AI and society

© Bruce Adams/Daily Mail

Image soruce

Outline

- Current concerns
  - Privacy
  - Bias and fairness
  - Manipulation
- Medium-term concerns
  - Safety
  - AI weapons
  - Jobs
- Long-term concerns
  - Superintelligence
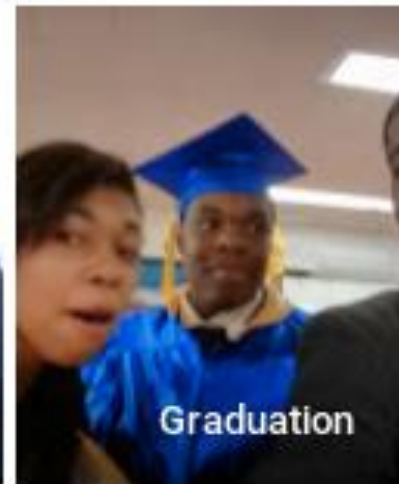
AI and privacy

Image source

AI and privacy

- Types of data collected
  - Relatively insensitive: shopping, browsing, web search history, social media, personal preferences
  - Sensitive: face, identity, financial and medical records
  - Very sensitive: potentially damaging surveillance data

- Entities collecting data
  - "Big data" companies: Facebook, Google, Apple, Amazon, etc.
  - Stores, employers, health ensurers, banks, etc.
  - Government, law enforcement, hostile parties

AI and privacy

- Concerns
  - Personal data being inadvertently revealed or falling into the wrong hands
  - Personal data being misused by the parties who collected it
  - Personal data enabling individuals to be manipulated or controlled with or without their knowledge
- Potential solutions
  - Technological: encryption, differential privacy, anonymizing tools
  - Regulation

AI, bias, and fairness

https://bits.blogs.nytimes.com/2015/07/01/google-photos-mistakenly-labels-black-people-gorillas/

"A hot, blond girl getting criticized by her boss."

E. van Miltenburg, "Stereotyping and Bias in the Flickr30K Dataset," Proceedings of the Workshop on Multimodal Corpora: Computer vision and language, collocated with LREC 2016

AI, bias, and fairness

- Concerns
  - AI will inadvertently absorb biases from data
  - Making important decisions based on biased data will exacerbate bias: especially for law enforcement, employment, loans, health insurance, etc.
  - Even well-intentioned applications can create negative side effects: filter bubbles, targeted advertising
  - Outcomes cannot be appealed because AI systems are opaque and proprietary
- Potential solutions
  - Regulation and transparency: e.g., right to explanation
  - More inclusivity among AI technologists: AI4ALL

AI, bias, and fairness

- Readings
  - [AI NOW report](#) (2017)
  - Weapons of math destruction
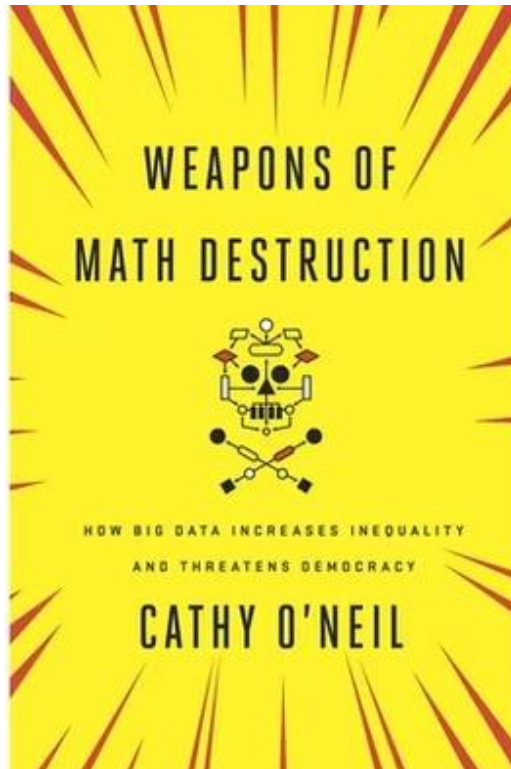
AI and manipulation

- [Fake news](#)
- [Deep Fakes](#)

AI safety

- Robustness to changes in data distribution
- Avoiding catastrophic "corner cases"
- Robustness to adversarial examples or attacks
- Avoiding negative side effects in reward function
- Avoiding "reward hacking"

- Reading: Concrete AI safety problems

AI weapons

AI weapons

Australian and Canadian AI Experts Call for Autonomous Weapons ...
Futurism - Nov 8, 2017
In two letters addressed to the heads of state in Australia and Canada, hundreds of experts in the field of artificial intelligence (AI) have urged for the ban of "killer robots," artificially intelligent weapons with the ability to decide whether a person lives or dies. They join a growing crowd of scientists who have ...

When AI rules, one rogue programmer could end the human race
BGR - Nov 8, 2017

Artificial intelligence will soon be used to create 'weapons of mass ...
International Business Times UK - Nov 8, 2017
Canadian AI experts urge for global ban on killer robots
International - BetaKit - Nov 8, 2017

'Slaughterbots' film shows potential horrors of killer drones
CNNMoney - Nov 14, 2017
The film is the researchers' latest attempt to build support for a global ban on autonomous weapon systems, which kill without meaningful human control. They released the video to coincide with meetings the United Nations' Convention on Conventional Weapons is holding this week in Geneva, ...

