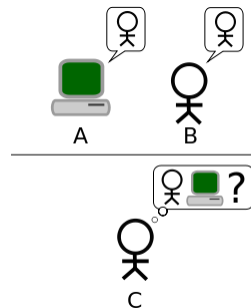


Thinking and Autonomy

Prasad Tadepalli

School of Electrical Engineering and Computer Science
Oregon State University

Turing Test (1950)



- The interrogator C needs to decide if he is talking to a computer A or a human B.
- If he is not able to distinguish A from B in 'a reasonable time,' then either
 - Both A and B think, or
 - Neither of them think
- It is more polite to say that both of them think

Quiz Question

1. What definition of AI is captured by the Turing Test?

- a) Thinking humanly
- b) Acting humanly
- c) Thinking rationally
- d) Acting rationally

Prediction

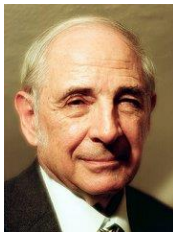
Turing predicted that computers would pass the Turing Test (fool 30% of testers) by 2000. It did not happen in any test setting. When do you think that this might happen?

- a) By 2025
- b) 2026-2050
- c) 2051-2075
- d) Never

Arguments against Turing Test

- **Mathematical Objection:** There are many problems for which there are no algorithms, e.g., the halting problem. So machines cannot solve them.
 - True, but there is no evidence that people can solve them either.
- **Lack of Originality** (Ada Lovelace): A machine can't do anything original because it needs to be programmed
 - Machines can *learn* to do things that surprise their programmers
 - The interrogator can ask questions that require creativity and conduct quizzes for originality
 - "Can you write a poem?"
 - In your poem you say, "Shall I compare thee to Summer's night?" Why not Spring night? Why not Winter's night?
- **Lack of consciousness** (Prof. Jefferson, Prof. John Searle)

Searle's Chinese Room



- Imagine a native English speaker that does not know a word of Chinese locked in a room in China.
- He has a rule book that allows him to perfectly answer any question put in Chinese by manipulating Chinese characters.
- He is so fast and so good in following the rules that he answers every question as well as a native Chinese speaker.
- Does he understand Chinese?

Searle's Argument in Slow Motion

- Turing is equating input-output behavior with thinking (this is called "strong AI" by Searle, "functionalism" in philosophy)
- In principle, the same behavior can be caused by multiple means
- It is most likely that machines exhibit intelligent behavior in a very different way than humans
 - Chess and Go machines search millions of positions. The neurons are too slow for people to be able to do this.
- Hence behavior alone is not a sufficient indicator of thinking.
- Counterarguments to Searle's thought experiment
 - **System's reply:** the person may not understand, but the room does
 - **Robot reply:** if you replace the Chinese room with a robot that manipulates the real world (rather than symbols), the robot is indistinguishable from a person who understands Chinese
 - **Pragmatic reply:** It is not practical to follow the rules as quickly as a Chinese speaker would

Question

Is the Chinese room argument convincing, i.e., do you believe that answering questions just like a human does not necessarily imply 'thinking' as the word is generally understood?

- a) Yes
- b) No
- c) Not sure

Why does it matter?



- South Korea's Super Aegis II has a supergun with a range of 4 kilometers, and is powerful enough to stop a truck
- Costing \$40M, it was deployed in Korea's DMZ, and sold to many Arab countries
- Safety features
 - It is required to issue a warning before shooting. Has a voice that can reach 3 kilometers.
 - The shooting requires an OK and a password from a human operator
 - Safeguards are largely self-imposed
- Are these safeguards sufficient? Are we entering a new era of "killer robots"?

The Maven Project

- Google had a contract from Pentagon to build an AI system for doing visual recognition to better target their drone weapons.
- Thousands of Google employees protested against the project.
 - *"Google should not be in the business of war.."*
- Last week Google decided not to renew the project.

Two Relevant Quotes

I am incredibly happy about this decision, and have a deep respect for the many people who worked and risked to make it happen. Google should not be in the business of war.

— Meredith Whittaker, Google employee

The same software that speeds through video shot with armed drones can be used to study customers in fast-food restaurants or movements on a factory floor... Google may want to act like they're not in the business of war, but the business of war long ago came to them.

— Peter Singer, author of *Like War*.

Is there a robot friend in your future?

- Child care robots can entertain kids, answer their questions
- Could assist staff at elder care centers
- Monitor health, provide companionship to kids and elders

"These machines do not understand us. They pretend to understand... To me, it lessens us. It is an inappropriate use of technology..."

— Sherry Turkle, professor of social studies of science of technology, MIT

AI and Democracy

Consider an app that advises you on how to vote in the elections. It takes into account the kind of things you read, like, and not like. It reads newspapers, blogs, etc. from your perspective and makes a recommendation.

Which one of these most reflects your opinion?

- a) I am going to follow its advice
- b) I am going to consider its advice
- c) I am not going to use the app
- d) I am not sure

Ethical Questions

- Trust: Can we trust the AI system? What justifies the trust?
- Legal liability: Who is responsible if something goes wrong?
- Psychological dependence: Is it healthy to treat machines as if they are humans?
- Costs & benefits: Who benefits from AI? Who bears the costs of increased automation?