Friday Conclusions on AI:
Pro: Intelligence is an Emergent Property
Con: The Chinese Room
The Future of AI

Final Lecture

There are limitations to what algorithms can do, even in their own realm of pure information......

Decision Problems

- Undecidable
- NP-Hard
- NP-Complete
- NP
- P

Hey Gixel - We're compiling a comprehensive list of fetishes. What turns you on?

Anything not on your list?

Oh... hm.
Strong AI Thesis: A properly programmed digital computer could have a mind in the same sense that a human has a mind.

This is not (yet) something we can test scientifically, but we can discuss the current thinking pro and con.

PRO: Perhaps intelligence is an Emergent Property, and so current computers are not intelligent because they are not complex enough.

If a single neuron is not intelligent, but 100 trillion neurons can be, perhaps the same can be said of computer circuitry.

And the complexity of computers is growing exponentially....
Con: What arguments have been raised against the Strong AI Thesis?

The most famous such argument is John Searle’s Chinese Room Thought Experiment:

Suppose a computer is invented, out of hardware which runs a problem which convincingly passes the Turing Test in Chinese.

There is no difference between this machine and a room in which a man sits, following the instructions of the algorithm (the “rulebook”) for how to interpret Chinese sentences. Native Chinese speakers submit questions through a slot in the door, and he looks up the rules for how to answer these questions. The native Chinese speakers are convinced there is an intelligent human behind the door.

The man does not understand Chinese, and there is nothing in the room that “understands” Chinese, and there is obviously nothing about this scenario which suggests the room is intelligent, has a “mind,” or is “conscious.”

Therefore a computer does not either.

This has sparked a lively debate since it was introduced in 1980, in fact, some have suggested that the philosophical field of “Cognitive Science” be renamed “Responses to Searle’s Chinese Room Experiment”!

The Systems Response [Basically the Emergent Properties argument]:

The man in the room doesn’t understand Chinese, but the room as a whole does.

Searle’s Answer to the Systems Response: Simulation is not the same thing as the real thing; a mind is a property of a physical being.

Later, speaking about consciousness (which seem to be necessary for a mind):

“Computational models of consciousness are not sufficient by themselves for consciousness. The computational model for consciousness stands to consciousness in the same way the computational model of anything stands to the domain being modelled. Nobody supposes that the computational model of rainstorms in London will leave us all wet. But they make the mistake of supposing that the computational model of consciousness is somehow conscious. It is the same mistake in both cases.

— John R. Searle, Consciousness and Language, p. 16
The Non-Classical Computer Reply: The problem is with the design; perhaps electrical neural networks or quantum computers would be different.

Searle: Nope! These can be simulated by digital computers, so we're back where we started.

Recall the discussion about analog vs. digital information. When is the resolution of the digital version “good enough”?

The Robot Response: The problem is that the computer is not connected to the environment – have sensors which detect hear, see, and feel as well as humans. Put the Chinese Room inside the head of a robot and it will be a mind!

Searle: That's no different than saying, if I memorize the rule book for how to speak Chinese, and follow it, then everyone will think I speak and understand Chinese. But I don't understand a thing. It is still a simulation!
The “Ship of Theseus” or “Brain Replacement” Response:

Imagine that engineers have invented a tiny computer that simulates the action of an individual neuron. What would happen if we replaced one neuron at a time? Replacing one would clearly do nothing to change conscious awareness. Replacing all of them would create a digital computer that simulates a brain. If Searle is right, then conscious awareness must disappear during the procedure (either gradually or all at once). Searle’s critics argue that there would be no point during the procedure when he can claim that conscious awareness ends and mindless simulation begins.

Searle’s Answer: While going through the brain prosthesis, “you find, to your total amazement, that you are indeed losing control of your external behavior. You find, for example, that when doctors test your vision, you hear them say ‘We are holding up a red object in front of you; please tell us what you see.’ You want to cry out ‘I can’t see anything. I’m going totally blind.’ But you hear your voice saying in a way that is completely out your control, ‘I see a red object in front of me.’ ... [Y]our conscious experience slowly shrinks to nothing, while your externally observable behavior remains the same.”

We do not know the final answers to any of these questions, of course, but Artificial Intelligence seems to be on everyone’s mind....
Isaac Asimov's "Three Laws of Robotics"

1. A robot may not injure a human being or, through inaction, allow a human being to come to harm.

2. A robot must obey orders given it by human beings except where such orders would conflict with the First Law.

3. A robot must protect its own existence as long as such protection does not conflict with the First or Second Law.