Conclusion

School of EECS
Washington State University
Philosophical Issues

- **Weak AI**
  - Machines can act *as if* they were intelligent

- **Strong AI**
  - Machines can *actually be* intelligent (i.e., think)

- Can we tell the difference?
- Is even weak AI achievable?
- Should we care about achieving strong AI?
- Are there ethical implications?
Weak AI?

- **Turing Test**
  - Can the machine convince a human that it is human via written English

- **Loebner Prize**
  - ([www.aisb.org.uk/events/loebner-prize](http://www.aisb.org.uk/events/loebner-prize))

- **AI XPRIZE** ([ai.xprize.org](http://ai.xprize.org)): $5M

- **Mitsuku** ([mitsuku.com](http://mitsuku.com))

Alan Turing (1912–1954)

The Singularity Is Near (2012)
Arguments Against Weak AI

Disability
  ◦ But a machine can never...
    • Beat a master at chess (√)
    • Compose a symphony (~)
    • Laugh at a joke
    • Appreciate beauty
    • Fall in love

Response
  ◦ Engineer different approaches (planes vs. birds)
  ◦ If we can understand how humans do it...
Mathematical objection
- Godel’s incompleteness theorem
  - In any formal system there are true sentences that cannot be proven
  - “This sentence is not provable” is true, but not provable

Response
- Formal systems are infinite, machines are finite
- Inability to prove obscure sentences not so bad
- Humans have limitations too
Arguments Against Weak AI

- Informality
  - Human behavior too complex to model formally

- Response
  - Usually assumes overly-simplistic models (e.g., propositional logic)
  - Learning can augment the model
Strong AI?

- Machine thinks like a human
- How do we define human thinking?
  - Machine has to know it passed the Turing test
  - Consciousness argument
- Mental state = physical (brain) state
- Mental state = physical state + ?
- Arguments ill-defined
- What is consciousness?
Strong AI?

- Functionalism: Yes
  - Brain maps inputs to outputs
  - Can be modeled as a giant lookup table
  - Brain in a vat

- Biological naturalism: No
  - Lookup tables are not intelligent
  - Searle’s Chinese room argument

- Does it matter?
Ethical Issues

- Losing jobs to automation
  - By most measures AI has created more (better) jobs than it has replaced

- AI used for illicit ends
  - Most advanced technologies used for good and bad
  - AI doesn’t kill people; people kill people
  - Autonomy
  - Surveillance and privacy

Ethical Issues

- AI responsibility
  - Generally, human experts are responsible for relying on AI decisions
  - Autonomous AI liability falls to the human designers
  - Can an AI system be charged with a crime?

“I, Robot” (2004)
Ethical Issues

- Stephen Hawking (2014)
  - “Success in creating AI would be the biggest event in human history. Unfortunately, it might also be the last.”

- Bill Gates (2015)
  - “I am in the camp that is concerned about super intelligence.”

- Elon Musk (2017)
  - “AI is a fundamental risk to the existence of human civilisation.”

- Henry Kissinger (2018)
  - “… whose culmination is a world relying on machines ungoverned by ethical or philosophical norms.”
Asimov’s Three Laws

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.
Google’s Five Problems

- Avoid Negative Side Effects
  - How can we ensure that an AI system will not disturb its environment in negative ways while pursuing its goals?

- Avoid Reward Hacking
  - How can we avoid gaming of the reward function?

- Scalable Oversight
  - How can we efficiently ensure that a given AI system respects aspects of the objective that are too expensive to be frequently evaluated during training?

- Safe Exploration
  - How do we ensure that an AI system doesn’t make exploratory moves with very negative repercussions?

- Robustness to Distributional Shift
  - How do we ensure that an AI system recognizes, and behaves robustly, when it’s in an environment very different from its training environment?

Ethical Issues

- End of human race
  - An unchecked AI system makes a mistake
  - Utility function has undesired consequences
  - Learning leads to undesired behavior
  - Singularity
  - Friendly AI

- "Colossus: The Forbin Project" (1970)
- "The Matrix" (1999)
- "I, Robot" (2004)
- "Transcendence" (2014)
Ethical Issues

- Robot/AI rights

“Bicentennial Man” (1999)  
“A.I. Artificial Intelligence” (2001)  
“Ex Machina” (2015)
Political Issues

“Preparing for the Future of Artificial Intelligence” October 2016

whitehouse.gov/blog/2016/10/12/administrations-report-future-artificial-intelligence
AI is the science and engineering of making intelligent machines

- Intelligent agents
- Search
- Logic
- Probabilistic reasoning
- Learning
- Natural language
Intelligent Agents

- Rational agent acts to achieve best possible outcome (i.e., maximize expected utility)
- Type of agent depends on type of environment
Search

- Goal-based, problem-solving agent
- Uninformed search
  - E.g., Iterative-deepening depth-first
- Informed (heuristic) search
  - E.g., A*
  - Designing good heuristics
- Local search
  - E.g. Hill climbing, stochastic
- Adversarial (game) search
  - E.g., Minimax with alpha-beta pruning
Logic

- Propositional logic
  - $\text{At(Wumpus,1,3)}$
- First-order logic
  - $\exists x, y \text{ At(Wumpus}, x, y)$
- Inference
- Resolution
- Theorem proving

- Planning combines search and logic
Probabilistic Reasoning

- Probability
- Conditional probability
- Probabilistic inference
- Bayes’ rule
- Bayesian network
  - Exact inference
  - Approximate inference

Thomas Bayes (1701–1761)
Learning

- Improving performance at some task through experience
- Supervised learning methods
  - Nearest neighbor
  - Naïve Bayes
  - Decision tree
  - Neural network
- How to choose the right model?
  - Overfitting
- Unsupervised learning (clustering)
- Reinforcement learning
Natural Language

- Natural language processing
  - Text classification
  - Information retrieval
  - Question answering
  - Information extraction

- Natural language communication
  - Grammars and parsing
  - Semantic interpretation
  - Generation
  - Speech recognition and synthesis

Jarvis…?
Other Topics

- Perception (vision)
  - Image processing
    - Edge and region detection
    - 3D cues: motion, stereo, texture, shading
  - Object recognition
  - Deep learning
  - cloud.google.com/vision/

- Robotics
  - Sensors and actuators
  - Mapping and localization
  - Navigation
  - Object manipulation
What’s Next for AI: The Russell & Norvig View

- Rational hybrid agent with probabilistic reasoning and learning
- Hierarchical knowledge representations to cope with scale
  - Knowledge from WWW
- “Compile” knowledge to solve specific, simpler problems
- Real-time AI
- **Bounded optimality**: Agent acts as best it can given its resources
What’s Next for AI

- Artificial General Intelligence (AGI)
  - Integration of techniques
  - Cognitive architectures
- Brain simulators
- Reverse engineering the brain
  - Neuroscience
- Deep learning
- Big data

- More AI movies!
Thank you!

Please complete your course evaluation.