Data Ethics and Professionalism

ISEA Session 15

Patrick C. Kennedy
University of Oregon
May 9, 2025

Overview

1. Data ethics
2. Reproducibility
3. Communicating results

Data ethics

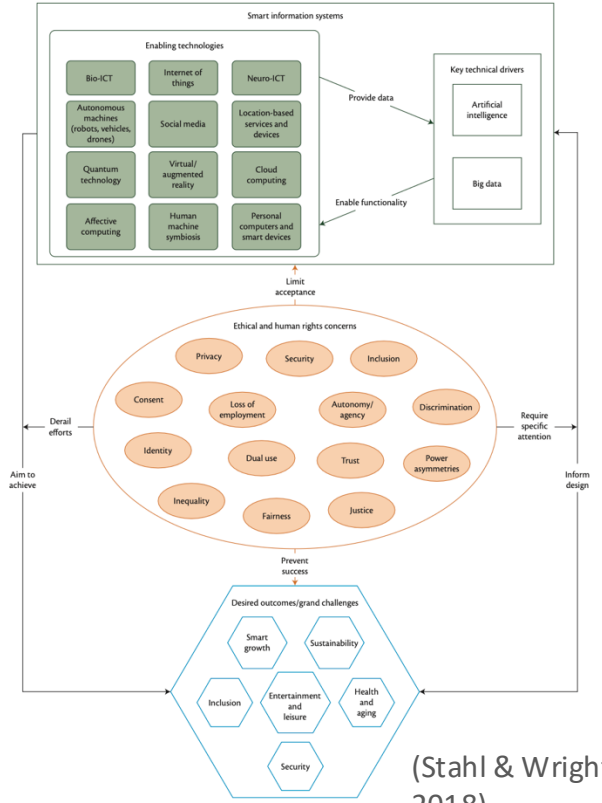
IES Institute of Education Sciences
AmplifyLearn.AI
UNIVERSITY of WASHINGTON eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS
UNIVERSITY OF OREGON

Motivation

- **Potential and peril**
  - **Massively accelerated rate at which data are produced**
  - **How data are analyzed & understood changing rapidly**
- **Artificial intelligence (AI)**
  - **The use of technology to mimic human cognitive capabilities** (Stahl & Wright, 2018)

Smart information systems

- Technologies that use AI, machine learning, & big data (Stahl & Wright, 2018)
- Main takeaway:
  - Wide range of tech
  - Driven by AI & big data
  - Influenced by many concerns
  - That affect desired outcomes



(Stahl & Wright, 2018)

Smart information systems (SIS)

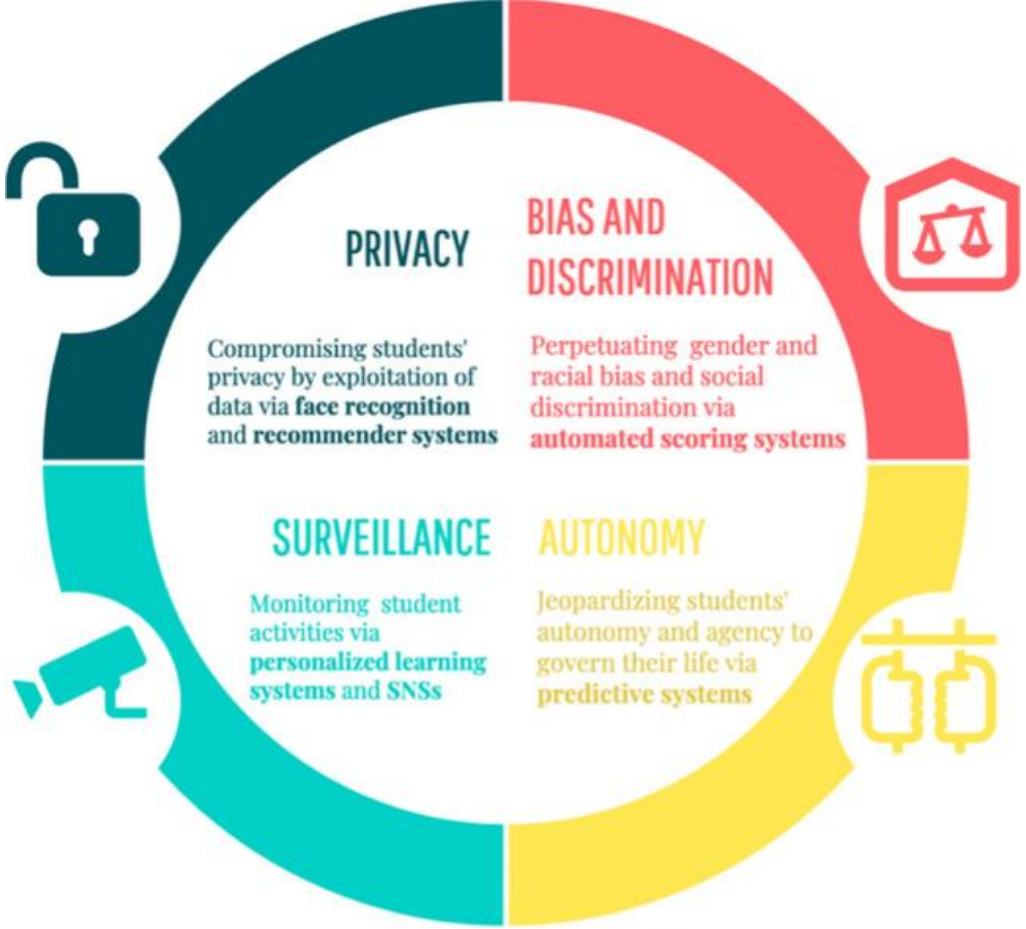
- Modern SIS are ubiquitous
  - Amazon recommendations/Alexa
  - Google search/Google Translate
  - Facebook/Instagram
- Other (education) examples?

