COSC244

Lecture 12: Data Ethics

Alistair Knott

Dept. of Computer Science, University of Otago

Ethics (and Law) in Computer Science

There are several different areas in CS where ethical issues arise.

- Data ethics:
  - What *personal data* should you be able to obtain / use / store?
  - What data should you *disseminate* / *make public*?
- Artificial Intelligence (AI) ethics:
  - What smart algorithms should you be able to run on people's data? How should we ensure these algorithms are *fair*, and *transparent*?
- Internet ethics:
  - How should you behave on the internet? (As a user/provider of information?)
  - How should behaviour on the internet be *regulated*?
- Intellectual Property (IP) law:
  - How can people *protect* the works they create? (Esp. software...)
  - How can people *disseminate* their works to others?

Ethics/Law in our CS curriculum

- Computers, ethics and society (COMP150)
- Internet ethics/law (COMP112)
- AI ethics (COSC343, COSC420)
- Machine Learning ethics (COSC471)
- Computer vision ethics (COSC450)
- Data ethics/law (COSC244)
- Intellectual Property law (COMP112, COSC345)

Ethics and Law

Nearly all the questions I posed at the start use the word ‘should’.

- What personal data should you be able to obtain / use / store?
- What data should you disseminate / make public?
- What algorithms should you be able to run on people’s data?
- How should we ensure these algorithms are fair, and transparent?
- How should you behave on the internet?
- How should behaviour on the internet be regulated?

The word ‘should’ can be understood in two ways:

- What is required *ethically*? (I.e. what’s right???)
- What is required *legally*? (In some particular country.)

In today's lecture

I'll discuss two topics today.

- One relates to privacy: people have certain rights to privacy...
- One relates to freedom of information: people have certain rights to know things.

I'll first discuss the ethical issues surrounding these topics...
Then I'll discuss how they're dealt with in NZ Law.

Privacy and freedom of information: Ethical issues

Can you give some examples of information you should be allowed to keep private?

Can you give some examples of information you are entitled to know?

Ethical issues: Consent

In practice, the government (and companies) very often want to use your personal data. And you might well want them to.

A way to manage this is for these organisations to ask for your consent.

Consent is an important ethical principle:

- You might *want* someone to infringe on your privacy. . .
- To negotiate this, they need your consent.

Ethical issues: conflicts between privacy and freedom

Can you think of cases where the right to privacy *conflicts* with the right to freedom of information?

Can you think of ways we might *resolve* these conflicts?

Ethical issues: conflicts between privacy and profit

Personal data is *extremely valuable*! Increasingly so.

- Many internet companies see customer data as their principal commercial asset. (For example?)

  If you *consent* to a company exploiting your data commercially, they can do it. (But do users know what they're consenting to?)

- Government agencies also have a lot of commercially valuable data. (For example?)

Conflicts between privacy and the ‘social good’

Covid monitoring apps are a good current example.

- If everyone uses monitoring apps, this helps control the epidemic. . .
- But contact tracing involves transmission of personal data. (E.g. location data.)

Do we want to compromise privacy for the sake of public health?

Covid monitoring apps: (I) Use

Monitoring apps have been used for two purposes in the covid epidemic:

- For quarantine enforcement (in self-isolation)
- For contact-tracing (of all citizens).

For both kinds of app, use can be:

- Voluntary
  Citizens can choose to download an app (or choose not to)
- Obligatory
  Phone locations are monitored for all citizens
  - E.g. in South Korea, for quarantine
  - E.g. in China, for quarantine and contact tracing

Covid monitoring apps: (II) Information gathering

Two methods for *gathering* information:

- Active: The user has to 'sign in' (e.g. to shops, cafes, workplaces)
  Venues have to collaborate, by providing a QR code
- Passive: contact registering happens automatically
  E.g. through bluetooth, phone localisation

In passive methods, there are two ways of registering contact:

- Indirect: people register at businesses, workplaces
  E.g. NZCOVIDTracker, Rippl
- Direct: contacts between people are registered 'peer-to-peer'
  (Only possible with bluetooth)
  E.g. The Australian govt's 'Covidsafe' app

Covid monitoring apps: (III) Information storage

Two basic ways of *storing* contact data:

- Local storage:
  - Each person's contacts are stored on their own phone
  - Other people are *anonymised* in these records
  - It's normally the user's choice to report sickness
  - Contact information is only accessed centrally when tracing contacts of people known to be sick
  - Positive contacts can be reported centrally...
  - or only shown to users
- Centralised storage
  - Everyone's contacts are stored on one (government) database
  - We don't rely on users reporting sickness
  - Single point of failure...
  - Will the government use this database *for other purposes*?

Covid monitoring apps: (IV) Public or private?

