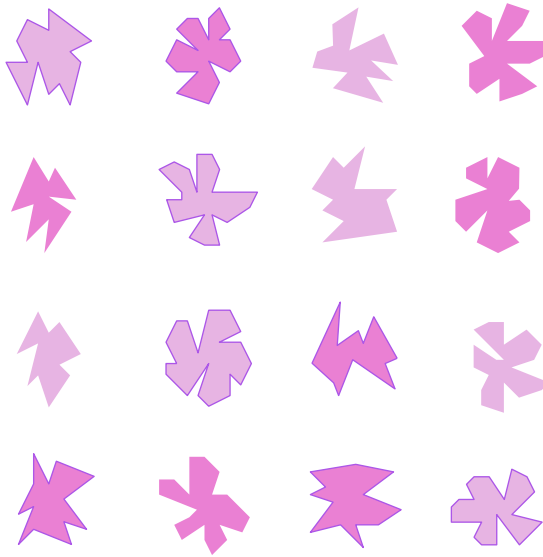
To acquire knowledge or skills by study, instruction, or experience.

1.2.1 Categorization

One way to understand and differentiate between different things in the world is through *categorization*. When we observe and interact with the world around us, we come to realize that some things are similar and some things are different in terms of their properties, such as how they look and behave, and what they can do for us. We tend to organize our understanding of the world by forming categories.

Example 1.1 Two kinds of purple objects

Imagine a world that consists of purple objects that all look fairly similar. While all objects are different, some are similar to each other in various ways, and so we can start to categorize them into different kinds. If you look at the following 16 examples of purple objects, can you tell me how many different kinds there are?



If you examine the objects carefully you will likely discover, that there are many different ways they can be categorized. They appear to come in different shades of purple, some are outlined, and the shapes also comes in different varieties.

Learning by categorization is an example of unsupervised learning: There is no teacher that tells us what to look for, and what the correct answer is. All we can do is observe the world around us, and start to look for patterns and regularities that help us structure it.

1.2.2 *Learning by example*

Another way to learn is *by example*. By looking at examples of objects we can learn to recognize them, and learn to generalize our knowledge by figuring out which properties and features are important.

Example 1.2 Kiptic or spurgle?

If I ask you if this unknown purple object is a *kiptic* or a *spurgle*, you probably wouldn't know what to answer.



In a way, you are seeing the object in the same way as a computer would: With no prior information to put the observation into context.

In order to answer the question, you need to know more about kiptics and spurgles. One way to get that information is to simply show you some labelled examples of kiptics and spurgles. Once you have seen a few examples of each type of object, you can use that information to reason about the purple object above.

Take a look at the examples in Figure 1.3. Can you now determine what type of object we are dealing with here?

Most people would agree that the *kiptic or spurgle* problem requires intelligence to solve. By looking at the labelled examples you probably formed a mental representation of the two types of objects. Then, you somehow compared the unknown purple object to this mental representation in order to classify it as a kiptic or a spurgle.

1.2.3 The perception-action cycle

We might see learning as a process that transfers external information into the mind of the learner, where it is represented in such a way that it can be accessed and utilized. In this view, learning is a flow from an external environment into the mind of the learner. However, perhaps this view on learning is too limited? Another view is that learning takes place as an interaction between the learner and the environment. In general, humans do not learn simply by being told—we learn by doing. This view of learning as a continuous circular flow of information that occurs when the learner interacts with her environment is known as the *perception-action cycle*.

In the perception-action cycle (see Fig. 1.2) the learner has a mental representation of how the world works. When presented with a new situation, the mental model guides the learner to choose an appropriate action. By performing an action, the learner interacts with the world, which results in some outcome. The learner perceives the outcome and uses it to update her mental representation if necessary. This feedback can be positive or negative: If the world behaves as she expected, the learner can use this information to strengthen her existing beliefs, but if the outcome was unexpected, the learner must more

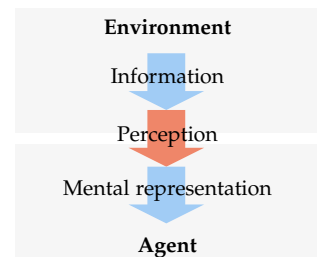


Figure 1.1: Simple learning

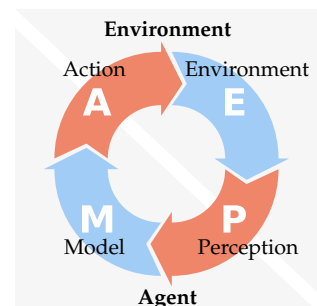


Figure 1.2: The perception-action cycle

fundamentally adjust or reorganize her world model.