Edtech SIS

- Personalized learning systems (PLS) based on students' needs & skills (Regan & Jesse, 2019)
- Facilitates collection of more, and more granular, info about students' educational experience
- Ubiquitous: used in nearly all US schools

Just privacy?

- Privacy as multi-faceted concept incorporating multiple distinct ethical concerns (Solove, 2008)
- Ethics discussions in US focus on privacy (Regan, 1995) & avoiding *inappropriate* access to data
- Reality is more nuanced



(Akgun & Greenhow, 2021)

Information privacy

- Limit info collected to that required for purpose
- Addressed by fair information practice principles: notice, consent, choice, and transparency
- Focusing only on information privacy is ineffective in a big data context

Surveillance

- Monitoring & analyzing student activity
  - Time on page
  - Keystrokes or response patterns
  - Location, time of day, other students doing same task
  - Cross-matched with other data (e.g., how much a student moves, or time spent on social networking)
  - Predictive analytics to determine patterns, strengths and weaknesses, and advice about how to personalize

Autonomy

- Algorithms jeopardize autonomy by steering people to make specific decisions
- PLS may influence instruction without giving students a choice

Discrimination & Bias

- SIS can
  - Perpetuate prejudices and accentuate social inequities
  - Create new forms of inequality
  - Introduce potential for bias from:
    - Those who design the systems
    - The algorithms themselves
- Algorithmic complexity can make identification of bias and discrimination difficult

“Average person” in US states

Anonymity

- The ability to remain anonymous if individuals so choose
- Nearly impossible with modern data sets
  - A handful of characteristics are likely sufficient to identify individuals

Ownership

- To what extent does data generated about and by students as they use PLS belong to the school or district vs. companies?
- Should there be limits on how companies use this data (e.g., to improve their offerings)?

What can be done?

- Edtech ethics is more than just privacy
- Responsible Research and Innovation (RRI)
  - Efforts to ensure that SIS processes & outcomes are acceptable, desirable, and sustainable
  - Built on principles of technology ethics, technology assessment, science and technology studies, and philosophy of technology

RRI

- Consensus is elusive
- Common foci include:
  - Stakeholder engagement
  - Openness and transparency
  - Willingness to be flexible and responsive
  - Integrating RRI into projects, funding, & support environments

Discussion

- Join the breakout room most relevant to you
  - 1) information privacy; 2) anonymity; 3) surveillance
  - 4) autonomy; 5) discrimination/bias; 6) ownership
- Discuss implications of that topic in your work
  - What can we, as educators with data science training, do to bring awareness to or help mitigate the issues?
  - How can we help develop a culture of responsibility among stakeholders for the processes and outcomes they develop and implement in schools?

BREAK

> 5-minute break

Reproducibility

Reproducibility

- **What** does it mean to be reproducible?
- **Why** does reproducibility matter?
- **How** is reproducibility achieved?

Reproducibility: What

- Conceptual
  - Replicating a study with new, independent data
    - Expensive
    - Hard(er) to get funded and published
    - Methodological challenges
- Computational
  - When others reproduce study results given only the original data, code, & documentation
    - Retains many advantages while minimizing the largest barrier (i.e., costs)

Reproducibility: Why

- **Benefits those who do it**
  - **Encourages robust documentation**
  - **Makes revisions easier**
  - **Promotes modularity and reuse of code**
  - **Provides an indication of rigor, trustworthiness, & transparency**
  - **Increases citation rates**

Reproducibility: Why

- Benefits the larger community
  - Makes findings more accessible
  - Allows others to learn from your work
  - Facilitates follow-up studies
  - Leads to faster progress
  - Provides protection when mistakes occur

Reproducibility: How

- **Plan**
  - Develop a well-defined question
  - Write and register study protocol
  - Justify the proposed sample size
  - Construct a data management plan
  - Proactively address sources of bias
- **Execute**
  - Avoid questionable practices
  - Interpret significance carefully
  - Make research open
- **Report**
  - Report all findings
  - Follow relevant reporting guidelines

