Big data, machine learning & decision making
Nov 14 2018 @ CSAP Horn Fellows Meet....

The View from tech

Jon Crowcroft,

http://www.cl.cam.ac.uk/~jac22

How Big is Big Data?

- Coverage
  - High fraction of domain subjects (eg. UK NHS patient records)
- Too much for humans
  - (true since calculators took over from human computers)
- Complex/Complicated
  - More dimensions than you can shake a stick in/at
- Takes a while to get to grips with
  - Wrangling. 90% of effort
  - Missing record, variants, etc

Machine learning

- Not AI (in the sense of GI)
  - i..e not about doing what people do at all
- ML = stats with/on a computer
  - Regression, random forest, PCA
  - Possibly (but not commonly) artificial/convolutional/deep Neural Networks
  - Increasingly, even those are model trained (GANs/multitask etc)
- What could go wrong with Big Data + ML
  - Loss of human expertise
  - Bugs
  - Bias

Fairness, transparency, privacy

- Fairness
  - Sentencing in US court
  - Triple bias: chance of being arrested, being found guilty, being given custodial sentence, all correlated with race – latter is a cause of correlation. Feedback loop is a) wrong b) unethical
- Transparency
  - Do you understand why you weren't given a loan?
  - Counterfactual reasoning helps, but may not be enough
  - Depends on ML technology
- Privacy
  - How mich do **they** know about *you*?
  - Surveillance Capitalism (and Government)

Agency, legibility, negotiability

- Agency
  - Do you have seat at table at all?

- Legibility
  - Do you grok the data/ml/decision
  - Can you use your understanding?

- Negotiability
  - Do you get to say yes/no or more?
  - Did you make a decision, or did the decision make you?

Evidence based policy?

- Or eminence based policy?
- Or policy based evidence?

- Decisions often incorporate evidence resistant ideas
- Or can't be made at policy level anyhow

Any questions?

- Feel free to catch me here@lunch or any time
- Or e-mail me on: jon.crowcroft@cl.cam.ac.uk
- If you must: jcrowcroft@turing.ac.uk