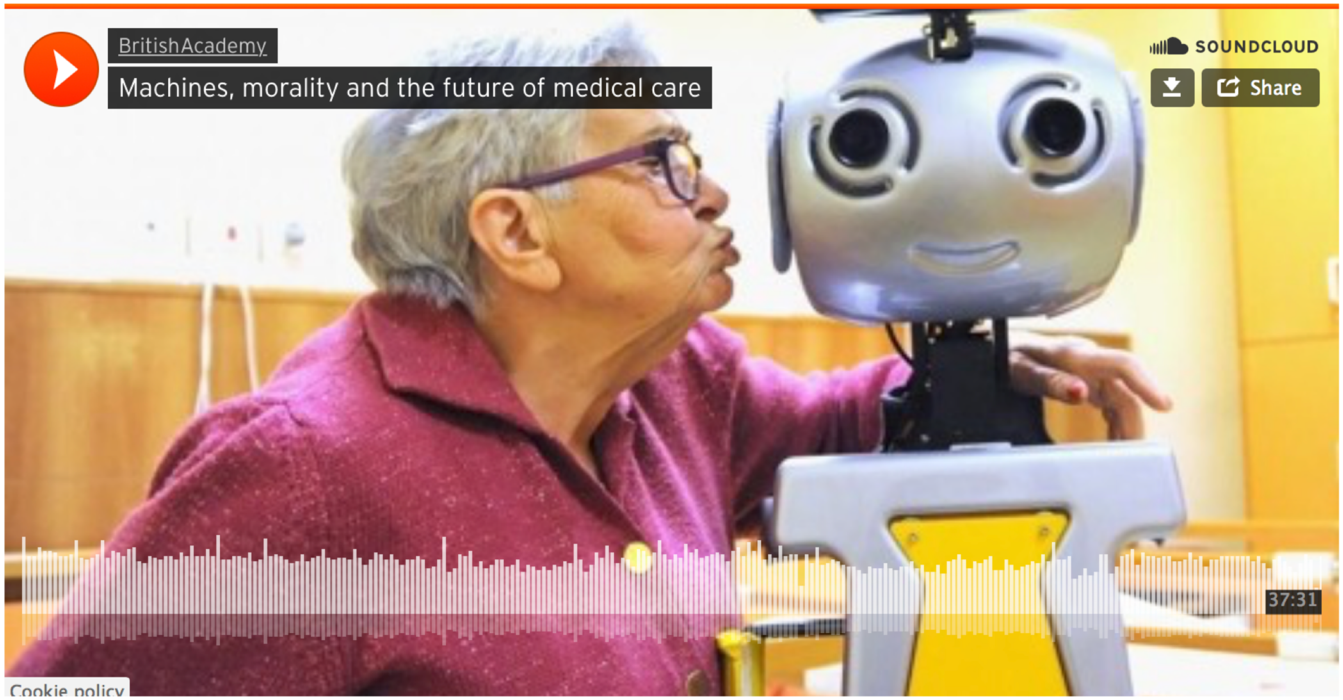
CS440/ECE448 Lecture 27: Societal Impacts of AI



Slides by Svetlana Lazebnik, 12/2017

Modified by Mark Hasegawa-Johnson, 4/2019

Image source: https://www.britac.ac.uk/audio/machines-morality-and-future-medical-care

Outline

- AI as a Tool
  - Privacy vs. Convenience
  - Health
  - Jobs
- Autonomous AI
  - Bias and fairness
  - Safety
  - AI weapons
  - Superintelligence

AI as a Tool

By Sémhur - Own work, CC BY-SA 4.0,
https://commons.wikimedia.org/w/index.php?curid=8247730

Privacy vs. Convenience

Image source

Privacy vs. Convenience

- Types of data collected
  - Relatively insensitive: shopping, browsing, web search history, social media, personal preferences
  - Sensitive: face, identity, financial and medical records
  - Very sensitive: geospatial location as a function of time
- Entities collecting data
  - "Big data" companies: Facebook, Google, Apple, Amazon, Microsoft, etc.
  - Stores, employers, health insurers, 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 without their knowledge
- Potential solutions
  - Technological: encryption, differential confidentiality, anonymizing tools
  - Regulation: require the use of a technology; forbid disclosure

Example Problem: Inadvertent Revelation

“Passports, however, use a different technology known as RFID (or Radio Frequency Identification), the same type used to tag clothing, pets, even artificial replacements for hips and knees. When embedded in a U.S. passport, the chip can be scanned only by someone at close range with an RFID reader, usually within a couple feet…

“Yes, someone nearby could read what’s in your wallet. That’s why I keep my passport in an RFID-shielded wallet,” said G. Mark Hardy, president of National Security Corp., based in Rosedale, Md., which provides cybersecurity expertise to government and corporate clients.

