

Richard W. Hamming



Learning to Learn

The Art of Doing Science and Engineering

Session 6: Artificial Intelligence I

Topic Outline



Can Machines Think?

Can Machines Think?



Computers manipulate *symbols*, not *information*.

We find it hard to even define a concept like information. Symbols, on the other hand, are almost arbitrary.

What are the limits of computers?

Are there things that humans can do that computers can't?

Philosophical Background: Turing Test



Hamming doesn't discuss this, but a little background:

The Turing experiment lets a user query via a teletype an entity in a locked room. If the user can't tell the difference between a machine and a human, we can think of the machine as being "intelligent" in some sense (maybe not literally!)

Can Machines Think?



John Searle proposes the "Chinese Room" thought experiment.

You sit in a locked room. You have a set of instructions (in English), someone drops in slips of paper (in Chinese) and, following the instructions, you match the input symbol and respond with another slip of paper (in Chinese).

This is analogous to the Turing experiment, with the English instructions as the "program," and the slips of paper as questions and answers.

You don't speak Chinese.

Do you "understand" Chinese in this experiment? Or are you just manipulating symbols?

Can Machines Think?



Searle's thought experiment is sometimes thought of as the difference between "Strong AI" and "Weak AI."

Strong AI proponents say that an appropriately programmed computer is not a simulation of a mind; it *is* a mind.

Weak AI advocates believe that the computer is only a simulation of the mind.

This is in part a black box/white box difference.

Humans are doing computation in the strong AI view, and that computation is intelligence; it is just that the computation is too complex for us to describe and understand at the present time.

Thermostats are "thinking" in a limited way.

Can machines think?



The question of materialism vs. dualism also quickly raises its head.

Materialism holds that everything is the result of physical phenomena. "Consciousness" is really just a byproduct of fancy chemistry and physics.

Dualism holds that there is a "spirit" separate and distinct from physical phenomena.

Rene Descartes was famous for this position: "cogito ergo sum (I think therefore I am)."

Penrose



Penrose (*Shadows of the Mind*) suggests 4 extreme positions:

- A. All thinking is computation; feelings of conscious awareness are evoked by computation.
- B. Awareness is a feature of the brain's physical action; any physical action may be simulated computationally, but the simulation does not evoke awareness.
- C. Physical action of the brain evokes awareness, but this awareness cannot even be simulated computationally.
- D. Awareness cannot be explained by physical, computational, or any other scientific terms.

Penrose



Penrose's categories A-C can be thought of as compatible with materialism, while D is dualist (at least).

A corresponds to Searle's strong AI description.

B says we can, in principle, make a Von Neuman machine pass a Turing test, but it would not be conscious.

C is still materialist, but suggests we couldn't make a Von Neuman machine pass a Turing test. But we might with some other, man-made mechanism, such as neural nets or biological computers.

D is compatible with a religious, mystic, or Cartesian dualist outlook.

Games



Games are often used as test cases in AI.

Games have clear rules, and we can determine when a participant has "won" or "lost", or at least gauge the participants effectiveness

Other situations are not as well defined as games. The rules are not clear, and objectives are fuzzy.

Games



Some games can be programmed such that they exhibit aspects that resemble human thinking.

- Cannibals & missionaries, theorem proving

The first attempts at a "general problem solver" used a small number of rules.

Later attempts increased this to 5,000 rules or more, and applied the rules to a specific problem domain with mixed results.

These are called rule-based systems.

Nature of human thought



But perhaps we can't think everything we know. Perhaps there are some thoughts our minds are physically incapable of holding, given the limitations of our biology.

Examples: can a bat hold certain ideas that humans can? Can bats form experiences that humans cannot conceive?

If this is true, this may present problems for rule-based systems.

Games



In the same way, this may correspond to position C in Penrose.

While humans may be able to create a device that mimics human thought, that device may not be a Von Neuman machine.

It may be that Von Neuman machines or Turing machines cannot express the things necessary for intelligence.

Games



Computers have been programmed to play checkers, and in fact have beaten state champions.

They also displayed a form of learning--by having various parameters that could be tuned, and playing games against itself until a superior system, the computer showed something resembling learning, in a genetic algorithm.

Learning?



Is this an example of learning by machines?

The program is telling the machine how to learn. But how is this different from a geometry teacher "programming" your mind with some axioms?

Tic-Tac-Toe



You can program a computer to play TTT (on a 4X4 grid) with a relatively small number of rules.

If you have three men in a row, play it as a win.

If you can't immediately win, block an opponent's immediate win.

If you have a fork, play it.

If your opponent has a fork, block it.

Beyond this the rules are somewhat hazy; you may choose to pick squares in certain high value spots on the grid. These are known as heuristics.

Objections to AI



Some people say "I wouldn't trust a computer with my life."

But in reality this is done all the time, via traffic controls, pacemakers, fly-by-wire systems, etc.

Computers are exceptionally good at vigilance tasks and fast computation.

This seems to be a less common objection these days.

Religion



Some people are hostile to the concept of machine intelligence, because they believe it is an essential part of humanity, and only God can create such things.

This is compatible with Penrose position D.

AI



Some define “intelligence” as “that which humans can do but machines can’t.”

This is a bit problematic, since it may constantly shift. A few years ago, the chess world champion was beaten by a machine. Does this mean the definition of AI changed at that point?

Duality



You can also think of AI as being analogous to the duality of photons. They are not either a particle or a wave, but both at the same time.

Likewise, you can think of machines as being both intelligent and not, at the same time.

Personal outlook regarding AI



Whatever your beliefs, you should be able to coherently defend them. If you can’t do this, you are likely to be badly led astray in the real world.