Science Fiction and the Future of AI

Miles Brundage
ASU
April 16, 2015

Overview

- What's AI?
- Hot topics and voiced concerns
- Science fiction case studies
- Open issues

What is AI?

- In the popular imagination
  - ≋ everything and/or nothing?
- In science fiction
  - ≋ Helpful tool
  - ≋ Existential threat

What Do Experts Say?

- History – 1956 - Dartmouth conference
- Gripes over popular conceptualizations (if it works, it isn't AI)
- Machine learning, AI, robotics, etc.
- Turing's legacy
- Common definitions
- Benchmarks
  - Turing Test – what was Turing's point?
  - Visual Turing Test
  - Etc.

AI vs. Robotics

- AI: focused on cognition
  - ≅ Includes perception and action, but applications needn't be in “real world”
- Robotics: machines that perceive, think, and act
  - ≅ More focused on physicality
- Much overlap/cross-pollination/collaboration

Examples of Recent Progress

- Watson
  - Defeated world champion at Jeopardy!
- Deep reinforcement learning
  - DeepMind's Atari results
- Machine vision
  - From handcrafted features to learned features
  - Image classification to natural language descriptions

Research Areas

- Machine learning
- Natural language processing
- Machine vision
- Logic
- Planning
- Cognitive architectures
- Etc.

Some Hot Areas

- Deep learning
- Integrated systems
- Robot learning from demonstration
- Swarm robotics
- Human computation
- Computational creativity
- Etc.

Musk et al. voice concerns

- Musk and DeepMind
- Hawking, Gates, etc.
- Media portrayal
- Future of Life Institute RFP

Where is AI heading?

- Progress in many sub-fields
- Integration challenges
- More interactive
- More research on safety, ethics
- Transfer learning
- Neuroscience/AI mutual influences

Superintelligence

- Definition
- Importance of values
- Multipolar scenarios
- Approaches
  - ≅ Capability control
  - ≅ Motivation selection
- Debates
- See Bostrom 2014

AI and Science Fiction

- AI-like technologies in sci-fi predate AI as a field
- Erewhon, Rossom's Universal Robots, etc. as early examples
- Some inspired by sci-fi to work on AI/robotics

Science Fiction Case Studies

- Origin story
- Design/capabilities
- Impact

Star Trek

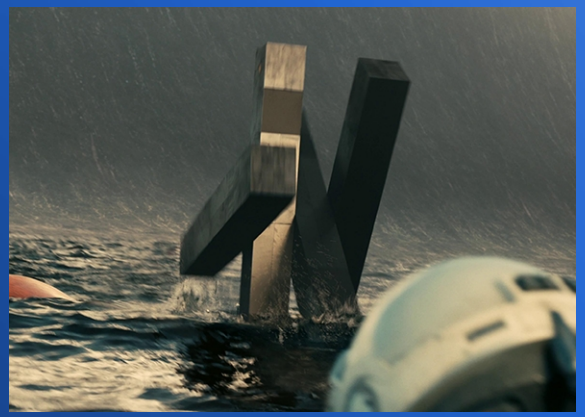
- Data and Lore
  - ≋ Role of emotions in cognition
- The Borg
- Central computer as model technology
  - ≋ “It comes up often.” - Amit Singhal, Google

Asimov

- Robots as a service
- 3 Laws of Robotics
  - ≅ Plus zeroth
- Future of work

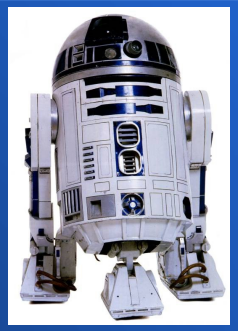
Interstellar

- TARS



- Role model for real AI/robots?
  - Transparency
  - Lack of survival instinct
  - Natural language
  - Unique form factor

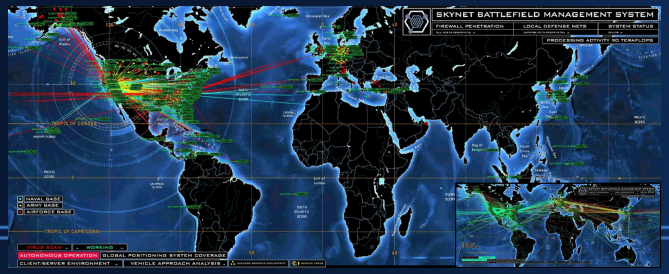
Star Wars

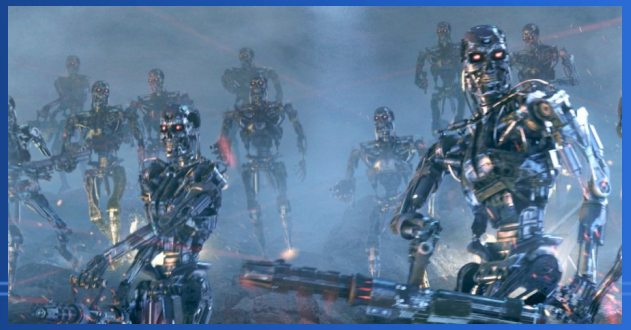


- Droids are ubiquitous
- R2D2 & C3PO
- Extended universe—IG-88 wants to replace us with machines...
  - ≅ Explanation for why robots are generally dumb?

Terminator

- Skynet
- Human failures leading to its takeover
- Implausibilities
  - ≅ Timeframe/extent of surprise
  - ≅ Subsequent stupidity
  - ≅ Humanoid form as dominant





Terminator (cont'd)

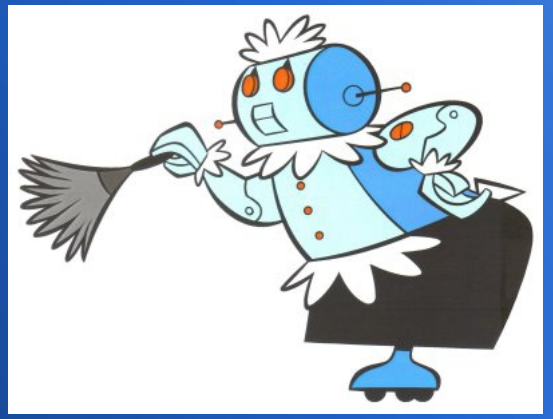
- “My CPU is a neural net processor – a learning computer.”
- Sarah Connor Chronicles and intelligence explosion

Battlestar Galactica

- First Cylon War
- Second Cylon War
- Caprica and Blood & Chrome
- Centurions and Raiders

The Jetsons

- Rosie
  - Cheap demo model
- Household robots
  - High end, low end
  - Challenges
- The future of work



Lessons from history of AI/sci-fi

- Importance of thinking ahead, and its (partial) tractability
  - ≅ Asimov got some things right
- Positive visions can inspire researchers
- Importance of AI designs viz-a-viz risks/opportunities
- Many possible human-machine configurations

Open Issues re: Future of AI

- Uncertainty re: future technical progress
- Future of work
- Security threats
- Human identity and Floridi's “fourth revolution”

Thanks!

- Questions/comments?

- Acknowledgments: Thanks to Spencer Green and Jonathan Jackell for helpful comments.