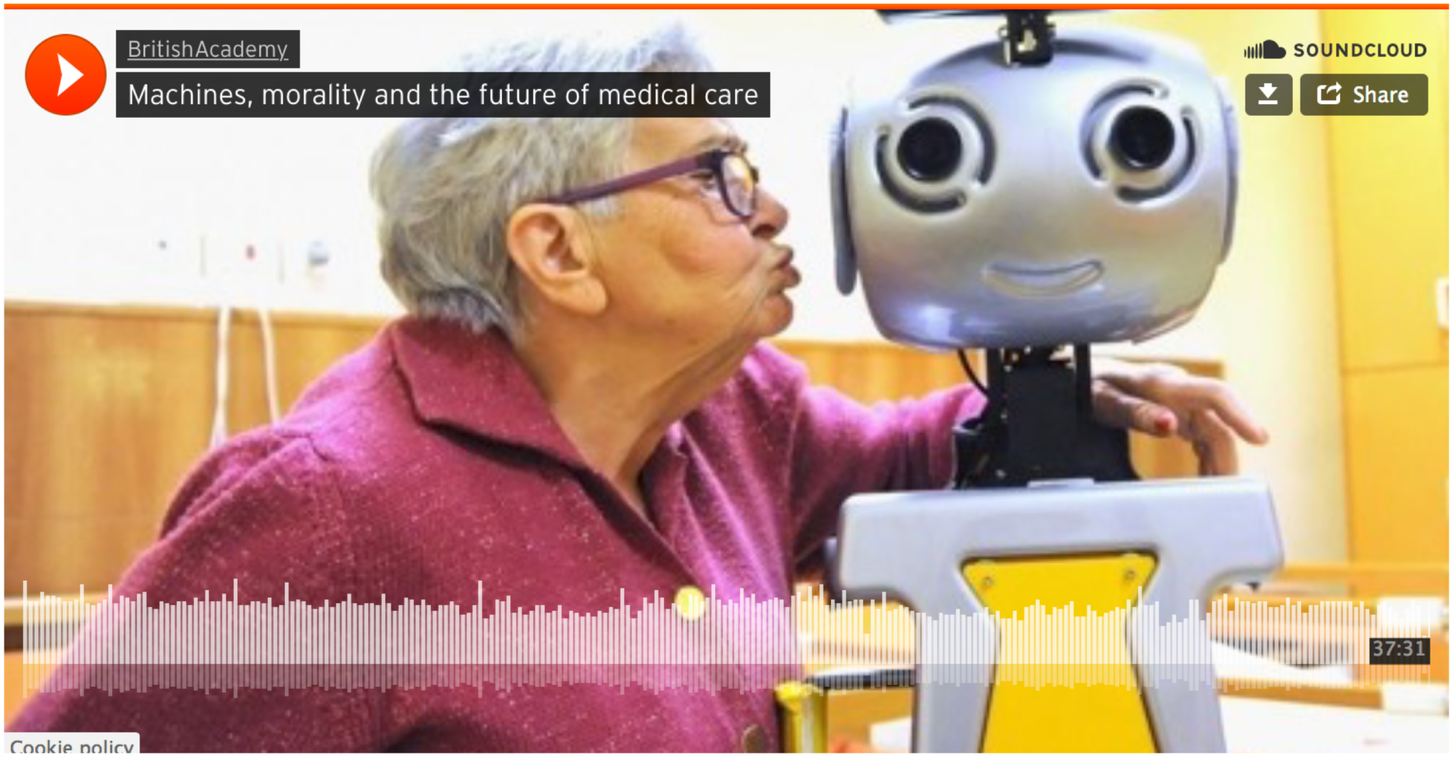
But, he said, “it’s less likely to happen, at this point in time, because it’s so much easier to do fraud some other way.””

