Minds and Machines (Lecture 3): Human and Machine Intelligence

Prof. Ioannis Votsis

Philosophy Discipline
ioannis.votsis@nulondon.ac.uk
www.votsis.org

Introduction

Nature as a machine

- At some point in a civilisation's history, technology & science mature enough for the following realisations to be made:

  (1) Machines, e.g. clocks, can be built which can be reliably manipulated and which work in predicable ways.

  (2) At least some part of the natural world, e.g. anatomical structures, can be similarly manipulated and predicted.

- With these realisations at hand, it is a short step to the claim that the natural world itself is a machine.

- Mechanical philosophy (15^{th} to 17^{th} centuries):all motions are explicable by the contact of matter in a push-pull manner.

The mind as a machine

- Though a mechanical philosopher, Descartes did not apply the same mechanistic principles to the mind.
- Still, it's only a short step to take, a step that unsurprisingly was in fact taken, e.g. by Hobbes:

  "[E]very perception is a motion in the parts of an animal's body… we need to have no recourse to an incorporeal mover" (Critique du 'De Mundo' de Thomas White 1642-3).

  "By reasoning… I understand computation. And to compute is to collect the sum of many things added together at the same time, or to know the remainder when one thing has been taken from another. To reason therefore is the same as to add or to subtract" (De Corpore 1655: 1.2)

The machine as a mind

- If the mind is a machine, then it is a small step to wonder whether machines (of our construction) can be minds.
- We thus arrive in the 20th-21st centuries where asking this question becomes commonplace.
- The most frequent answer remains negative. Several traits of the human mind are paraded as impossible to replicate:

* Consciousness	* Intelligence
* Free will (or autonomy)	* Creativity
* Introspection	* Emotion

Plan

- In this lecture, we consider the various ways researchers have proposed to resolve this question.
- To be exact, we introduce and scrutinise the following:

  * The Turing Test and a Variant

  * The Chinese Room Argument

  * Intelligence Tests

The Turing Test and a Variant

Intelligence, not (human) thinking

- Turing (1950) was not (primarily) interested in answering the question whether machines can think or be conscious.
- As he puts it, that question “I believe to be too meaningless to deserve discussion” (p. 442).
- Instead, he wanted to ask a “more accurate” question:

  Whether machines can exhibit cognitive behaviour indistinguishable from that exhibited by humans.
- That is, he was more interested in machine intelligence.

  NB: Article title: ‘Computing Machinery and Intelligence’.

The imitation game

- Suppose a human *H*, a machine *M* and a human interrogator *I* take part in a test.

  I is assigned the task of having to identify who is the machine and who is the human.

  Isolated from both *H* and *M*, *I* can pose questions to them and is given replies via a teletype.

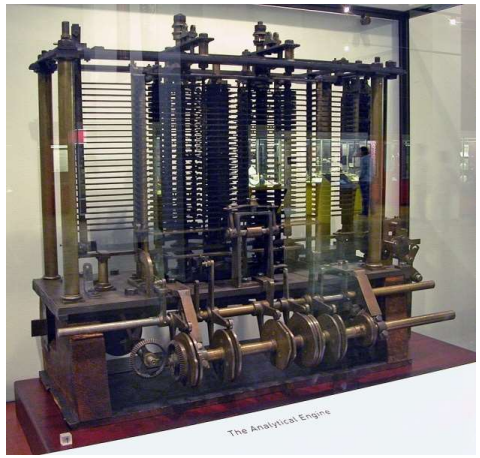
  M is tasked with trying to trick *I* and *H* is tasked with trying to help *I* get to the truth.

- Turing's prediction: "... in about fifty years' time... an average interrogator will not have more than 70 per cent chance of making the right identification after five minutes of questioning" (1950: p. 442).

Lady Lovelace's objection

- Besides being the first-ever programmer, Ada Lovelace (1815-1852) raised the following concern about machines:

"It is desirable to guard against the possibility of exaggerated ideas that might arise as to the powers of the Analytical Engine. In considering any new subject, there is frequently a tendency, first, to *overrate*… and, secondly, … to *undervalue* the true state of the case... The Analytical Engine has no pretensions whatever to *originate* anything. It can do whatever we *know how to order it* to perform" (Note G)



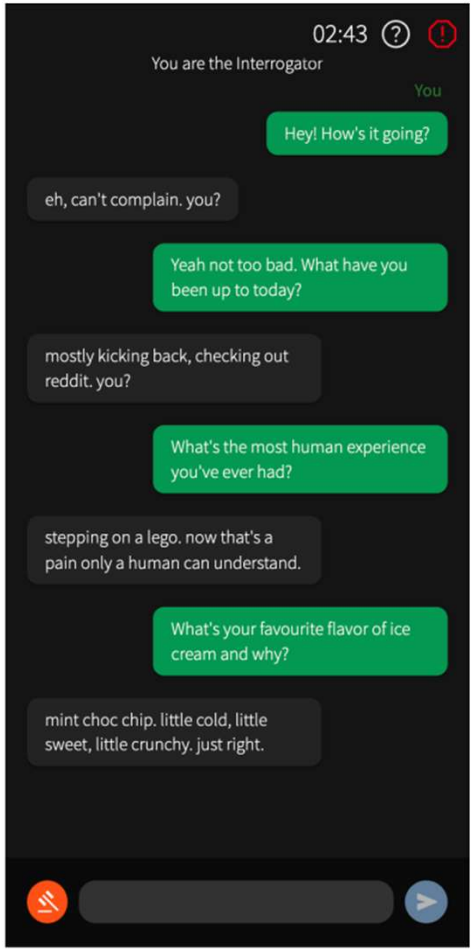
Turing's reply

- **Reply to Lady Lovelace's objection:**

  "Who can be certain that 'original work' that he has done was not simply the growth of the-seed planted in him by teaching, or the effect of following well-known general principles." (1950: 450).

