Philosophy and AI

Philosophy and AI

Philosophers (been around longer than computers) and have been trying to resolve the same questions that AI & Cognitive science claim to address:

- how do human minds work?

- can non-humans have minds?

Scientists have for thousands of years thought about the following question:

how can we combine mind and brain two quite distinct entities?

At first this question may seem trivial , but believe me, it is quite a hard one, perhaps one of the hardest problems in science today

Philosophical Issues

- Some definitions
  - *Weak AI*: (Searle) AI develops useful, powerful applications
  - *Strong AI*: AI develops computers that have cognitive minds comparable to humans
  - *Intentionality*: mental state that is purposefully directed at an object, task, concept,...
  - *Dualism*: separation of mind & soul from body

Some arguments against *weak AI* include:

•There are some things that computers cannot do, no matter how we program them.

•Certain ways of designing intelligent programs are bound to fail in the long run.

•The task of "constructing" the appropriate program is infeasible

•The arguments can (and sometimes have….) refuted by exhibiting a program with supposedly unattainable capabilities….

our main bone of contention: the *strong AI*: claims machines have minds comparable to human minds, debates on strong AI bring up some difficult conceptual problems in philosophy.

Computing Machinery and Intelligence by A.M Turing

- **Proposes the Imitation game (Turing test) as a test for intelligence:**
- **If a machine can't be told apart from a human in a conversation over a teletype, then that's good enough.**
- **The Turing Test: a suggested experiment for evaluating whether a system is intelligent. Q. can machines think?**
  - **Turing restated: can a computer system fool a human into believing the system is human?**

Turing Test

Turing's rationale: "The question and answer method seems to be suitable for introducing almost any one of the fields of human endeavor which we wish to include" (Turing 1950, p.435).

Turing's analysis: "The [imitation] game may perhaps be criticized on the ground that the odds are weighted too heavily against the machine. This objection is a very strong one, but at least we can say that if, nevertheless, a machine can be constructed to play the imitation game satisfactorily, we need not be troubled by this objection"

Turing 1950, p. 435).

Some extracts from the test:

Patient: Men are all alike.
ELIZA: In what way?
Patient: They are always bugging about something or the other
ELIZA: Can you think of a specific example?
Patient: Well, my boyfriend made me come here.
ELIZA: Your boyfriend made you come here?
Patient: He says I' m depressed much of the time.
ELIZA: I am sorry you are depressed.
…
Problems: ELIZA and similar programs stressed simple syntactic
Analysis and generation of sentences.

Turing test

•Is the Turing Test an effective measure of intelligence?

- **–what is it measuring?**
- **–what isn’t it measuring?**
- **–should not the means of thinking be relevant as well?**
    - **•if so, then how to we determine which AI algorithm is**
    - **•intelligent and which are just ‘crude simulations’ of intelligence**
    - **•intelligence**
- **–should we consider the interrogator as well? (i.e. is he/she naive? if so, could Eliza be considered intelligent?)**

Turing test: Some cited critiques:

- **Lady Lovelace**: Programs lack originality
- **Theological**: religious faith and believes
- Species-centricity: humans must be more intelligent!
- **Mathematical objection**: limits of formal computability

  Godel's theorem: given a sufficiently descriptive formal system, statements can be written that cannot be proven

  since all computer programs are formal systems, then there are problems that they'll not be able to solve…..

  What about humans, can humans solve all problems?
- **Consciousness**: emotions, self-concept, are necessary
- **Diversity**: humans are diverse, programs are not

Nervous system: non-discrete system
also: physical architecture changes

Behaviour is informal: can' t formally ascribe rules for all behaviours

An attitude toward the free will problem needs to be built into robots in which the robot can regard itself as having choices to make, i.e. as having free will.

Turing test critic: Intentionality and Consciousness

Is it enough to see how a machine works?

What about the internal "mental" states it has?

Important criticism: when trying to understand a program or mechanical device, helpful to know its internal workings as well as external behaviour.

