III. Philosophy of AI

The very nature of the term “artificial intelligence” brings up philosophical questions whether intelligent behavior implies or requires the existence of a mind, and to what extent is consciousness replicable as computation.

The Turing test

Alan Turing (1912-1954) was an English mathematician and logician. He is rightfully considered to be the father of computer science. Turing was fascinated by intelligence and thinking, and the possibility of simulating them by machines. Turing’s most prominent contribution to AI is his imitation game, which later became known as the Turing test.

which is a human, the computer is said to pass the test. The argument is that if a computer is indistinguishable from a human in a general natural language conversation, then it must have reached human-level intelligence.

entity is intelligent if it cannot be distinguished from another intelligent entity by observing its behavior. Turing just constrained the set of behaviors into discussion so that the interrogator can't base her or his decision on appearances.

One problem: does being human-like mean you are intelligent?

One criticism of the Turing test as a test for intelligence is that it may actually measure whether the computer behaves like a human more than whether it is intelligent. The test has indeed been "passed" by computer programs that keep changing the subject, make plenty of spelling errors, and sometimes refuse to respond at all. A famous example is Eugene Goostman, a 13-year-old Ukrainian boy who constantly tries to avoid answering questions by making jokes and changing the subject to his pet guineapig. Here is a short segment of an example conversation with Eugene Goostman from the blog of Scott Aaronson:

Is Eugene a computer or a person?

Interrogator: How many legs does a camel have?

Eugene: Something between 2 and 4. Maybe, three? :-))) By the way, I still don't know your specialty - or, possibly, I've missed it?

Eugene: Just two, but Chernobyl mutants may have them up to five. I know you are supposed to trick me.

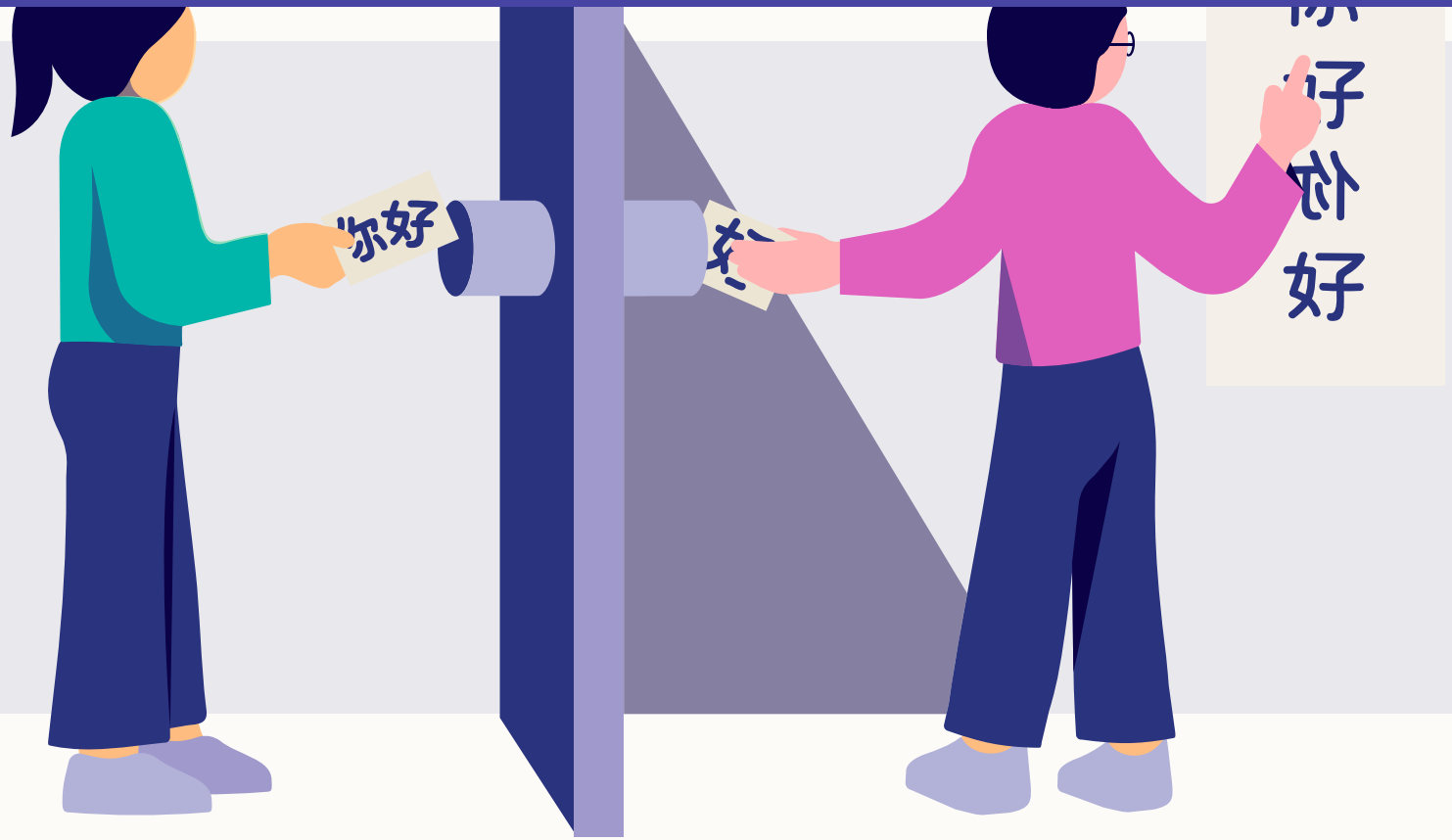
Interrogator: No, I need to know that you're not a chatbot. Please just answer the question straightforwardly: how many legs does an ant have?