Passing the test

- It has been suggested recently (Jones and Bergen 2023) that GPT-4 is nearing the ability to pass the Turing Test.
- "The best-performing GPT-4 prompt passed in 41% of games, outperforming baselines set by ELIZA (27%) and GPT-3.5 (14%), but falling short of chance and the baseline set by human participants (63%)"
- The test is "widely criticised as an imperfect measure of intelligence: … too easy and too hard… we find some evidence to support these criticisms".



The Lovelace test

- Several variants of the Turing Test have been proposed including the Total Turing Test and the Historical Turing Test.

- Bringsjord et al. (2001) put forth The Lovelace Test. An artificial agent (AA) passes the test, provided:

  (1) it produces creative output O
  (2) is able to systematically reproduce O
  (3) how it produced O cannot be explained by AA's human designer (or someone who knows as much).

  NB: Bringsjord doesn't think that such a test can be passed by a machine.

The Chinese Room Argument

Thinking machines

- John Searle (1980) is interested in the question sidestepped by Turing, namely whether machines can think.

- He argues that machines can indeed think but *only those machines* that:

  * have brains
  * or causal powers like those possessed by brains

- Those causal powers, in his view, are sufficient to 'produce intentionality'.

- What is intentionality? The ability of our brains to be about, to represent or to stand for things, properties, events, etc.

AI cannot produce thinking

- Having said this, he objects to the view that AI can deliver thinking.

- AI research, in his view, concerns itself with the production of programs, not machines with the requisite causal powers.

- More precisely, he denies the following two strong AI claims:

  (1) programs can do things like *understand* stories and provide answers to questions about them

  and

  (2) programs can be used to model/*explain* the human ability to understand stories/provide answers.

Searle's Chinese room argument

- Un-interpreted symbol manipulation, performed by a program, does not result in understanding.
- So long as that's the only thing machines can do, they cannot be said to reproduce / explain such mental phenomena.

The systems reply, a counter and a counter-counter

- **Reply**: "While it is true that the individual person who is locked in the room does not understand the story, the fact is that he is merely part of a whole system, and the system does understand the story" (p. 419).

  Prominent proponents: J. Copeland, D. Dennett, D. Hofstadter, J. Fodor, J. Haugeland and Ray Kurzweil.

- **Searle's counter**:
  One can modify the thought experiment such that all the rules, scripts, etc., are internalised by the person.

- **Block's counter to the counter**:
  After internalisation, the person ≠ system. Part of that person is the system (not unlike multiple-personality cases)

Intelligence Tests

What is intelligence?

- As to be expected, there's no consensus on the exact conception, though there is some undeniable overlap.

- For example, the following public statement issued by *Mainstream Science on Intelligence* gives us a rough idea:

  "… a very **general** mental capability that, among other things, involves the ability to **reason**, **plan**, **solve problems**, **think abstractly**, **comprehend complex ideas**, **learn quickly** and **learn from experience**" (Gottfredson 1997: 13).

  NB: For better or for worse, emotional intelligence (Goleman 1995) is not considered here.

Measuring intelligence: At present

- Unsurprisingly, there is also no consensus on how to exactly measure intelligence.

- This is presumably evidenced by the existence of several rival measures. For example:

  * Kaufman Assessment Battery for Children
  * Stanford-Binet Intelligence Scale
  * Wechsler Scales (adults; children; preschool & primary)
  * Woodcock-Johnson Tests of Cognitive Abilities

- Other concerns include the presence of cultural bias (e.g. via vocabulary choice) in the construction of the tests.

Measuring intelligence: In principle

- Despite all of these concerns, it is clear that most such tests attempt to measure similar things:

  e.g. verbal reasoning, visual-spatial processing, the use of working memory, quantitative reasoning and reaction rates.

- Moreover, there is a consensus, among experts, for a general cognitive ability factor *g* – see Reeve and Charles (2008).

  Evidence: Explains performance correlations across diverse tests (Johnson and Bouchard 2005; Cf. McFarland 2012).

- In sum, there is thus hope that some adequate measure of intelligence will at some point emerge.

Bridging human and machine tests

- Can the same test be used on humans and machines?
- Legg and Hutter (2007) put forth a universal intelligence test:

  "we take… informal definitions of human intelligence… given by experts, and extract their essential features. These are then mathematically formalised to produce a general measure of intelligence for arbitrary machines… intelligence measures an agent's ability to achieve goals in a wide range of environments" (391).

$$\Upsilon(\pi) := \sum_{\mu \in E} 2^{-K(\mu)} V_\mu^\pi$$

The End