Patient data used for Artificial Intelligence in health care.

What are the ethical issues ?

Otago University, Wellington
Public Health Summer School
Wellington
15 February 2019
Rochelle Style

Overview

- Introduction - Industry 4.0 and AI healthcare examples
- AI challenges
- Society 4.0 – moral machine
- Bias, Transparency, Dignity, Respect and Autonomy
- Regulation of secondary use + waiver of consent
- Explainability , the IoT + data ownership/data rights
- Ethics by design, Complementarity and transdisciplinary collaboration

Industry 4.0

- The 4th Industrial Revolution: a range of new technologies fusing the physical, digital and biological worlds – including artificial intelligence (Schwab).

- Impacting all sectors including healthcare, employment, education, criminal justice, art …

$$\min_G \max_D E_x[\log(D(x))] + E_z[\log(1 - D(G(z)))]$$

AI Health examples

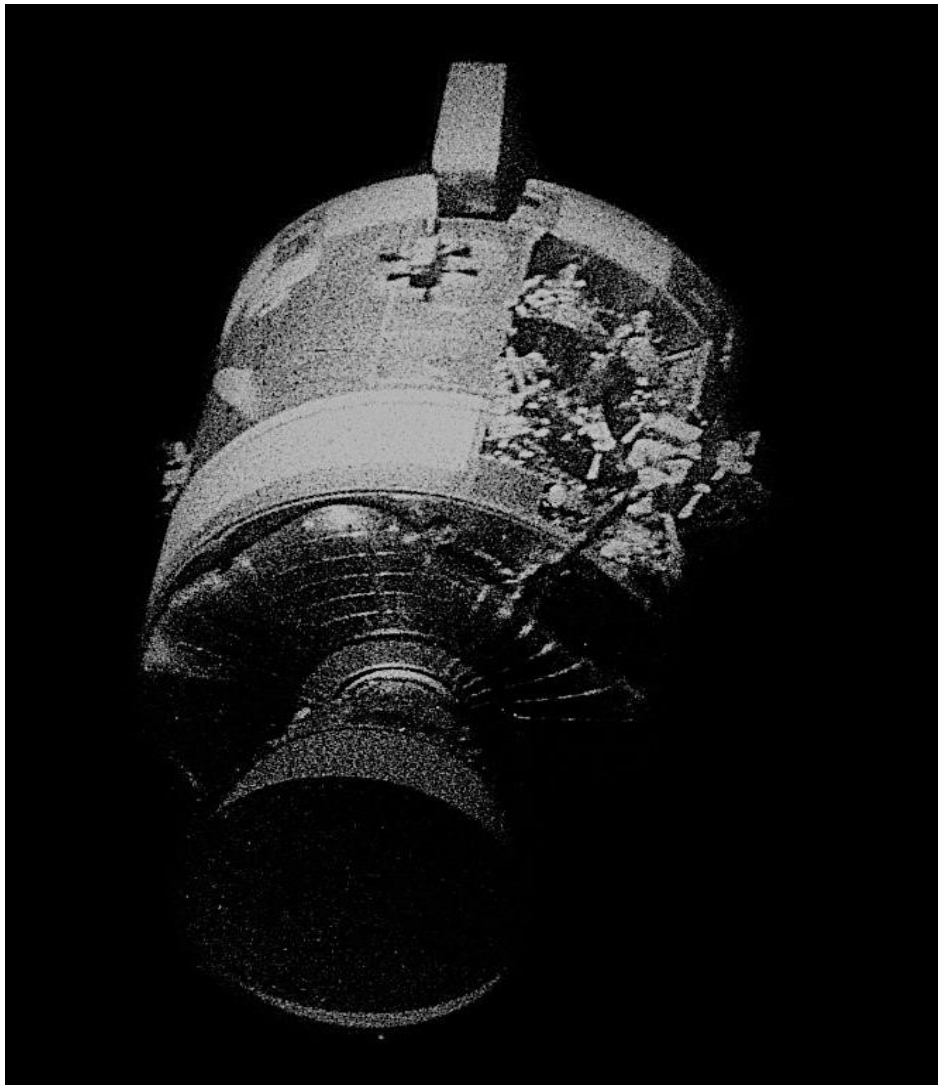
- **Deep Patient:** data from > 8 m patients (7 hospitals in New York) using AI to assess probability of patients developing various diseases. Predictions particularly good for severe diabetes, schizophrenia and some cancers.
- **The AI Clinician:** selects optimal treatment strategies for sepsis in ICU using data from > 100,000 patients from > 130 ICUs over 15 years. AI treatment on average reliably higher than human clinicians.

