Weaponising jon's research - ACS Transferable Skills 2025 - Retrospective Ethics

Jon Crowcroft,

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

Attendance Question:- "What had happened to AI before the story in Dune (book or films) unfolds or Is AI an existential threat to Huminity in Fictional work?"

Do no harm…

3 examples of normative ethics fail

1. DPI
2. DSI
3. CA
4. Oh, ok 4 - recombinant malware
5. Oh, ok 5 - contact tracing
6. Oh, ok 6 - tbd...

1. DPI

- Endace Deep Packet Inspection
- Measurement of ISP - Sprint Backbone:
- https://www.cl.cam.ac.uk/~jac22/out/konstantina-papagiannaki.pdf
- >NSA (surveillance)
- https://en.wikipedia.org/wiki/The_Snowden_Files

Could have underpowered capture

- But would have been unsuccessful business
  - Not led to innovation in protocols
  - And sunk without trace (pun intended)
- Could have controlled product via patent.
  - And had corporate ethical position…
  - If we'd thought of it

2.DSI

- Distributed Simulation Internet
- Multicast&realtime multimedia
- https://www.cl.cam.ac.uk/~jac22/otalks/ballardie-thesis.ps.Z
- >DSI (war)
- https://www.ncbi.nlm.nih.gov/pubmed/7643020

Dual use is Internet #101

- Hard to do a lot about
  - Much comms (and crypto) has dual use origin
- Arguably, training soldiers better, might lead to less violent death
  - Personal choice…

- Pugwash: nuclear weapon research self-ban

3. CA

- MyPersonality>Cambridge Analytica
- Facebook graphs&processes:
- https://www.cl.cam.ac.uk/research/srg/netos/papers/p955-quercia.pdf
- >CA (democracy)
- https://www.theguardian.com/news/2018/mar/26/the-cambridge-analytica-files-the-story-so-far

Definitely foreseeable

- Much written about re-purpose of tech
  - Author has read books on topic
  - Should have thought about control of results
  - Or at least put out a warning..

4. Recombinant malware

- Vigilante observes behaviour of malare & builds self-certifying alert+patch

https://www.microsoft.com/en-us/research/publication/vigilante-end-to-end-containment-of-internet-worms/

- So now the bad guy doesn't just have polymorphic worms:-

Can build recombinant malware…bad bad bad

https://www.nature.com/articles/455290a

This example was mitigated

- So while potential for harm is high,
  - Solutions in place
- General solution maybe Asilomar Protocols?

5. Contact Tracing Apps

- Back 16 years, we wrote FluPhone
  - Detects proximity of other people via Bluetooth
  - Uploads contacts to secure site
  - used to track infection exposure during H1N1 epidemic
  - Informed consent, data private
- What's the problem?

centralised data for contact trace

- could be appropriated and misused
  - e.g. compliance checking or later, commercial exploitation
    - like fb friend list only worse:)
  - Enter GAEN decentralised solution
    - Why trust google/apple more than NHS?
    - fails to deliver on epidemiology data

What to do about it?

- Not do research?
- Warn people about poor uses of ideas?
- Patent/protect tech to control?
- Something new?
- Pugwash for Computing

What could possibly go wrong…?

- Digital Identity System deployment +
- Evaluating operational readiness using chaos engineering simulations on Kubernetes architecture in Big Data

  https://ieeexplore.ieee.org/document/9993998

Attendance Q: Does SF AI actually spell the end of humanity?

- And the answer is:
  - Not in Terminator, I Robot/Foundation, Humans
  - Not in Colossus, WarGames
  - And only in individual cases in
  - Ex Machina, Westworld, M3GAN, Alien
  - Robbie the Robot is your friend
    - (Forbidden Planet, Lost in Space)

6. AI example of Good&Bad

- Say we construct a Synthetic Face Generator (e.g. DCGAN)
  - Train on real (specific) people
  - For we think "good" reasons – e.g.
- Use to find a face that challenges limits of
  - Automatic face recognition compared with
  - Human face recognition
- Illustrate visually the limits of algorithms and humans
  - E.g. unreliable witnesses and police use for arrests
  - Learn about neurodiversity (prosopagnosia)

What could possibly go wrong? Who are the adversary?