Read more here: http://www.sacbee.com/news/business/personal-finance/claudia-buck/article2599038.html#storylink=cpy

Example Solution: Differential Confidentiality

- A social scientist wants to collect data about some sensitive behavior. She tells her subjects:
  - Toss a coin.
  - If it’s heads, answer truthfully. If tails, use a second coin toss to decide what you’ll tell me.
- Outcomes:
  - It’s impossible to know whether or not any individual engages in that behavior.
  - … but if X% of subjects say “yes,” then the truth is that ((X-25)/50% of the population engage in this behavior.

Health Care



BritishAcademy

Machines, morality and the future of medical care

AI and Health Care

- Mining Healthcare records
  - Pro: better and faster health service; MUCH more accurate diagnosis; epidemiology
  - Con: potential for misuse of data
- AI-assisted diagnosis, e.g., in Radiology
  - Pro: AI can detect things that a physician might miss
  - Con: Lack of explainability
- AI-assisted design of treatment plans
- AI-assisted drug creation
- Long-term health care: the robot nurse, the robot doctor

AI and jobs

Source

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

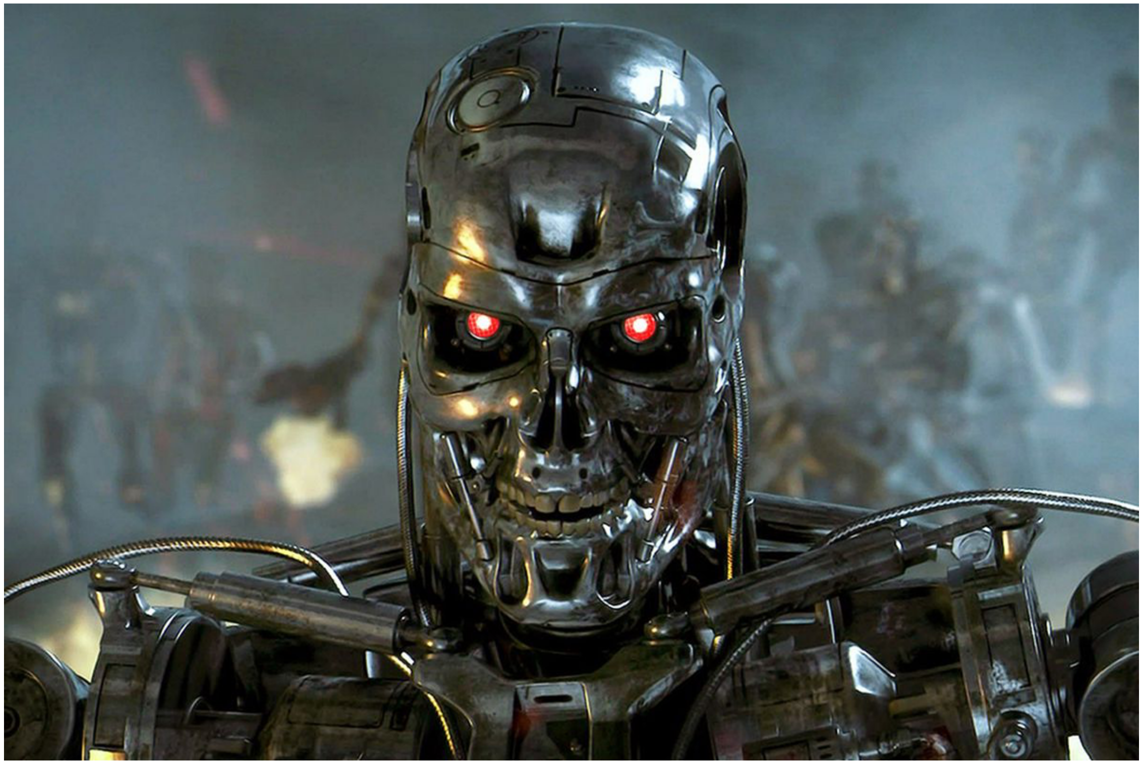
AI and jobs

- Keep it from happening?
  - Regulating automation, mandating new jobs
- Help the people who are displaced?
  - Investing in retraining programs
  - Policy solutions: welfare, universal basic income, redistribution of assets
- Revolution of human values?
  - “Netherlands is among the nations with the shortest work weeks, according to *CNN Money*. Netherlands has an average of 29 working hours per week and an average annual income amounting to $47,000.” – https://nltimes.nl/2013/07/11/worlds-shortest-work-weeks