Killer robots are almost a reality and need to be banned, warns ...
Telegraph.co.uk - Nov 14, 2017
Ban autonomous killer robots, urge AI researchers
Radio Canada International - Nov 14, 2017
'Slaughterbots' Video Depicts a Dystopian Future of Autonomous ...
In-Depth - Seeker - Nov 15, 2017

The UN is worried about killer robots. We should be, too.
News & Observer - Dec 4, 2017
Agreements banning nuclear weapons from space are likewise a precedent, as are those prohibiting the use of laser weaponry to blind people, enacted by the Convention on Certain Conventional Weapons in 1995. We need a ban on autonomous offensive weapons in a similar way. As with nuclear ...

AI weapons

- Reading
  - [Robotics: Ethics of artificial intelligence](#) (Nature, May 2015)
  - [Humans, not robots, are the real reason artificial intelligence is scary](#) (The Atlantic, August 2015)

AI and jobs

AI and jobs

- Why we should worry
  - Oxford report: 47% of American jobs at high risk of automation in the next two decades
  - In the past couple of decades, manufacturing employment has dropped even as output kept rising; labor force participation among working-age males has been dropping
  - Truck driver is the most common job in over half the states
- Why we shouldn't worry
  - Productivity growth is currently low, as is business investment spending
 - Historically, automation has destroyed jobs but added more new jobs

AI and jobs

- Possible solutions
  - Regulating automation, mandating new jobs
  - Policy solutions: welfare, retraining, universal basic income, redistribution of assets
  - More profound revolution of human values!

AI and jobs

- Reading
  - [Moshe Vardi talk](#) (YouTube)
  - [AI NOW report](#) (2017)
  - [Technological unemployment](#) (Wikipedia)
  - [A world without work](#) (The Atlantic, July 2015)
  - [The automation paradox](#) (The Atlantic, Jan. 2016)
  - [AI will transform the economy. But how much, how soon?](#) (New York Times, Nov. 2017)
  - [Welcoming our new robot overlords](#) (New Yorker, Oct. 2017)
  - [The great tech panic: robots won't take all our work](#) (Wired, Aug. 2017)

Superintelligence
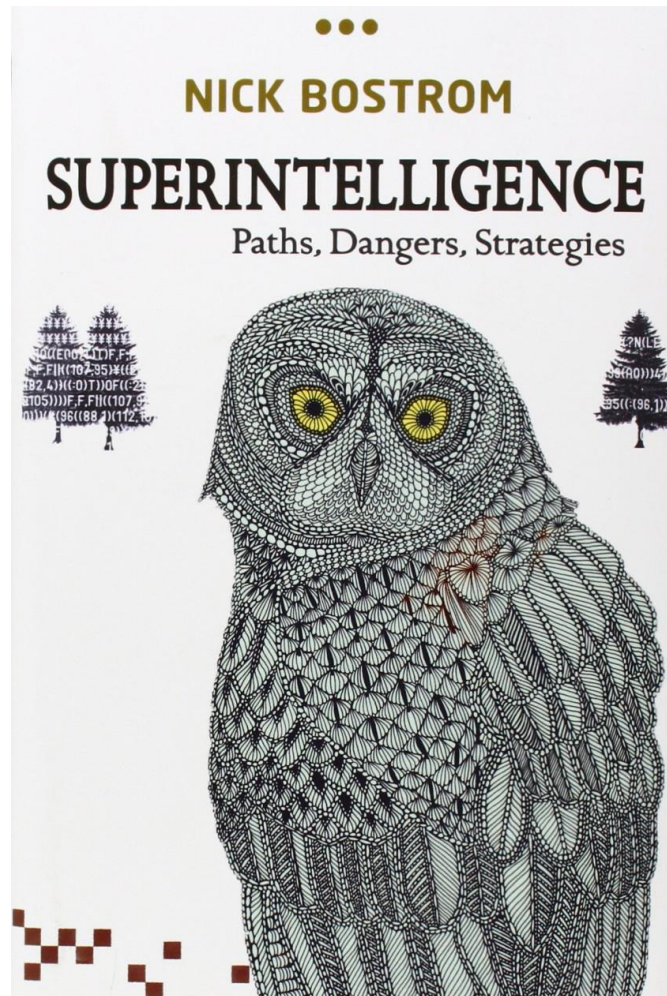
Superintelligence

- Why we should worry
  - Recent dramatic progress in AI – we could be at the "knee" of an exponential growth curve
  - No fundamental reasons why general human-level AI cannot be achieved
  - Positive feedback loops will kick in once a certain level of intelligence has been achieved
  - Historically, people have not seen disruptive innovations coming or underestimated their probability
  - Pascal's wager

Superintelligence

- Why we shouldn't worry
  - Technological obstacles to general AI are still too great (and have historically been underestimated by AI scientists)
  - The notion of "superintelligence" is too simplistic: intelligence is multifaceted, embodied, collective, reliant on (and limited by) the physical world
  - We have no evidence that recursively self-improving intelligence is possible
  - It is unclear why "superintelligent" AI would develop or single-mindedly pursue destructive goals
  - High intelligence is neither necessary nor sufficient for controlling the physical world effectively

Superintelligence

- Readings
  - [Superintelligence: The Idea that Eats Smart People](#) (Maciej Ceglowski)
  - [Discussion of "superintelligence"](#) (Neil Lawrence)
  - [The Myth of a Superhuman AI](#) (Kevin Kelly)
  - [Seven deadly sins of predicting the future of AI](#) (Rodney Brooks)

AI ethics

- We should be aware of all these issues when developing AI technologies!
  - Privacy violations
  - Potential for deception, misuse and manipulation
  - Exacerbating bias and unfair outcomes
  - Lack of transparency and due process
  - Threats to human rights and dignity
  - Weaponization
  - Unintended consequences