Reproducibility: Analysis

- Before: Plan data storage & organization
  - Location & format, data structure, metadata
- During: Use coding best practices
  - Clean, well commented code; code review; document environment & parameters
- After: Finalize & share results
  - Include input data, scripts, program versions, parameters, and important intermediate results
  - Choice of repository, prioritize DOI for citations

Reproducibility: Resources

- UW eScience Reproducible and Open Research
  - http://uwescience.github.io/reproducible/
- Coursera Reproducible Research course
  - https://www.coursera.org/learn/reproducible-research
- Reproducibility and Replicability in Science
  - https://www.ncbi.nlm.nih.gov/books/NBK547537/

Communicating results

IES Institute of Education Sciences
AmplifyLearn.AI
UNIVERSITY of WASHINGTON eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS
UNIVERSITY OF OREGON

Clear communication

- > Motivate the contribution
- > Contextualize the work (what's known)
- > Balance interpretability and accuracy

A Checklist for Communicating Science and Health Research to the Public: https://www.nih.gov/about-nih/what-we-do/science-health-public-trust/checklist-communicating-science-health-research-public

Clear communication

> **Use visuals that can be easily understood**

A Checklist for Communicating Science and Health Research to the Public: https://www.nih.gov/about-nih/what-we-do/science-health-public-trust/checklist-communicating-science-health-research-public

IES Institute of Education Sciences

AmplifyLearn.AI

UNIVERSITY of WASHINGTON eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

UNIVERSITY OF OREGON

Clear communication

- > Be respectful
- > Avoid potentially offensive terms
- > Include other relevant resources
- > Provide citations and sources

A Checklist for Communicating Science and Health Research to the Public: https://www.nih.gov/about-nih/what-we-do/science-health-public-trust/checklist-communicating-science-health-research-public

IES Institute of Education Sciences

AmplifyLearn.AI

UNIVERSITY of WASHINGTON eScience Institute
ADVANCING DATA-INTENSIVE DISCOVERY IN ALL FIELDS

UNIVERSITY OF OREGON

Innovative communication

- Get the basics right
  - Define objectives, specify your audience, frame your message, and develop a dissemination plan
- Use websites, social media, and unique identifiers to make your work visible
- Foster participation and collaboration

(Ross-Hellauer et al., 2020)

Innovative communication +

- Embrace open science
  - Principles of equitable participation and transparency that enable others to collaborate in, contribute to, scrutinize and reuse research, and spread knowledge as widely as possible
- Think beyond traditional research outputs

(Ross-Hellauer et al., 2020)

Innovative communication ++

- > Engage stakeholders
- > Think data visualization
- > Reflect and respect diversity
- > Find and use the right tools
- > Evaluate, evaluate, evaluate

(Ross-Hellauer et al., 2020)

Assignment

- **Write a brief reflection (300-500 words) on the current or anticipated role of ethics & reproducibility in your own work.**
  - **What areas are most pertinent to you? How might principles of reproducibility help address (some of) these issues?**

References

Akgun, S., & Greenhow, C. (2022). Artificial intelligence in education: Addressing ethical challenges in K-12 settings. *AI and Ethics*, 2(3), 431–440. https://doi.org/10.1007/s43681-021-00096-7

Alston, J. M., & Rick, J. A. (2020). A beginner's guide to conducting reproducible research. *Bulletin of the Ecological Society of America,* 102(2), e01801. https://doi.org/10.1002/bes2.1801

Regan, P. M. (1995). Legislating privacy: Technology, social values, and public policy. Chapel Hill: University of North Carolina Press.

Regan, P.M., & Jesse, J. (2019). Ethical challenges of edtech, big data and personalized learning: twenty-first century student sorting and tracking. *Ethics and Information Technology*, 21, 167–179. https://doi.org/10.1007/s10676-018-9492-2

Ross-Hellauer, T., Tennant, J.P., Banelyte˙, V., Gorogh, E., Luzi, D., Kraker. P., et al. (2020) Ten simple rules for innovative dissemination of research. PLoS Comput Biol 16(4): e1007704. https://doi.org/10.1371/journal.pcbi.1007704

Schwab, S., Janiaud, P., Dayan, M., Amrhein, V., Panczak, R., Palagi, P. M., Hemkens, L. G., Ramon, M., Rothen, N., Senn, S., Furrer, E., & Held, L. (2022). Ten simple rules for good research practice. *PLoS Computational Biology*, 18(6), e1010139

Solove, D. (2008). Understanding privacy. Cambridge: Harvard University Press.

Stahl, B.C. & Wright, D. (2018). Ethics and privacy in AI and big data: Implementing Responsible Research and Innovation. *IEEE Security & Privacy*, 16 (3), 26-33. https://doi.org/10.1109/MSP.2018.2701164