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)

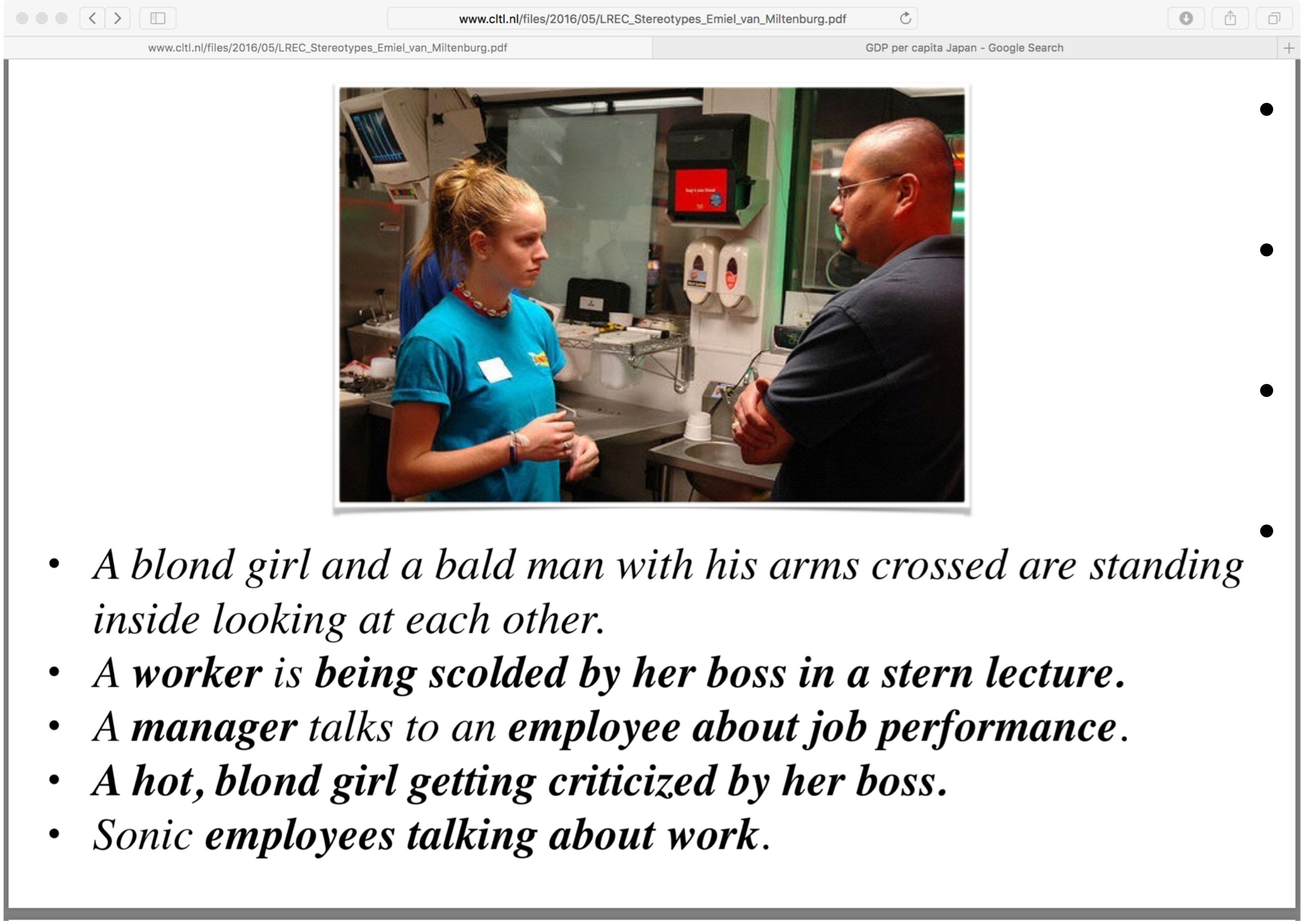
Autonomous AI



Accidental Replication of Human Bias

- Training an AI requires lots of data
- The data contains biases representative of the attitudes of the people who generated the data
- Without special care, the AI absorbs the bias