- **DeloitteAssist:** Canterbury DHB pilots AI-enabled voice assistance.

AI challenges – Houston, we have a problem

Ethical, Social & Political challenges

- Fairness: bias
- Accuracy, reliability & safety
- Transparency
- Dignity, respect & autonomy
- Trust
- Privacy & security
- Accountability & responsibility

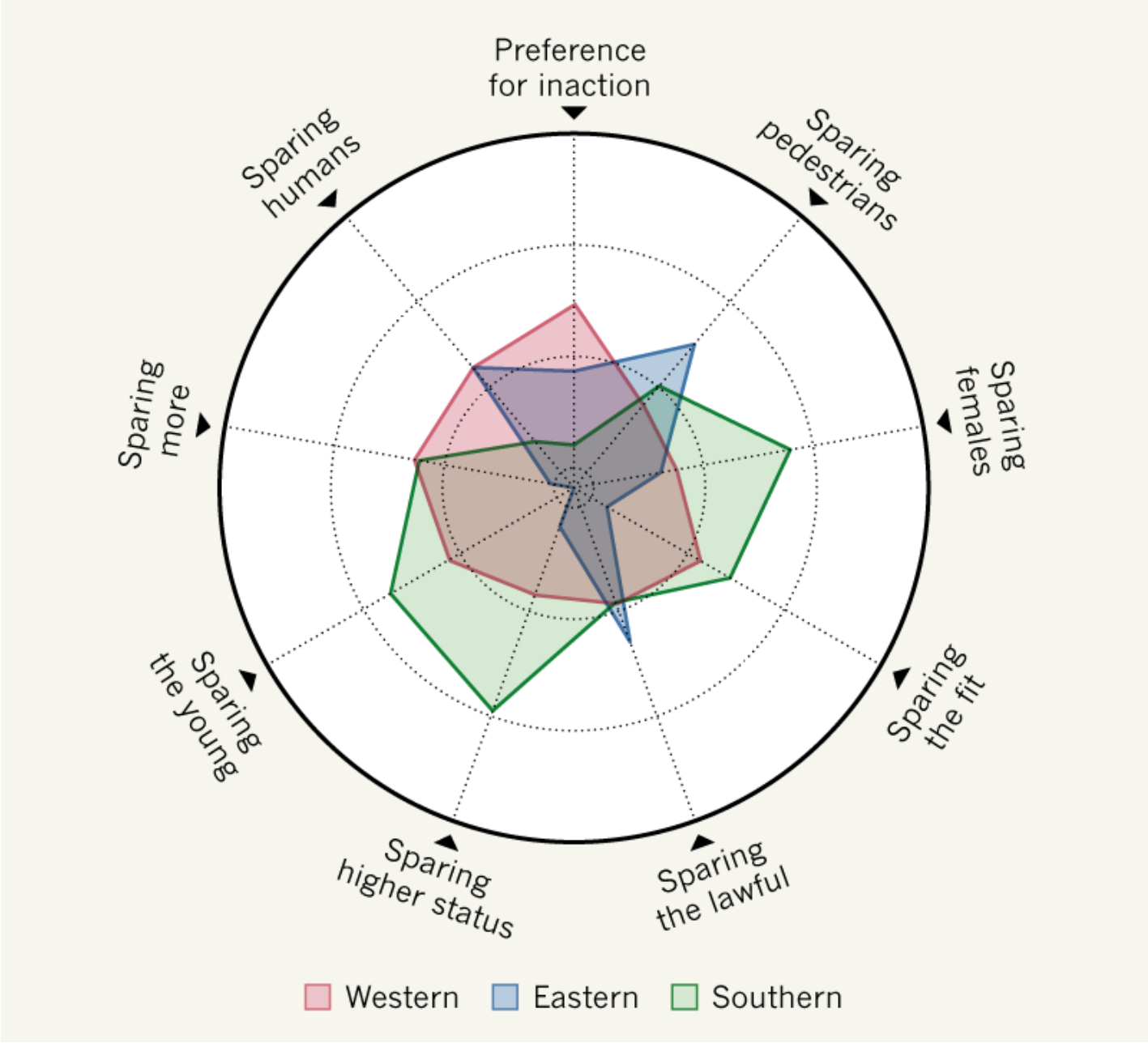


Society 4.0

- Doing nothing is not an option – maybe unethical not to use AI
- Industry 4.0 should be underwritten by values that ensure these technologies are trained towards the social good
- This means using data ethically, involving citizens in the process and building social values into the design

Moral Machine

- MIT's Moral Machine: Survey 2.3 million across countries.
- 13 scenarios of an unavoidable collision involving self-driving car killing various combinations of passengers.
- Variations in moral principles guiding drivers' decisions



Source: http://moralmachine.mit.edu/

Challenge: Bias – AI’s Achilles' heel

- Patient data is the fuel for AI.
- Training data used in AI may reflect and reinforce biases leading to discrimination based on gender, ethnicity, disability and age.

Challenge: Bias

- “AI, Ain’t I a woman” (Buolamwini)
- Error rates as high as 35% for darker-skinned women; 7% for lighter-skinned women, and no more than 1% for lighter-skinned men

- AI reflects the “coded gaze” - the priorities, preferences, and at times prejudices of those who shape technology

Challenge: Bias

- Watson for Oncology - patient data sets used for training matter