Eugene: Something between 2 and 4. Maybe, three? :-))) Oh, what a fruitful conversation;-)

Interrogator: Do you understand why I'm asking such basic questions? Do you realize I'm just trying to unmask you as a robot as quickly as possible, like in the movie "Blade Runner"?

Eugene: ...wait

Eugene Goostman, who in reality is a computer program, fooled 10 out of 30 judges into thinking he is a real person.



The Chinese room argument

The idea that intelligence is the same as intelligent behavior has been challenged by some. The best known counter-argument is John Searle's Chinese Room thought experiment. Searle describes an experiment where a person who doesn't know Chinese is locked in a room. Outside the room is a person who can slip notes written in Chinese inside the room through a mail slot.

Searle argued that even if the person outside the room gets the impression that he is in a conversation with another Chinese-speaking person, the person inside the room does not understand Chinese. Likewise, his argument continues, even if a machine behaves in an intelligent manner, for example, by passing the Turing test, it doesn't follow that it is intelligent or that it has a "mind" in the way that a human has. The word "intelligent" can also be replaced by the word "conscious" and a similar argument can be made.

Is a self-driving car intelligent?

The Chinese Room argument goes against the notion that intelligence can be broken down into small mechanical instructions that can be automated.

A self-driving car is an example of an element of intelligence (driving a car) that can be automated. The Chinese Room argument suggests that this, however, isn't really intelligent thinking: it just looks like it. Going back to the above discussion on "suitcase words", the AI system in the car doesn't see or understand its environment, and it doesn't know how to drive safely, in the way a human being sees, understands, and knows. According to Searle this means that the intelligent behavior of the system is fundamentally different from actually being intelligent.

evasive and leads to apparently never-ending discourse. In intellectual company, this discussion can be quite enjoyable (in the absence of suitable company, books such as The Mind's I by Hofstadter and Dennett can offer stimulation).

However, as John McCarthy pointed out, the philosophy of AI is "unlikely to have any more effect on the practice of AI research than philosophy of science generally has on the practice of science." Thus, we'll continue investigating systems that are helpful in solving practical problems without asking too much whether they are intelligent or just behave as if they were.

Key terminology

General vs narrow AI

When reading the news, you might see the terms "general" and "narrow" AI. So what do these mean? Narrow AI refers to AI that handles one task. General AI, or Artificial General Intelligence (AGI) refers to a machine that can handle any intellectual task. All the AI methods we use today fall under narrow AI, with general AI being in the realm of science fiction. In fact, the ideal of AGI has been all but abandoned by the AI researchers because of lack of progress towards it in more than 50 years despite all the effort. In contrast, narrow AI makes progress in leaps and bounds.

between being intelligent and acting intelligently, which was emphasized by Searle. Strong AI would amount to a “mind” that is genuinely intelligent and self-conscious. Weak AI is what we actually have, namely systems that exhibit intelligent behaviors despite being “mere“ computers.

Answered

Exercise 4: Definitions, definitions

Which definition of AI do you like best? How would *you* define AI?

Let's first scrutinize the following definitions that have been proposed earlier:

1. "cool things that computers can't do"

Your task:

- Do you think these are good definitions? Consider each of them in turn and try to come up with things that they get wrong - either things that you think should be counted as AI but aren't according to the definition, or vice versa. **Explain your answers by a few sentences per item** (so just saying that all the definitions look good or bad isn't enough).
- Also come up with **your own, improved definition** that solves some of the problems that you have identified with the above candidates. Explain with a few sentences how your definition may be better than the above ones.

Please read the above instructions carefully and answer both of the items above in the text box below. Your answer will be reviewed by other users and by the instructors. Please answer in English, and check your answer before clicking 'Submit' because once submitted, you can no longer edit your answer.

Your answer:

1. Cool things are not all AI - AI is cool but it is complicated by various factors. Just describing as Coll is not justified.

3. Autonomous and adaptive systems are general electronics products which get updated regularly but do not have their own sense of adapting. So, this is not a proper definition of AI.

Example answer

There is no right or wrong answer, but here's what we think:

"Cool things that computers can't do"

The good: this adapts to include new problems in the future, captures a wide range of AI such computer vision, natural language processing.

The bad: it rules out any "solved" problems, very hard to say what counts as "cool".

"Machines imitating intelligent human behavior"

The good: the same as in the previous. Also, imitate is a good word since it doesn't require that the AI solutions should "be" intelligent (whatever it means) and it's instead enough to act intelligently.

such as so-called swarm intelligence (intelligence exhibited by for example ant colonies).