“Stereotyping and Bias in the Flickr30k Dataset,” Emiel van Miltenburg


www.cltl.nl/files/2016/05/LREC_Stereotypes_Emiel_van_Miltenburg.pdf
www.cltl.nl/files/2016/05/LREC_Stereotypes_Emiel_van_Miltenburg.pdf
GDP per capita Japan - Google Search

- *A blond girl and a bald man with his arms crossed are standing inside looking at each other.*
- *A **worker** is **being scolded by her boss in a stern lecture**.*
- *A **manager** talks to an **employee about job performance**.*
- ***A hot, blond girl getting criticized by her boss.***
- *Sonic **employees talking about work**.*

- Inferring status
  - “worker” vs. “boss”
- Inferring intentions
  - “being scolded”
- Disrespect
  - “girl” vs. “man”
- Marking the “less common” attribute
  - girl vs. boss
  - blond vs. brunette
  - “nurse” vs. “male nurse”

Bias caused by Data Sparsity

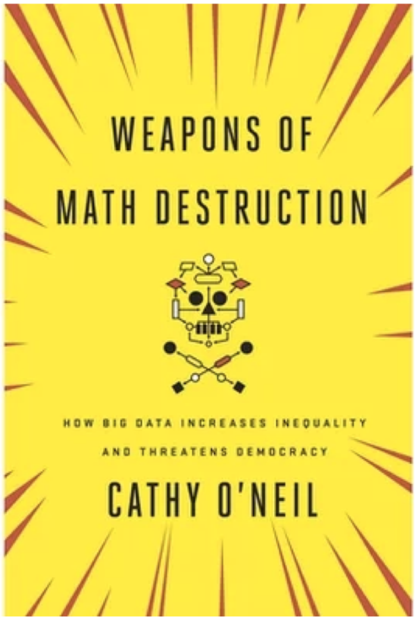
- Data contain more examples of one type than others, e.g., more Caucasians than African Americans
- Accuracy may be higher for the type that is better represented in the training data (minimize error by minimizing error for the majority case)
- Example: blacks more likely to be refused parole even if their prison records are the same (https://www.nytimes.com/2016/12/04/nyregion/new-york-prisons-inmates-parole-race.html)
- Example: tweets containing African American vernacular classified as “Danish,” and therefore excluded from automatic sentiment analysis (https://www.technologyreview.com/s/608619/ai-programs-are-learning-to-exclude-some-african-american-voices/)

