The Future of Artificial Intelligence

CSE 415: Introduction to Artificial Intelligence
University of Washington
Spring, 2019

Outline

Common Sense, Ontologies
Dangers of AI
Asimov's Three Laws of Robotics
Will they be like us?
Tools vs Agents
Technological Challenges
Social Challenges

Big Issues in AI

Is there a key concept, technique, or technology that will enable widespread use of AI?

Common sense...
Computers lack it.
Why?
a. It involves much knowledge that humans take for granted, but can be difficult to codify.
b. Computers aren't very "grounded" in the world (Computer vision isn't integrated with NLU yet);
Thus they miss context that people assume.

Big Issues in AI (cont)

For computers to have common sense, they need:

a. **A large knowledge base about the way things work in the world, and about how people think and communicate.**
b. **Ability to perceive contexts, such as see the environment.**
c. **Ability to learn efficiently like people do.**

Some believe that a convergence of technologies will lead to a great breakthrough around the year 2041. (Ray Kurzweil: ***The Singularity is Near***).

IBM's Watson system demonstrates that integration of multiple computing technologies can win at Jeopardy (like a Turing test).

Pros and Cons of Achieving High Levels of AI

Pros...
Powerful tools, solutions to tough problems, better standards of living(?)

Cons...
Tools might be used against people; technology may create worse problems than it solves; standards of living might get worse; we might feel we lose some aspect of our humanity.

Robots Revolt in R.U.R. – a 1920 play

Isaac Asimov's Three Laws of Robotics (1940)

First Law: A robot may not injure a human or through inaction, allow a human to come to harm.

Second Law: A robot must obey the orders given it by human beings, unless such orders would conflict with the first law.

Third Law: A robot must protect its own existence, as long as such protection does not conflict with the first or second law.

Are the 3 Laws the Answer?

Carrying out the laws requires very sophisticated judgment.

Extending the Laws(?!)

Zeroth law: A robot may not injure humanity or through inaction allow humanity to come to harm.
(due to Asimov, Olivaw, and Calvin).

Will They Be Like Us?

Like us, AI systems...
...will talk to us in our languages.
...will help us with our problems.
...will have anthropomorphic interfaces.

Unlike us, AI systems...
...will compute and communicate extremely quickly.
...will have bounds for learning and retention of knowledge that will soon surpass ours.
...might not be well modeled by the psychological models that work for people.

Tools vs Agents

Agent: Takes responsibility, takes initiative, interacts with others on behalf of a client.

Tool: Responds directly to its user. Does not take responsibility. Does not take initiative. Does not normally interact with others on behalf of a client.

Technological Challenges

Giving computers "common sense" is still an unrealized goal.

Human language is diverse (there are many languages, dialects, and idiolects) and often ambiguous. Computers don't yet understand it.

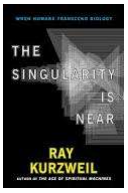
General AI systems that can redesign themselves.

Achieving the Transition to the Next Epoch of Civilization

Ray Kurzweil, in *The Singularity Is Near*, says that it will happen in the next century; maybe in the next generation.

AI will soon reach a "singularity" in which its effectiveness suddenly increases enormously, due to a virtuous circle of improvements.

Achieving the Transition to the Next Epoch of Civilization



Aspects of the Singularity

Convergence: It results from multiple exponential trends – computer fab. Technology, brain science, genome science, and A.I.

Immortality: A person could just "upload" her/his personality to the system (cloud? Matrix?)

Superhuman intelligence will be benevolent.

AI will be the vehicle to spread human civilization throughout the universe.

Helping the Singularity

Networked A.I. services such as Siri, Cortana, and Alexa can quickly deploy the latest A.I. techniques since they are server-based.

Networked robots, driverless cars, and other agents, can learn in parallel (by the thousands or millions), and immediately share the results of that learning with each other.

Although people can learn in parallel by the billions, they cannot share their knowledge as efficiently as networked robots can.

Singularity: Questions

What aspects do you consider believable?

Does it help to make a goal out of it?

How would economics relate to it?

Won't politics or religion get in the way?

How do we make sure it will be good?

Where are the best opportunities to contribute?

Social Challenges

Users need to understand the limits of their tools and agents. (Expert systems tend to be brittle)

AI applications need to be created that help bring harmony to the world rather than which intensify battles.

AI applications are needed which enhance the economy rather than reduce economic competition.

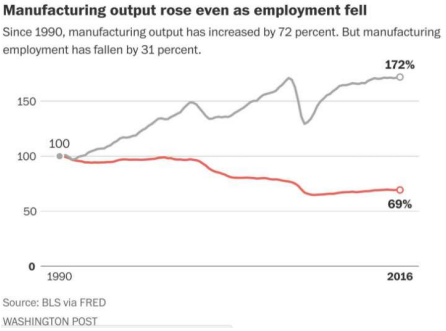
AI extends the reach of automation and threatens to eliminate, if not change many white-collar jobs.

AI raises the bar for information literacy and computer literacy.

Social Challenges

Robotic systems work on a car at the Jaguar Land Rover factory in March in Solihull, England. (Leon Neal/Getty Images)

Social Challenges



Preventing an A.I. Apocalypse

Several leading scientists and technologists have called for caution in the development of A.I. (Stephen Hawking, Elon Musk, Bill Gates)

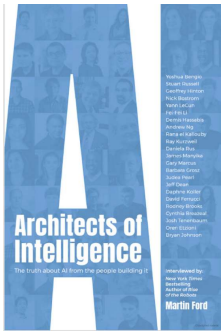
"A significant problem is that unfriendly artificial intelligence is likely to be much easier to create than friendly AI. " (Wikipedia)

In January 2015, Nick Bostrom joined Stephen Hawking, Max Tegmark, Elon Musk, Lord Martin Rees, Jaan Tallinn, and numerous AI researchers, in signing the Future of Life Institute's open letter speaking to the potential risks and benefits associated with artificial intelligence. The signatories
" …believe that research on how to make AI systems robust and beneficial is both important and timely, and that there are concrete research directions that can be pursued today.[13][14]"

Preventing an A.I. Apocalypse

Research is needed today on how to make sure future A.I. systems are *beneficial*.

For Further Reading



(C) Packt Publishing, 2018