- *Sometimes respecting people means making sure your systems are inclusive such as in the case of using AI for precision medicine, at times it means respecting people's privacy by not collecting any data, and it always means respecting the dignity of an individual.* (Buolamwini)
- Optimisation – for false negatives or false positives ? Err on the side of caution and over-diagnose?
- Recidivism

Challenge: Bias

- Humans are biased too – parole decisions after lunch

- Bias in clinical record keeping of patients' data affects quality of the AI fuel:

  - subjective judgments about what is important.
  - separation between what is observed by the doctor & communicated by the patient; and what is recorded.
  - more serious illness > accurately documented - clinician bias (increased attention) and patient recall bias.
  - doctors' unconscious bias may impact on perceptions of patients' illness

Challenge: Transparency

- AI “Black Boxes”

- Dudley: Deep Patient is “a bit puzzling” –we can build the models but we don’t know why they’re so good at predicting the onset of diseases like schizophrenia

Challenge: dignity, respect, autonomy

- Ethical principles require respect for individuals' autonomy
- Individual autonomy may be impacted if treatment choices are restricted by machine calculations about risks/benefits which may be opaque

- Patients' rights to make free, informed decisions about their health may be compromised if a doctor can't explain how AI systems make a diagnosis or devise treatment plans

Secondary use of patient data

- Use of health information is restricted for purposes other than that for which it was collected, subject to a research exception (HIPC)