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

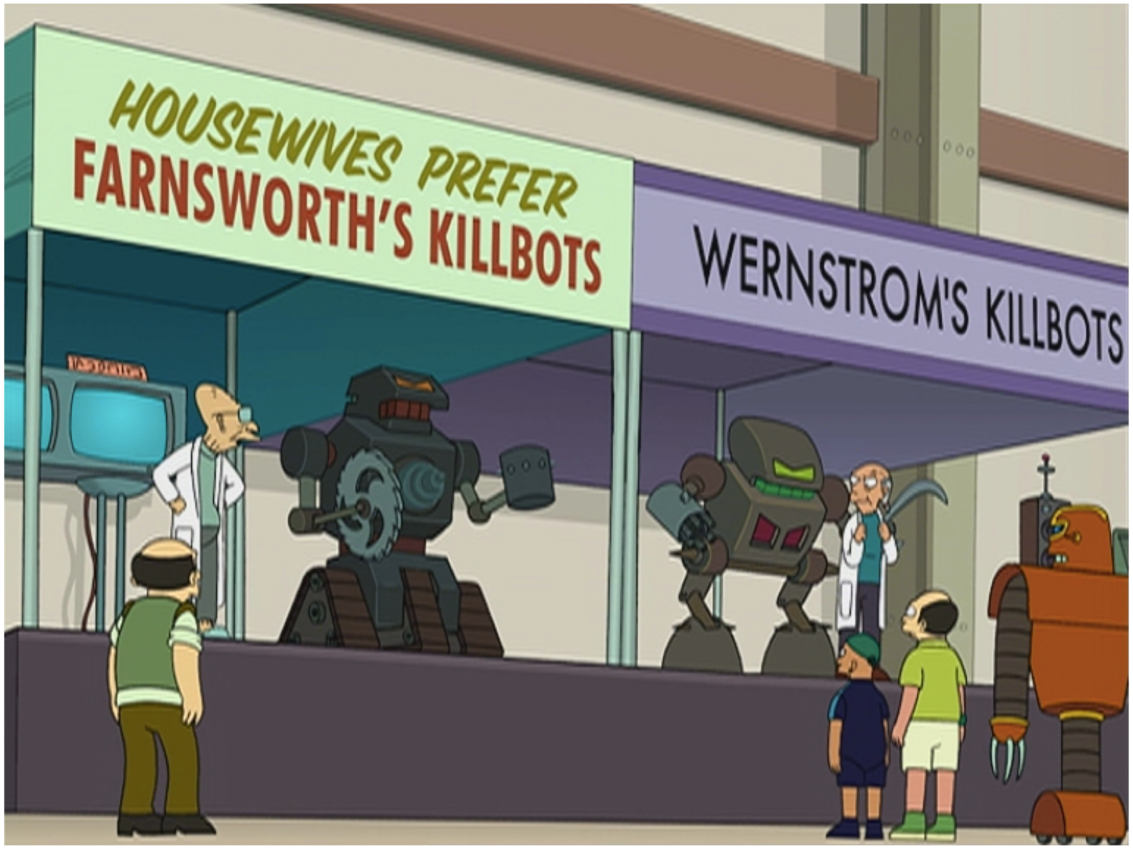


AI safety

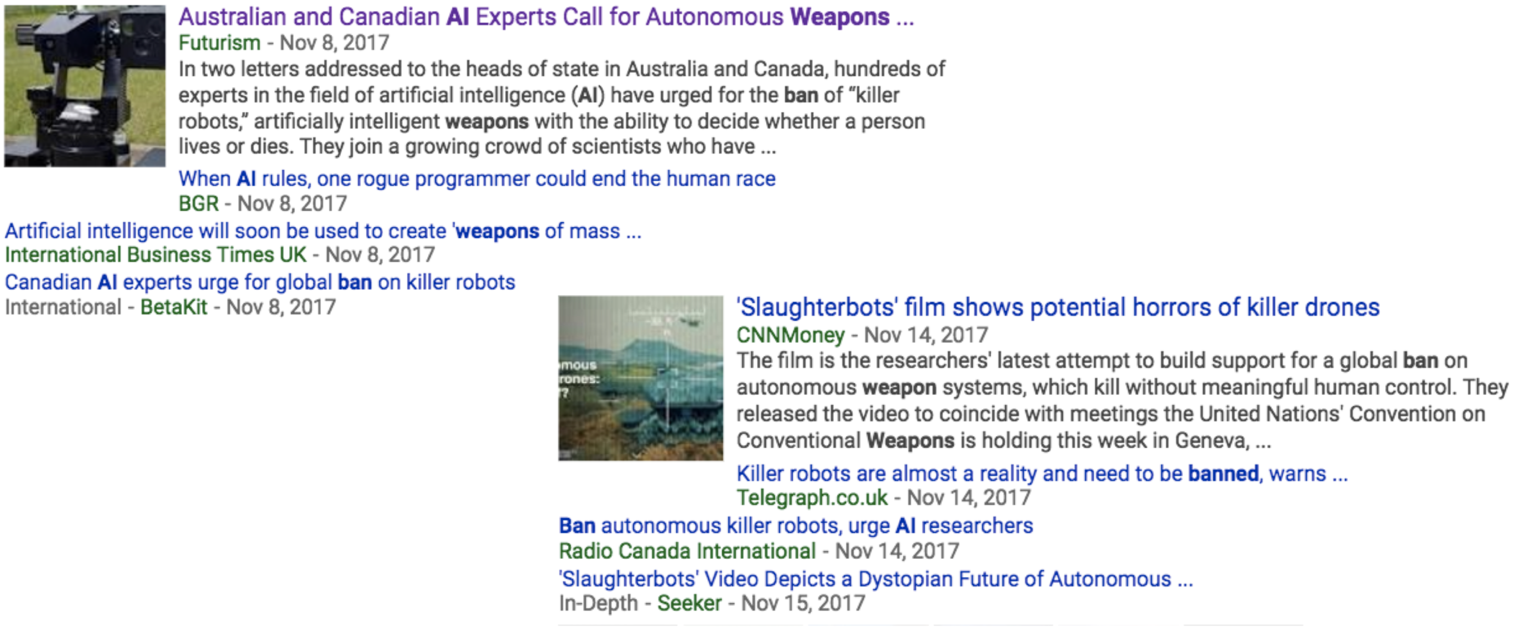
- Robustness to changes in data distribution
- Avoiding catastrophic “corner cases”
- Robustness to adversarial examples or attacks
- Avoiding negative side effects in reward function
- Avoiding “reward hacking”

