Ethics in AI

Melissa Goldstein

mgoldste@gwu.edu

The Issues

- Scandals (Cambridge Analytica)
- Google employee protests over military contracts
- Biased algorithms in Amazon's hiring processes
- Artificial intelligence has a trust problem. NYT 9/6/19

Discussion

- Data Analytics
- Ethical Principles
- Privacy
- AI Principles/Guidelines

Data Analytics

Data Analytics

- Both opportunities and risks
  - Privacy
  - Civil rights
  - Responsible and ethical use of data
- Complex, difficult to understand.
  - Can make it hard to know when one's privacy is violated or if one has been the subject of unfair bias

https://obamawhitehouse.archives.gov/blog/2016/06/08/ensuring-data-driven-wellness-programs-work-us-and-not-against-us

Milken Institute School
of Public Health
THE GEORGE WASHINGTON UNIVERSITY

THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

Data Analytics

- Intentional or implicit biases may emerge from data (inputs to the system) or algorithms themselves
- Decisions based on these systems can affect our lives:
  - Whether we qualify for credit or employment opportunities
  - Which financial, employment and housing advertisements we see
- Transparency?

https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

Data Analytics: Challenge

- *"*Potential to increase access to opportunity and help overcome discrimination, if fairness in computational systems and ethical approaches to data analytics are the norm"

- *"*Risks that [these] innovations could perpetuate discrimination and unequal access to opportunity as the use of data expands"

https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

Equal Opportunity by Design

- Designing data systems that promote fairness and safeguard against discrimination from the first step of the engineering process and continuing throughout their lifespan.

- Bias mitigation

- Avoiding "data fundamentalism"

https://obamawhitehouse.archives.gov/sites/default/files/microsites/ostp/2016_0504_data_discrimination.pdf

Milken Institute School
of Public Health
THE GEORGE WASHINGTON UNIVERSITY

THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

Ethical Principles

Ethics

The study of morality using the tools and methods of philosophy (+ law + policy + religious tenets).

Ethics (types)

- Applied ethics—The use of moral norms and concepts to resolve practical moral issues.

- Bioethics—Applied ethics focused on health care, medical research, and medical technology.

- Organizational ethics—Focused on the character traits, rights, values, and duties expressed in the following formats (among others):

  - Mission Statements

  - Policies and administrative practices

  - Business priorities

(Purtilo, Ethical Dimensions in the Health Professions, 4th ed., 2005)

Moral Principles

- An absolute principle applies without exceptions.

- A *prima facie* principle applies in all cases unless an exception is warranted (*i.e.*, applies until and unless it conflicts with another *prima facie* principle).

Moral Principles

- Autonomy (Respect for Persons)
- Nonmaleficence
- Beneficence
- Justice

Autonomy

- Autonomy—A person's rational capacity for self-governance or self-determination

- Autonomy principle—Autonomous persons should be allowed to exercise their capacity for self-determination

Vaughn L. 2d ed. Bioethics (Oxford University Press: New York 2013).

Respect For Persons

- Act so as to treat humanity, in every case, as an end and never as a means only [Kant]

- Incorporates at least 2 ethical convictions

  - Individuals should be treated as autonomous agents.

  - Persons with diminished autonomy and thus in need of protection are entitled to such protection.

Beneficence

- *Primum non nocere*: First (or above all) do no harm
- Two general rules:

  1. Do no harm;

  2. Maximize possible benefits and minimize possible harms.

Justice

- Distributive justice—Justice regarding the fair distribution of society's advantages and disadvantages.

- Egalitarian theories of justice—Doctrines affirming that important benefits and burdens of society should be distributed equally.

Vaughn L. 2d ed. Bioethics (Oxford University Press: New York 2013).

Justice

- Libertarian theories of justice—Doctrines holding that the benefits and burdens of society should be distributed through the fair workings of a free market and the exercise of liberty rights of noninterference.

- Utilitarian theories of justice—Doctrines asserting that a just distribution of benefits and burdens is one that maximizes the net good (utility) for society.

Vaughn L. 2d ed. Bioethics (Oxford University Press: New York 2013).

Organizational Codes of Ethics

- Define appropriate and inappropriate conduct within the workplace and identify conduct that falls between those poles.

- Protect organizations from legal entanglements and employees from harassment and unfair treatment

- Leaders should support the code of ethics and should educate themselves and others about its application and benefits.

- Mechanisms to monitor employee adherence to the code must be developed.

