

Philosophy & Ethics of Artificial Intelligence

Master in Philosophy of Science

Philippe van Basshuysen

21 October 2019

Worksheet 1 – Defining AI

1. **Who am I? Who are you?**
2. **Discuss the following questions with your neighbour:**

* What is intelligence?
* What is artificial intelligence?
* Things that AI can do:

1. **Many different definitions in the literature and public imagination.**

Discuss the following definitions. Which aspects are good, bad, useful, generate counterexamples?

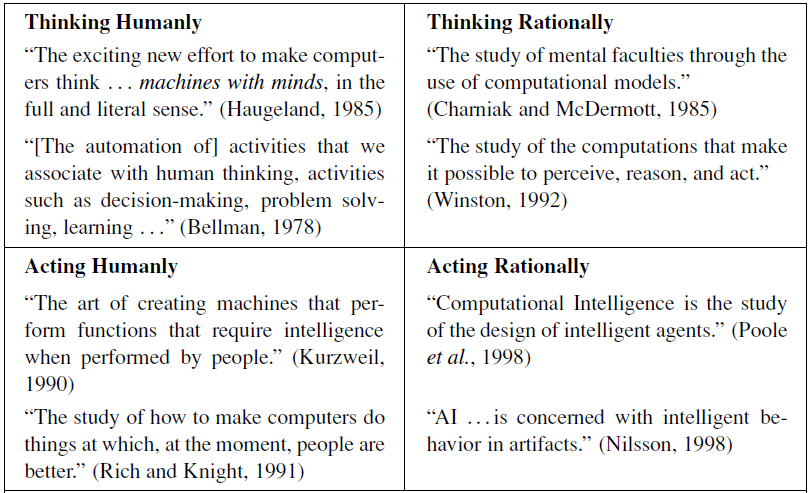
* “Cool things that computers can’t do.”
* “Artificial intelligence (AI) is the field devoted to building artificial animals (or at least artificial creatures that – in suitable contexts – appear to be animals) and, for many, artificial persons (or at least artificial creatures that – in suitable contexts – appear to be persons).” (SEP entry, Artificial Intelligence)
* AI is what passes the Turing Test.
* The characteristic properties of AI are *autonomy* and *adaptivity*:
  + Autonomy: “The ability to perform tasks in complex environments without constant guidance by a user.”
  + Adaptivity: “The ability to improve performance by learning from experience.” (*Elements of AI course*, <https://course.elementsofai.com/>)

1. **Are the following instances of AI or not? Explain why or why not.**

* Spreadsheet that calculates sums and other pre-defined functions on given data
  + Yes
  + No
  + Sort of
* Predicting the stock market by fitting a curve to past data about stock prices
  + Yes
  + No
  + Sort of
* A GPS navigation system for finding the fastest route
  + Yes
  + No
  + Sort of
* A music recommendation system such as Spotify that suggests music based on the user's listening behaviour
  + Yes
  + No
  + Sort of
* Big data storage solutions that can store huge amounts of data (such as images or video) and stream them to many users at the same time
  + Yes
  + No
  + Sort of
* Photo editing features such as brightness and contrast in applications such as Photoshop
  + Yes
  + No
  + Sort of
* Style transfer filters in applications such as Prisma that take a photo and transform it into different art styles (impressionist, cubist, ...)
  + Yes
  + No
  + Sort of

Exercise adapted from: *Elements of AI*

1. **Classifying Differing Approaches**



From Russell & Norvig, Artificial Intelligence: A Modern Approach (3rd edition, p. 2)

R&N take the approach “Acting Rationally” because it is more general than “Thinking Rationally”, and “The standard of rationality is mathematically well defined and completely general, and can be “unpacked” to generate agent designs that provably achieve it. Human behavior, on the other hand, is well adapted for one specific environment and is defined by, well, the sum total of all the things that humans do.” (p. 5)

Narrow AI vs. artificial general intelligence (AGI) vs. artificial superintelligence:

1. **Next Week: Agency and Responsibility**

*Required Readings:*

Andreas Matthias (2004). The responsibility gap: Ascribing responsibility for the actions of learning automata. *Ethics and Information Technology* 6, 175–183.

Christian List (2019). Group Agency and Artificial Intelligence. *Discussion Paper*

*Optional, additional readings:*

Migle Laukyte (2017). Artificial agents among us: Should we recognize them as agents proper? *Ethics and Information Technology* 19, 1–17.

Robert Sparrow (2007). Killer Robots. *Journal of Applied Philosophy* 24(1), 62-77.