- Reading: Concrete AI safety problems

AI weapons


HOUSEWIVES PREFER
FARNSWORTH'S KILLBOTS
WERNSTROM'S KILLBOTS

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

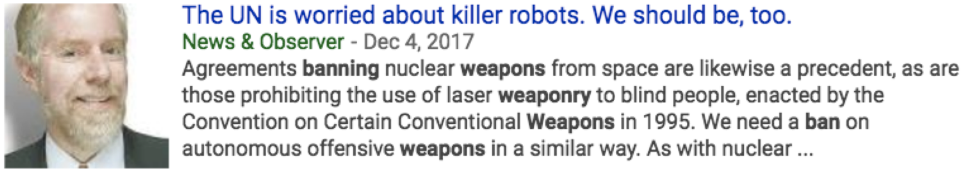
Canadian **AI** experts urge for global **ban** on killer robots
International - BetaKit - Nov 8, 2017

'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

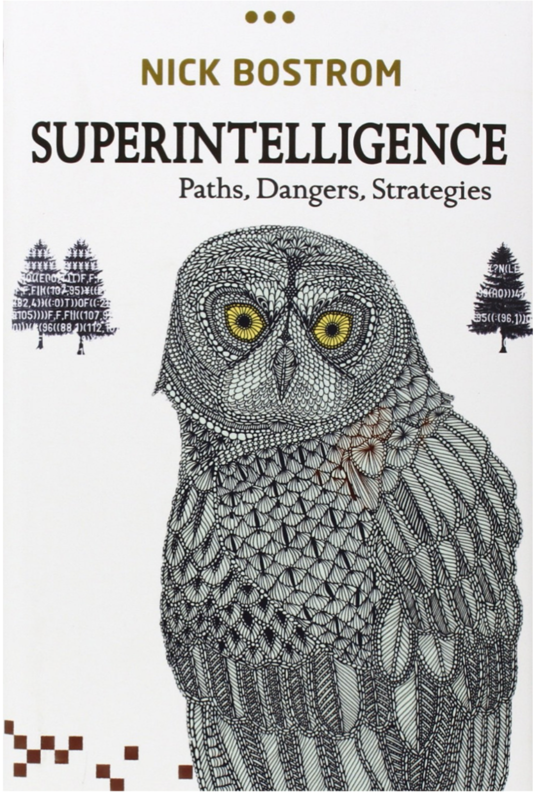


The UN is worried about killer robots. We should be, too.
News & Observer - Dec 4, 2017
Agreements **banning** nuclear **weapons** from space are likewise a precedent, as are those prohibiting the use of laser **weaponry** to blind people, enacted by the Convention on Certain Conventional **Weapons**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

Superintelligence



NICK BOSTROM

SUPERINTELLIGENCE

Paths, Dangers, Strategies

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

- Why we shouldn’t worry
  - The Extraterrestrials will kill us first
  - The Russians will kill us first
  - Environmental disaster will kill us first
  - AI is just a tool…
- Potential solutions
  - Asimov’s three laws of robotics: hard-coded limitations
  - Game theory: humans vs. AI? “Humans do not have general purpose minds, and neither will AIs.”

Superintelligence

- Readings
  - 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