(Dye, Leadership in Healthcare: Essential Values and Skills, 2d ed., 2010)

Privacy

Privacy

Issues to discuss:

- What is at stake?

- How do we define (health) privacy?

- What is different in a networked environment?

What Does Privacy Mean?

What people refer to as privacy can sometimes be more precisely described:

- Privacy: right to control own health information

- Confidentiality: privacy interest within specific legal relationship

- Security: safeguards placed on health information

Privacy

- Basic human right
- Negative and positive right
  - Negative: right to be left alone
  - Positive: right to control access to information about yourself
- Right protected by common law, US Constitution, state constitutions

Ethical Justifications

- Grounded in principle of respect for persons (autonomy) and rule of fidelity
  - Respect for persons (autonomy): consideration of an individual's capacity and desire for self-governance
  - Fidelity: promise keeping, loyalty, faithfulness in the health care provider-patient relationship

Health Information Privacy

- Unauthorized/erroneous disclosures
  - Embarrassment
  - Stigma
  - Discrimination in employment, insurance, government programs/benefits
- AND "the distinction between personal data and health care data is blurry"

  (Executive Office of the President, Big Data: Seizing Opportunities, Preserving Values (May 2014), 23-24.)

Privacy and Health Generally

- Individuals' privacy must be balanced with important uses of health information (e.g., research, public health)
- Infringing too much on privacy could reduce people's willingness to seek medical care, participate in research, and/or cooperate with public health authorities
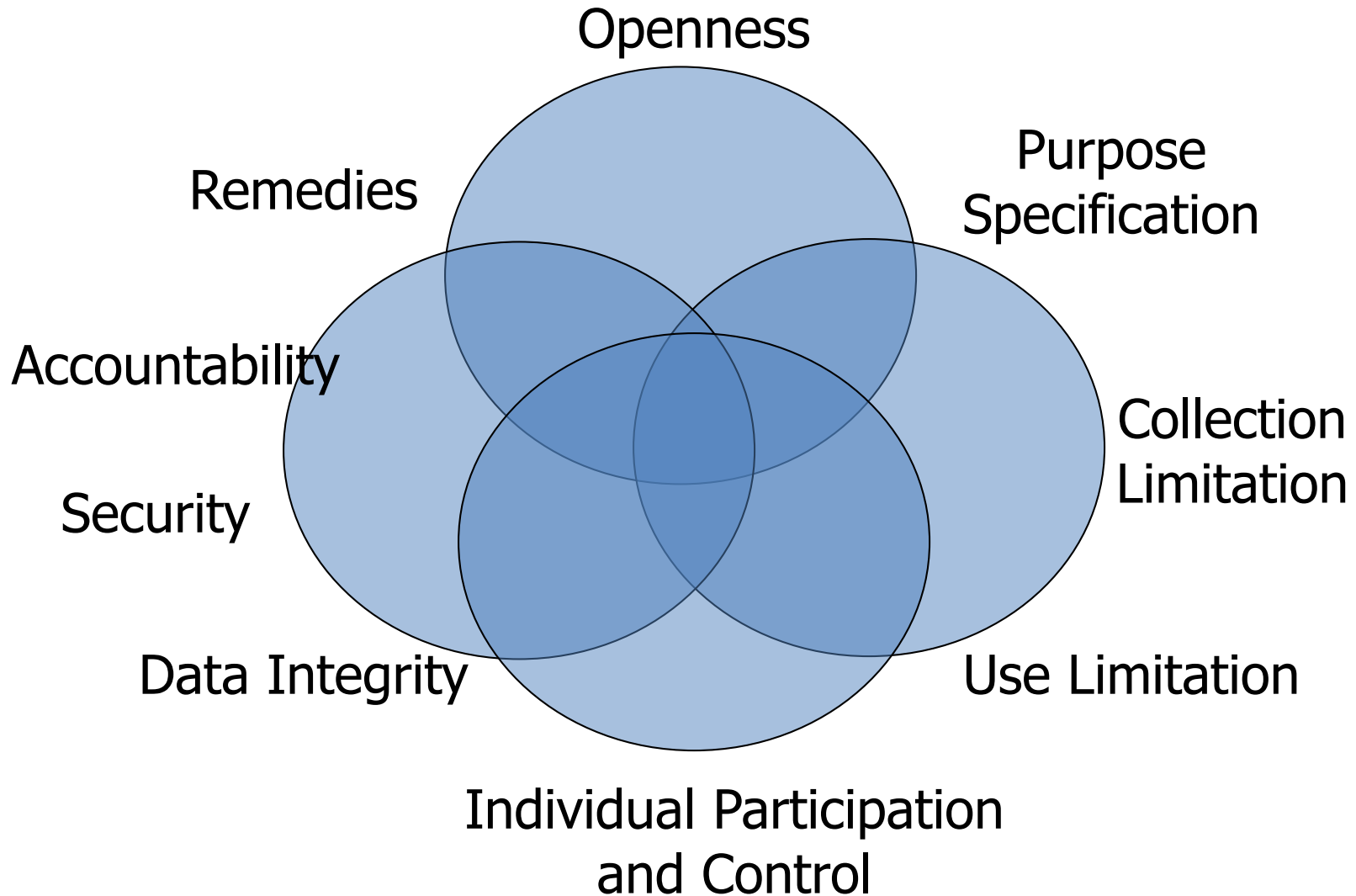
Privacy and Networked Environments