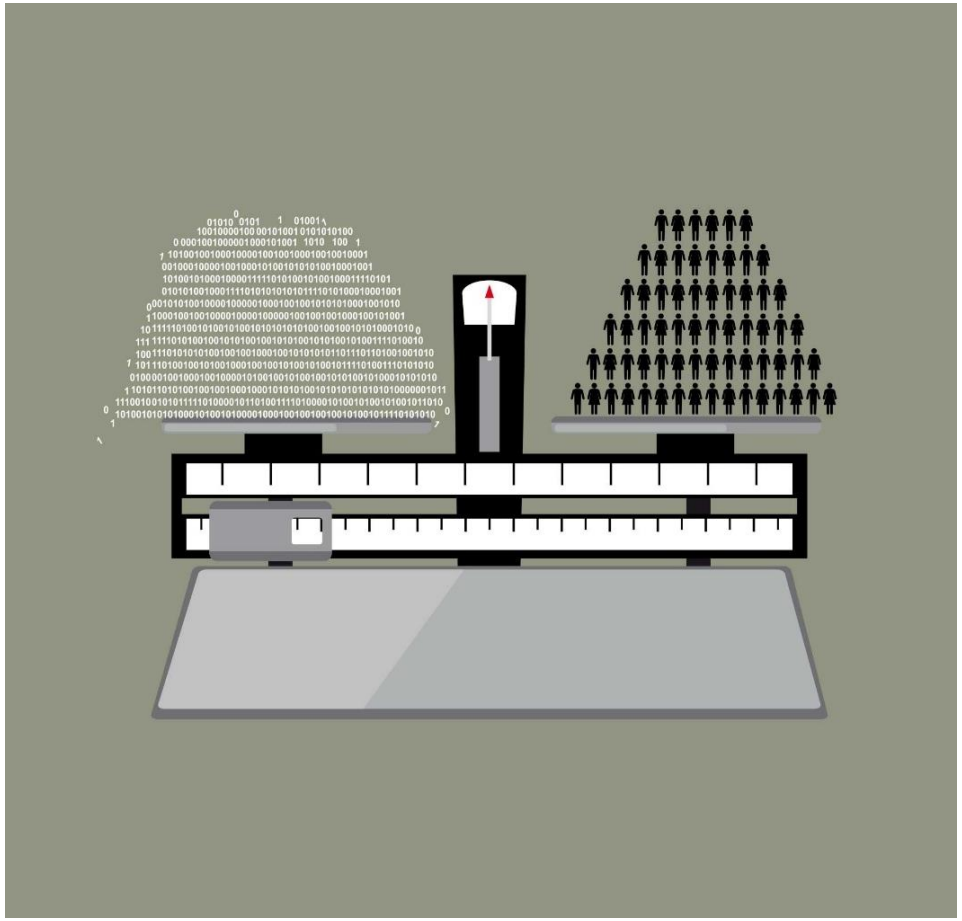
- Transparency is required, ensuring an individual is aware of the fact that identifiable health information is being collected, the purpose(s) of collection and the intended recipients (HIPC)

Secondary use of patient data

- Ordinary people who are the sources of the information should not be surprised about the way in which their data is being used. They should grasp that their personal information is being used/disclosed for purposes which they have reason to both expect and accept

Draft NEAC standards

- Researchers must identify the possible **benefits and risks of harm** of data use and carefully **balance** them against each other and consider how to **minimise and mitigate** any harms of data use



Consent

- Should people be consented for participation in research by sharing their data to fuel AI?
- *Streams*

- Patients would not have expected their information to be used in that way.
- *"The price of innovation does not need to be the erosion of fundamental privacy rights"*
- Models of Consent? Opt-in? Opt-out? Dynamic.

Waiver of consent to data use must be justified

- possible benefits outweigh possible harms, including to any participant, whānau, iwi and any other groups.
- no known or likely reason to expect that the participant(s) would not have consented if they had been asked.
- appropriate consultation has been undertaken with cultural or other relevant groups and those consulted support the proposed use.

DRAFT

Explainable AI (XAI)

- How do we explain to patients about their participation in AI healthcare?

- Clear, simple and easy-to-understand language (Stats NZ & OPC, GDPR, AI Universal Guidelines)

- To whom?

> *Finite, ignorant and vulnerable agents with limited cognitive capacities, limited abilities to choose and limited time within which to choose [and we should not] be expected to perform heroic or even impossible cognitive feats of [understanding];* (O'Neill)

XAI – what?

- What should be explained?
  - The statistical models used and how the training set has changed over time?
  - That bias has been detected in the dataset?
  - That the doctors don't know how the model works because it is a "black box"?
  - What the model is optimised for – eg, cost saving or saving lives?

Internet of Things

- Offer different treatment options and lower healthcare costs. Plethora of data collected.
- Medibank *LiveBetter* app:
  - collects data from social media accounts
  - collects “non-personal data”
  - uses tracking technologies & Google analytics which “promises insights and machine learning capabilities”
  - collected data supports research and may be used to develop & deliver new products
- Digital nudges: Clinicians and others getting AI nudges may trust them based on heuristics (eg, automation bias) rather than sound insights.