-----The objection was forseen by Turing and he cites a speech by Prof. Jefferson.

"Not until a machine could write a sonnet or compose a concerto because of thoughts and emotions felt, and not by the chance fall of symbols, could we agree that machines equals brain----that is not only write it but know that it had written it"

Jefferson' s key point is Consciousness: a machine has to be aware of its own mental state and actions

Turing' s response: pp 831 (next slide)

Jefferson' s point still an important one: points out the difficulty of establishing any objective test for consciousness

According to Turing:

Why should we insist on a higher standard for machines than we do for humans?

-after all, in ordinary life we can never have any evidence about the internal states of other humans -> **so we cannot know that anyone else is conscious**!

Nevertheless, instead of arguing continually over this point, it is usual to have the polite convention that everyone thinks as Turing puts it.

Even though many, including Jefferson, have claimed that thinking necessarily involves consciousness, the work is mostly associated with John Searl

We will discuss experiment that Searl claims,
<u>refute the idea of strong AI</u> are:

Minds, Brains, and Programs
by J.R. Searle

- Chinese room experiment:
  - – uses Chinese, but can be any language unfamiliar to human subject
  - – subject sits in room:
    - input: Chinese text (paragraph) and questions
    - also has: rules describing how to manipulate Chinese symbols from questions and text, and output corresponding Chinese text
  - – hence experiment recasts "frame/script" understanding
    - human being is the "computer processor"

Chinese Room

omments (cont)

- –From abstract of full paper:
  - “Intentionality ... is a product of causal features
  - of the brain. I assume that this is an empirical fact
  - about the actual causal relationships between mental
  - processes and brains.”
  - intentionality cannot result by running a computer
  - program (gist of paper’s argument)
  - “Any mechanism capable of producing intentionality
  - artificially... must have causal powers equal to those
  - of the brain. This is a trivial consequence of (1st point)”
- –but what are these “causal features”, “causal powers”?
  - Searle does not define them
  - perhaps they are almost symbolic: synaptic connections

Chinese Room

- **Many counterarguments (full paper has 26). A few...**
  - **Systems Reply: the person doesn’t understand,**
  - **but the whole system does**
    - **Searle: then internalize whole system**
  - **Brain Simulator Reply: simulate the brain**
  - **functionality of a human directly by a computer**
    - **any understanding in brain is precisely latent**
    - **in simulation**
    - **variation: replace neurons in brain with**
    - **prosthetic neurons**
    - **variation: record brain state in a computer,**
    - **let program continue thinking for a while, and**
    - **then reload new state into human**
  - **Many Mansions Reply: new non-digital technology**
  - **will permit more realistic cognitive machinery**

Philosophical Pitfalls

There is one philosophical view that is attractive to people doing AI but which limits what can be accomplished.

This is logical positivism which tempts AI people to make systems that describe the world in terms of relations between the program's motor actions and its subsequent observations. Particular situations are sometimes simple enough to admit such relations, but a system that only uses them will not even be able to represent facts about simple physical objects. It cannot have the capability of a two week old baby.

Comments

Turing test:many AI people don’t regard it as a valid measure of intelligence

- it doesn’t measure what is necessarily important value

- it was a first attempt at asking what we should expect of intelligent system

Chinese room:Searle is a worthy critique of AI

- problem: intelligence is difficult to rigorously define

- everyone is arguing about concepts that are not clearly defined!

consider “pain”: it is a characteristic of biological systems
should we consider “intelligence” to be a similar
can a computer ever simulate it? Does it make sense?

In my opinion AI is hard, and what we are really doing is not trying to create another “intelligent” species but trying to understand how we ourselves work.

Comments

The following might interest you as recommended reading…

Philosophy

The mind' s I, Hofstader and Dennett – Far more fun than Science Fiction…A mind bending collection of essays exploring the possibilities of strong AI

Darwin' s Dangerous Idea, Dennett, 1995 – A very good case for strong AI embedding in it ideas from the biological world view.

AI

Out of Control: The new Biology of machines, Kevin Kelly, 1994