In April, Google and Apple announced changes to Android & iOS to enable bluetooth-based (local) contact-tracing.

- Their system is called the exposure notification framework (ENF).
- It's *voluntary*, *local*, *passive*, *direct*.
- Users have to choose to report illness.

There are open-source equivalents: e.g. TCN ('temporary contact numbers').

These solutions are hackable—but large-scale hacking seems hard.

Who do you trust more with your health data:
Governments, or big tech companies?

Privacy and freedom of information in NZ Law

As regards privacy (and consent), the key NZ act is the Privacy Act (1993), governing the use of *personal data*, by private or public agencies. Here are the key points:

- There must be a *lawful purpose* behind gathering the data, connected to the agency's function. (Data can't be gathered 'for no purpose'.)
- Personal data must be gathered 'directly' from the person—and their consent is required. (But consent is *implied* if the person is *aware* that the data is being gathered, and they don't object.)
- Transparency: the agency is obliged to make the person aware that their data is being gathered—and aware of the purpose for which it's being gathered.
- The *method* by which data is collected must not be 'unreasonably intrusive'.

Privacy and freedom of information in NZ Law

The Privacy act (1993) continued:

- Security: there must be methods in place to *protect* the data that has been gathered (keep it safe).
- Data should not be used without *checking it for accuracy* (in cases where it influences important decisions, e.g. credit approval).
- The agency won't *keep* the data for longer than is needed (for the purpose it was collected for).

There's a Privacy Commissioner who oversees compliance to this Act (who people can complain to).

- Liability under the Act requires a breach of some principle, but *also 'some loss or harm to the individual'*. Harm can include humiliation, loss of dignity, hurt feelings but must be 'significant'.

Privacy and freedom of information in NZ Law

As regards freedom of information, the key NZ act is the Official Information Act (1982), covering information held by *government agencies*.

- The basic principle is that *information shall be made available unless there is good reason for withholding it*.
- Information can be withheld on several grounds—in particular, if it's necessary to protect the privacy of individuals.

Open data initiatives in NZ

There's a push in NZ (and other countries) towards 'open data' in government. (Suitably anonymised, of course.)

In NZ, this is done under the Open Government Data Programme.

- NZ is ranked 6th in the world in the Open Data Barometer global rankings 'for readiness, implementation and impact of open data'.

NZ's integrated data infrastructure project

NZ is also a global leader in 'integrated' government data.

The Integrated Data Infrastructure (IDI) project is a large database of information (over 166 billion facts) about people in NZ. It's drawn from several areas:

- Stats NZ (the census), Health, Housing, Justice, Education, Work and Income...

Data for individuals is *linked* across areas. (The Privacy and Statistics acts require linked data to be anonymised—which can be quite tricky to do.)

The IDI is only used 'for approved research projects that are in the public interest'. (But there's some talk that we might sell the data to other agencies—e.g. big healthcare providers. Is that okay?)

Integrated data in other countries

In the US, the rules about privacy are less restrictive: a lot of data integration is done by commercial companies.

- 'In the US, almost all personal data is for sale. For example, if you want to know where Jewish women live, you can simply buy this information, phone numbers included.'
  (Grassegger and Krogerus, 2017)

In China, the rules about privacy allow government agencies to gather/integrate personal data very freely.

- The Chinese government aims to use its data-gathering powers to position itself as a global leader in AI.

Bias in government data

It's quite easy for *bias* to creep into datasets (both public and private).

This bias can be self-perpetuating. Consider:

- Police have more patrols in Māori areas / poor areas.
- So they find more crime in these areas.
- So crime data disproportionately feature these areas.
- So police have more patrols in these areas...

Māori groups are responding to this issue with an interesting initiative: the Data Sovereignty Programme.

- This programme advocates for Māori involvement in the governance of data repositories, and for processes that ensure the quality of data about Māori. (The IDI is a particular focus.)

Australia's anti-encryption law

In Dec 2018, Australia passed a new law: any organisation which hosts data in Australia is required to give authorities access to their IT system, if the government requests it.

- That might mean they have to provide a backdoor for encrypted data.
- Amazon's local AWS cloud platform is based in Sydney: it hosts data for many of NZ's biggest organisations (e.g. Xero, Orion Health).

What are the benefits and risks in giving the government a back door to encrypted data?

Some URLs

NZ's Privacy Act:
`https://www.privacy.org.nz/the-privacy-act-and-codes/the-privacy-act`

NZ Government's Open Data Programme:
`https://www.data.govt.nz/`

NZ's Integrated Data Infrastructure:
`http://archive.stats.govt.nz/browse_for_stats/snapshots-of-nz/integrated-data-infrastructure.aspx`

The Māori Data Sovereignty network:
`https://www.temanararaunga.maori.nz/`