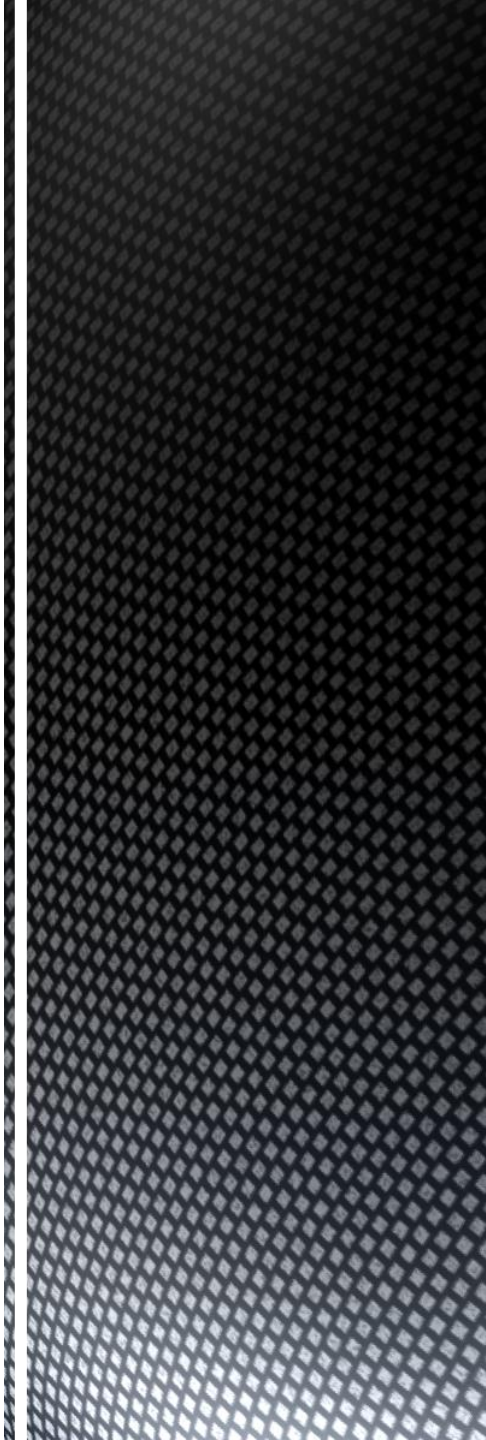
Philosophical Issues and Future Directions



Outline

- Weak AI: Can Machines Act Intelligently?
- Strong AI: Can Machines Really Think?
- Ethics and Risks of AI
- Agent Components
- Agent Architectures
- Are We Going in the Right Direction?
- What if AI Does Succeed?

Weak AI: Can Machines Act Intelligently?

- Field of AI founded on the assumption of weak AI
- Dijkstra (1984): “The question of whether machines can think… is about as irrelevant as the question of whether submarines can swim.”
- Thinking: Does it require a brain, or just brain-like parts?

Weak AI: Can Machines Act Intelligently?: The Argument from Disability

- Argument from disability claims that “machines can never do x”
- “It is clear that computers do many things as well as or better than humans, including things that people believe require great human insight and understanding. This does not mean, of course, that computers use insight and understanding in performing these tasks… but the point is that one’s first guess about the mental processes required to produce a given behavior is often wrong.”

Weak AI: Can Machines Act Intelligently?: The Mathematical Objection

- Godel's incompleteness theorem
  - For any formal axiomatic system F powerful enough to do arithmetic, it is possible to construct a so-called Godel sentence G(F) with the following properties:
    - G(F) is a sentence of F, but cannot be proved within F.
    - If F is consistent, the G(F) is true.
- Lucas claimed (1961) this proves machines to be inferior to humans because machines are formal systems limited to the incompleteness theorem
  - Problems:
    - 1. Based on supposition that computers are Turing machines, but Turing machines are infinite, computers not, therefore, not subject to Godel's incompleteness theorem
    - 2. "J.R. Lucas cannot consistently assert that this sentence is true." If he did, he would be contradicting himself, so he can't, therefore it is true. So Lucas is subject to the incompleteness theorem
    - 3. Computers have limitations on what they can prove, but there is no evidence that humans are immune to those limitations.

Weak AI: Can Machines Act Intelligently?: The Argument from Informality

- Qualification Problem:
  - The inability to capture everything in a set of logical rules
- Human behavior is too complex to be captured by a set of rules
- Dreyfus and Dreyfus criticize GOFAI
  - List of criticisms (that have been refuted in more recent developments)
  - Except… “embodied cognition” – the supposition that it makes no sense to consider the brain separately, and that computers lack the embodiment that humans (and other animals) have

Strong AI: Can Machines Really Think?

- Arguments about:
  - Consciousness
  - Phenomenology
  - Intentionality
- Turing's response:
  - There is no direct evidence that people think, and we shouldn't hold machines to a higher standard than people
  - "Instead of arguing continually over this point, it is usual to have the polite convention that everyone thinks."

Strong AI: Can Machines Really Think?: Mental States and the Brain in a Vat

- Intentional states:
  - Believing, knowing, desiring, fearing, etc. that refer to some aspect of the external world
- Brain in a vat thought experiment
  - “Wide content” – view of an omniscient outside observer
  - “Narrow content” – view from the inside, considers only the brain state

Strong AI: Can Machines Really Think?: Functionalism and the Brain Replacement Experiment

- Functionalism: a mental state is any intermediate causal condition between input and output
- Brain replacement experiment – replace neurons gradually such that behavior remains the same
  - At what point are you no longer you? Or is there such a point?

Strong AI: Can Machines Really Think?: Biological Naturalism and the Chinese Room

- Chinese room experiment
- Searle: “running the right program does not necessarily generate understanding”
- But creatures with neurons have been learning and deciding before consciousness evolved
  - The great mystery… why can a hunk of brain be a mind, while a hunk of liver cannot?

Strong AI: Can Machines Really Think?: Consciousness, Qualia, and the Explanatory Gap

- Qualia: Why is it that it feels like something to have certain brain states (e.g. eating a hamburger), while presumably it does not feel like anything to have other physical states (e.g. being a rock)?
- Suppose we have perfectly mapped and understood brain function – what neurons and processes happen under which circumstances
  - From our understanding, there is still no (external) proof that a person has consciousness
  - Explanatory gap

Ethics and Risks of AI

- People might lose their jobs to automation
- People might have too much (or too little) leisure time
- People might lose their sense of being unique
- AI systems might be used toward undesirable ends
- The use of AI systems might result in a loss of accountability
- The success of AI might mean the end of the human race