Week 11: Data Ethics

AF3214 Python Programming for Accounting and Finance

Vincent Y. Zhuang, Ph.D.
vincent.zhuang@polyu.edu.hk

School of Accounting and Finance
The Hong Kong Polytechnic University

R508, 8:30 am – 11:20 am, Wednesdays, Semester 2, AY 2024-25

What is Ethics and What is Computer Ethics?

- **Ethics** tells us about right and wrong
  - ➢ They are shared value / societal rule
  - ➢ Ethics is not law

- Wikipedia defines **computer ethics** as "a part of practical philosophy (concerned with) … how *computing professionals* should make decisions regarding professional and social conduct."

- James Moor, a famous Computer Science Professor, defined it as "the analysis of the nature and societal impact of computer technology and the corresponding formulation and justification of policies for the ethical use of such technology. "

- These definitions suggest a *strong tie between ethics and professional conduct* and an approach for influencing that conduct through *policies and rules*.
  - – Determine what are those policies and rules

Retracted: Risk Management in Financial Institutions - 2021 - The Journal of Finance:
https://onlinelibrary.wiley.com/doi/10.1111/jofi.13064
https://afajof.org/2021/07/answers-to-faq-about-the-recent-retraction-of-an-article-in-the-jf/ 2

Codes of Ethics and What is Data Ethics?

- **Specificity** in Big Data and Analytics:

    ➢ General codes of ethics and codes of conduct are *not* aimed at issues of concern with big data and analytics.

But…

- Data ethics (or big data ethics) refers to systemizing, defending, and recommending concepts of right and wrong conduct in relation to **data**, in particular **personal data**.

Let's look at some examples

Informed Consent and Data Ownership

Informed Consent

- Human Subject

  - ➤ Must be informed about the experiment
  - ➤ Must consent to the experiment voluntarily
  - ➤ Must have the right to withdraw consent at any time

Data Ownership

- Most of the time you don't own the data about you. The data belongs to the company who collected it.

- However, we might have some control over these data that aren't ours because they are about us

- We need to create the principle to reason about this control and that's the main concern of a discussion about the right to privacy.

- If the company goes bankrupt, another company buying it should keep the same privacy.

https://www.immd.gov.hk/eng/forms/forms/ops003.html

- Privacy, of course is the first concern that comes to so many minds when we talk about Big Data.

- How do we get the value we would like by collecting, linking, and analyzing data, while at the same time avoiding the harms that can occur due to data about us being collected, linked, and analyzed?

Privacy

- Can we define reasonable rules that all would agree to?
- How can we maintain anonymity? Is it very hard?
- People have different privacy boundaries.
- But different boundaries do not mean no boundary.
- As society adapts to new technologies, attitude can change.

https://breachforums.st

darknet

No option to exit

- In the past, one could get a fresh start by:
  - ➢ moving to a new place
  - ➢ waiting till the past fades (reputation can rebuild over time)
- Big Data is universal and never forgets
  - ➢ Even worse, now we have Blockchain…
- Data Science results in major asymmetries in knowledge

Wayback Machine and Right to be forgotten

- Archives pages on the web (https://archive.org/web/ - 300 billion pages saved over time)

  - almost everything that is accessible
  - should be retained forever

- If you have an unflattering page written about you, it will survive forever in the archive (even if the original is removed)

Right to be forgotten

- Laws are often written to clear a person's record after some years.

- Impacts search engines (not removed completely but hard to find)

Collection vs Use

- Privacy is usually harmed only upon use of data

- Collection is a necessary first step before use

- Existence of collection can quickly lead to use

- But collection without use may sometimes be right

    - e.g., surveillance

① JY-M730 显示器
② JY-D13 球型摄像头
③ JY-861 双镜头摄像头
④ JY-6732 后视摄像头
⑤ JY-660 侧视摄像头
⑥ JY-660 侧视摄像头
⑦ JY-810 前视摄像头
⑧ JY-D16 球型摄像头
⑨ JY-PT08 停车传感器
⑩ JY-DV820 MDVR
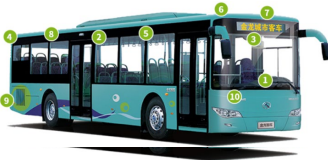
➢ By the time you know what you need, it is too late to go back

Loss of Privacy

- Due to loss of control over personal data

- We may be OK with others having certain data about us that we have chosen to share with others or that is public, but we really do not want others to share our data in ways that we do not approve.

https://gigazine.net/gsc_news/en/20240219-airline-bad-review-posts-passport-aero-dili/

Privacy is a basic human need, even for people who have nothing to hide.

https://source.android.com/docs/core/connect/wifi-mac-randomization-behavior

Sneaky Mobile App

- There was a time when App didn't tell you what kind of data they were collecting
- Many apps ask for far more permissions that they need
  - Might be used for future functionality
  - But most of the time just for adware
    - Picture management app that needs your location

wechat - Android Apps on Google Play

抖音Douyin- Apps on Google Play   Douyin – Chinese version of Tiktok

TikTok - Apps on Google Play

Xiaohongshu – A platform that combining fb, ig, X, Tiktok, YouTube

小红书 – 你的生活指南 - Apps on Google Play

Sneaky Web/Mobile App

Many transactions need ID

- You must provide an address to receive goods
- You must give your name for travel booking
- You must reveal your location to get cellphone service
- You must disclose details of your health care and lifestyle to get effective medical care

Welcome to Airchina!

Digital Footprints

<u>Enough history tells all</u>

- Search pattern for person can reveal your identity

- If we have a log of all your web searches over some period, we can form a very good idea of who you are, and quite likely identify you

  - Digital footprints:
    - https://support.google.com/accounts/answer/465?hl=en&co=GENIE.Platform%3DDesktop&oco=0

<u>De-identification</u>

- Given zip code, birth date, and sex, about 87% of Social Security Numbers (i.e., SSN) can be determined uniquely.

- Similar to Chinese Identity Card Number:
  - https://en.wikipedia.org/wiki/Resident_Identity_Card

Leakage Types and Data Validity

Four+ Types of Leakage

- ➤ Reveal identity
- ➤ Reveal value hidden attribute
- ➤ Reveal link between two entities
- ➤ Reveal group membership

Validity

- Bad Data and Bad Models lead to bad decisions

- If decision making is opaque, results can be bad in the aggregate, and can affect an individual

- For example: what if someone has a loan denied because of an error in the data analyzed?

The End