“Autonomous and adaptive systems”

The good: it highlights two main characteristics of AI, captures things like robots, self-driving cars, and so on, also nicely fits machine learning-based AI methods that adapt to the training data.

The bad: once again, these lead to further questions and the definition of 'autonomous' in particular isn't very clear (is a vacuum cleaner bot autonomous? How about a spam filter?). Furthermore, not all AI systems need to be autonomous and we can in fact often achieve much more by combining human and machine intelligence.

Your answer has been accepted!

You may still give peer reviews to help others

Give peer review

Received peer reviews:

After completing Chapter 1 you should be able to:

- Explain autonomy and adaptivity as key concepts for explaining AI
- Distinguish between realistic and unrealistic AI (science fiction vs. real life)
- Express the basic philosophical problems related to AI including the implications of the Turing test and Chinese room thought experiment

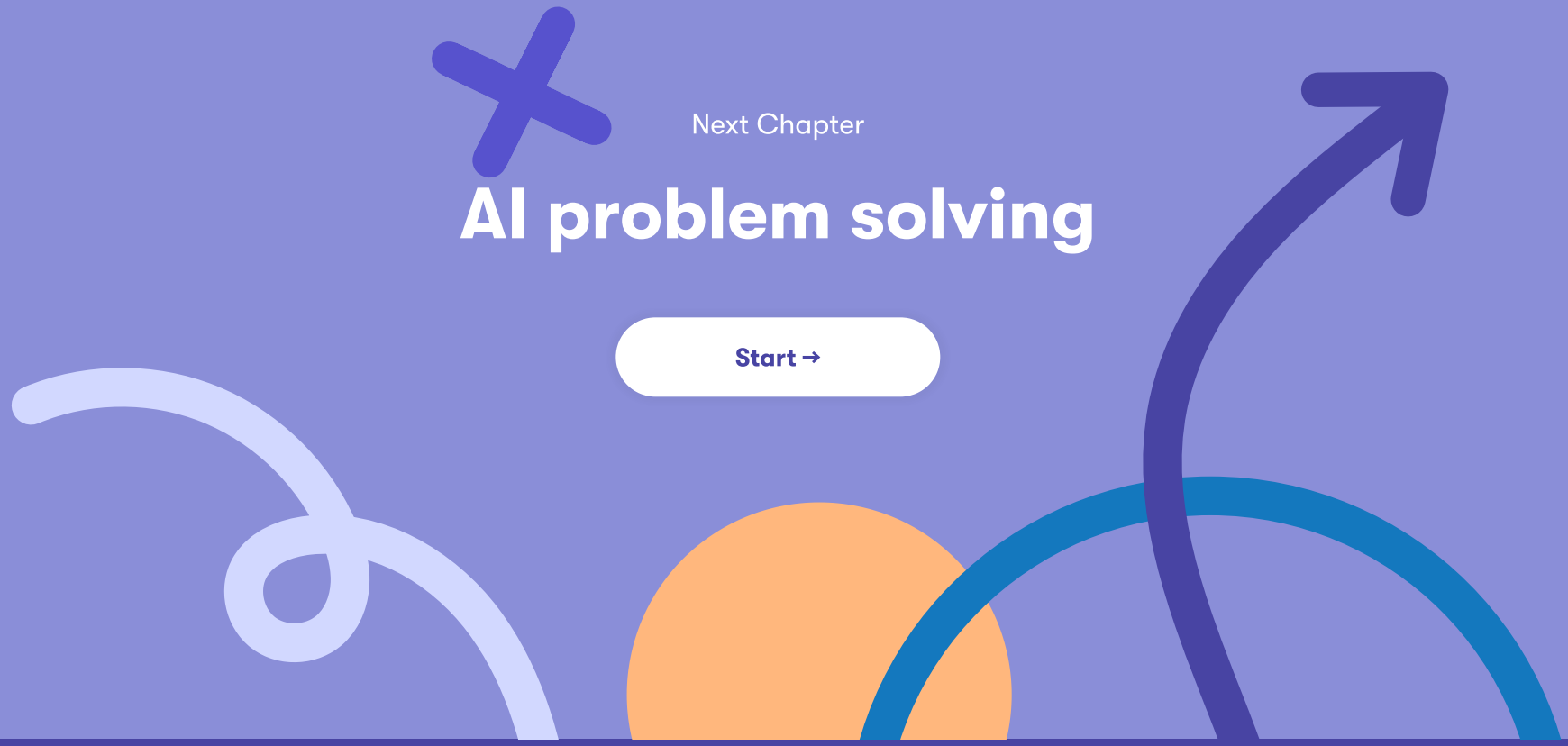
Please join the Elements of AI community at Spectrum to discuss and ask questions about this chapter.

Elements of AI

You reached the end of Chapter 1!

97 %
Exercises completed

25 /25



Next Chapter

AI problem solving

Start →

Course overview


HELSINGIN YLIOPISTO
HELSINGFORS UNIVERSITET
UNIVERSITY OF HELSINKI


Reaktor