1.3 Reasoning

Where *learning* is the process of acquiring knowledge and skills, *reasoning* is the process of manipulating the knowledge and adapting our skills to answer new questions and solve new problems. Reasoning is closely related the the concept of *rationality*: What we believe and how we act should be consistent with the information we have available and our own goals and objectives.

We can distinguish between three different modes of reasoning, called deduction, induction, and abduction.

Definition 1.3 Deduction, induction, and abduction

- Deduction* Reasoning from general premises to specific conclusions.
- Induction* Reasoning from specific facts to general rules.
- Abduction* Reasoning about the simplest or most likely explanation.

Example 1.3 A long line at the cafeteria

Deduction If we presume that there is always a long line at the cafeteria at noon, and we look at our watch to see that it is 12 o'clock, we can deduce that there is now a long line at the cafeteria. Given that the premise is true, the conclusion must also hold purely by logic; however, if there is a flaw in the premise, the conclusion might not be true.

Induction If we every day observe a long line in the cafeteria at noon, we might induce that it is a general rule that there is always a long line around at 12 o'clock. If, however, we one day go to the cafeteria at noon and there is no line, such a counterexample completely disproves the general rule.

Abduction If we happen to pass by the cafeteria and see that there is a long line, we might abduce that it is lunchtime. Since we know that most people tend to

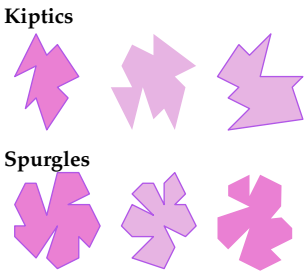


Figure 1.3: Some examples of kiptics and spurgles.

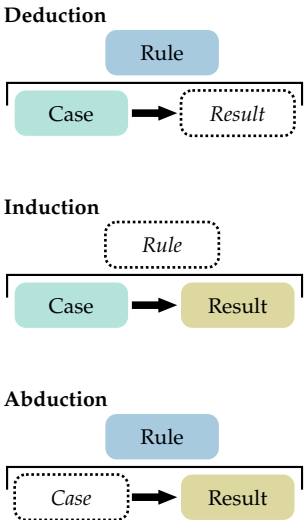



Figure 1.4: Deduction is reasoning from general rules to specific facts. Induction is reasoning from specific facts (observed cases and their results) to general rules. Abduction is reasoning about the most likely explanation.



go to the cafeteria around noon, it is the simplest and most likely explanation for the long line; however, our abduction might be wrong if there is a flaw in our explanation—maybe the cafeteria are giving out free sandwiches today at 10 o'clock or something.

We can distinguish between *intuitive* and *formal* reasoning. Intuitive reasoning is based on instincts, feelings, and unconscious knowledge, whereas formal reasoning is based strictly on unambiguous rules such as logic or mathematics. In practice, most human reasoning probably lies somewhere on the spectrum between these two extremes.

Problems

1. Come up with your own definition of the terms *intelligence* and *artificial intelligence*. Which aspects of these concepts do you think are most important?
2. Do you believe it is possible to create a thinking machine?
3. What is the most significant and profound example of the following topics you can think of?

<i>Superhuman AI</i>	Artificial intelligence that outperforms humans.
<i>Creative AI</i>	Artificial intelligence that emulates human creativity.
<i>Animals intelligence</i>	Intelligent behavior in animals (or perhaps plants).
<i>Augmented intelligence</i>	Enhancing human performance using artificial intelligence.

4. Is artificial intelligence strictly based on *formal reasoning* or can an AI be said to have intuition?