In a networked environment: "data shadow"

- Increased risks (mis-use and re-use)
- Demands systemized privacy protection (based upon fair information practice principles) to reduce risks and establish trust

https://www.markle.org/markle-common-framework-connecting-professionals

Privacy in a Networked Health Environment

Openness

Purpose Specification

Remedies

Accountability

Collection Limitation

Security

Data Integrity

Use Limitation

Individual Participation and Control

Health Information Privacy: Legal Status (LOTS OF LAW)

- U.S. Constitution – qualified right

- HIPAA

- Privacy Act of 1974

- Federal Regulations for the Protection of Human Research Subjects ("the Common Rule")

- State laws

- Special status for certain diseases (drug and alcohol treatment (42 CFR Part 2), HIV/AIDS, genetic conditions (GINA), etc.)

Health Information Privacy: Legal Status (continued)

- mHealth (mostly outside of HIPAA)
  - FDA (mobile medical apps)
  - Food and Drug Administration Safety and Innovation Act, 2012 (FDASIA) (required FDA (with ONC and the FCC) to develop recommendations on a risk-based regulatory framework)
  - FTC (authority to prevent "unfair or deceptive" acts related to commercial activity)
  - Others (NTIA, FCRA, Gramm-Leach-Bliley Act, etc., 21 Century Cures Act etc.)
- Medicare claims data (SSA, ACA)

Ongoing Issues

- BIG data
- Research/Public Health (who holds/controls the data?)
- Electronic Health Records (EHR)
- Personal Health Records (PHR)
- mHealth (mobile health apps, health-related websites and web-based services, portable sensors that connect to mobile devices, etc.)
- "Anonymizing" Data – genomics, pharmacogenomics
- AI…

AI Principles/Guidelines

Inventories

- Recent inventories of AI ethics guidelines have collected over 80 different ethical frameworks

- Private companies, research institutions, and public sector organizations

- Vary in content and detail

- Some common elements

Health Ethics and Policy Lab: "'AI' Should be Ethical"

- But debate about what constitutes 'ethical AI'

- And debate about which ethical requirements, technical standards, and best practices are needed for its realization

Milken Institute School
of Public Health
THE GEORGE WASHINGTON UNIVERSITY

THE GEORGE
WASHINGTON
UNIVERSITY
WASHINGTON, DC

Global convergence around 5 ethical principles

- Transparency, justice and fairness, non-maleficence, responsibility. and privacy
- Differences in
  - how the principles are interpreted
  - why they are deemed important
  - what issue, domain, or actors they pertain to, and
  - how they should be implemented

Conclusion

Findings highlight the importance of integrating guideline-development efforts with substantive ethical analysis and adequate implementation strategies.

AlgorithmWatch [AI Ethics Guidelines Global Inventory](#)

- All frameworks include similar principles on transparency, equality/non-discrimination, accountability, and safety

- Some add additional principles, such as the demand for AI be socially beneficial and protect human rights

AlgorithmWatch, etc.

- Most frameworks developed by coalitions, or institutions such as universities that invite companies and individuals to sign on

- Only a few companies have developed their own frameworks

- Almost all are voluntary commitments. Only three or four indicate an oversight or enforcement mechanism

- Most from 2018/2019

- Few include recommendations or examples of how to operationalize the principles

Conclusion

"The absence of internal enforcement or governance mechanism shows that, there is a lot of virtue signaling going on and few efforts of enforcement."

Questions/Comments?