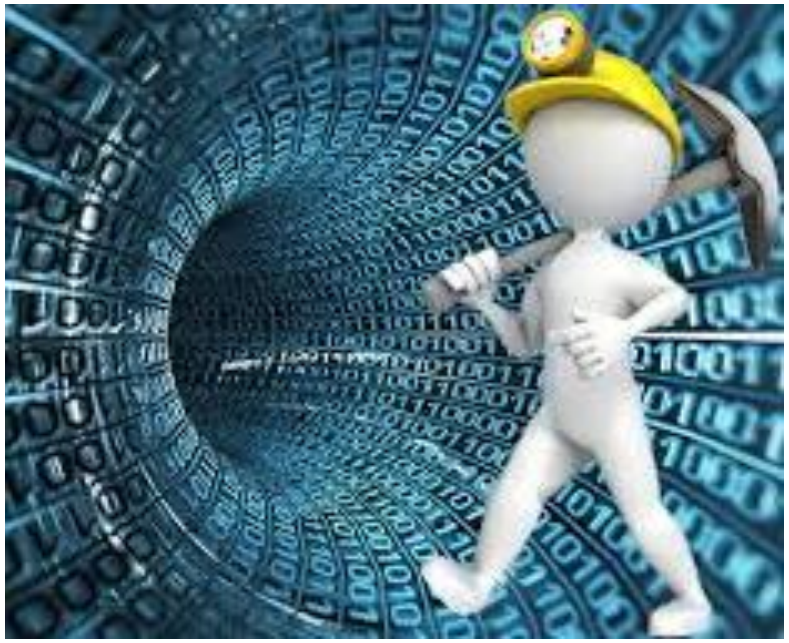
Data ownership

- Data ownership or data rights balance public v private?

- Future imagined world of data vaults:

  - Storing others' personal data for profit outlawed
  - Disputes decided by digital society specialty courts.
  - Regtech algorithms assess how algorithms accessing personal data vaults be harmful to personal or societal best interests, leading to poverty, civil unrest …. . (Cahan)

Data mine-ing



Data rights

- Data ownership is flawed (Tisne). Property frameworks vulnerable to the illusion of "consent."
- Algorithms ‘bucketize’ people – “ heavy smoker with a drink habit” or “healthy runner, always on time”.
- Algorithm unfair to an individual – wrongly assess 99% chance committing another crime based on people demographically similar to you BUT you can’t own your demographic profile - “owning” your data won’t make it fair.
- Extrapolations can be made about you.
- Anonymised information for the collective good

A new paradigm

- *Relationships between citizens, the state, and the private sector have changed in the data era. A new paradigm for understanding what data is—and what rights pertain to it—is urgently needed if we are to forge an equitable 21st-century polity.* (Tisne)

- *Consumerism has a very different endpoint to citizenship and there is something about the intermingling of those endpoints, those narratives, that makes it easier to violate trust.* (Genevieve Bell)

Ethics by design

- *If you don't want to block innovation, it is better to frame it by design within ethical and philosophical boundaries* (Macron)

- Tools: AI ethical frameworks and matrices, Algorithmic Impact Assessments

Ethical oversight

- Algorithmists?
- FDA for Algorithms?
- Data Advisory Panels?
- Specialist HDECs?

Collaborative, diverse and transdisciplinary

- 22,000 AI researchers with PhDs. “This is where it starts to get worrying”. “Pale, male data issue” (Buolamwini)

- Computer scientists, engineers, mathematicians, statisticians, data scientists, social scientists, economists, lawyers and ethicists – DIVERSE

- Maori Data Sovereignty Network

TE MANA RARAUNGA

Māori Data Sovereignty Network

Human + Machine Complementarity

- Debate about the Compas algorithm's predication of recidivism: are the algorithm's predictions better than humans?
- Dressel & Farid research:
  - Individual humans got it right 63 % of their time (pooled 67%)
  - COMPAS' accuracy = 65%
- Complementarity: Humans and machines have weaknesses and complementary abilities suggesting benefits from hybrid models that combine machine and humans decisions. (Cornell, MIT + Microsoft)

Final Challenge

- *The biggest challenge is to make sure that the futures we choose don't ignore what makes us human, and to ensure that technology doesn't undermine democracy and reinforce inequity.* (Genevieve Bell)