Computational Ethics

Human subjects research

People in AI technologies

- People create data
- People develop & deploy AI technologies
- People use AI technologies

People in AI technologies

- People create data
- People develop & deploy AI technologies
- People use AI technologies

⟶ Data and labels are noisy

People in AI technologies

- People create data
- People develop & deploy AI technologies
- People use AI technologies

⟶ Data and labels are noisy

⟶ How to use humans to get more/better labels? Let's use Amazon Mechanical Turk and get an answer?

History of using Human Subjects

- World War II medical experiments on prisoners in concentration camps and Nuremberg Code of 1947
- Tuskegee syphilis experiment
- Stanford prison experiment
- Milgram experiment
- National Research Act of 1974

Nuremberg Code of 1947

- World War II medical experiments on prisoners in Nazi concentration camps
- After the war ended, a series of trials were held against major war criminals
- The first trial in 1947 – The Doctors' Trial
  - "The United States of America v. Karl Brandt, et al.,"
  - 23 physicians from the German Nazi Party were tried for crimes against humanity for murder and torture in the atrocious experiments they carried out on unwilling prisoners of war
  - 16 were found guilty, of which 7 received death sentences and 9 received prison sentences ranging from 10 years to life imprisonment
- The verdict also resulted in the creation of the Nuremberg Code
  - a set of 10 ethical principles for human experimentation

https://history.nih.gov/display/history/Nuremberg+Code

Nuremberg Code of 1947

1. Voluntary **consent** is essential
2. The results of any experiment must be for the **greater good of society**
3. Human experiments should be based on previous animal experimentation
4. Experiments should be conducted by **avoiding physical/mental suffering** and injury
5. No experiments should be conducted if it is believed to cause death/disability
6. The **risks should never exceed the benefits**
7. Adequate facilities should be used to protect subjects
8. Experiments should be conducted only by qualified scientists
9. Subjects should be able to **end their participation at any time**
10. The scientist in charge must be prepared to terminate the experiment when injury, disability, or death is likely to occur

Shuster, Evelyne. 1997. "Fifty years later: the significance of the Nuremberg Code." New England Journal of Medicine 337, 20: 1436-1440.

United States Public Health Services study in Tuskegee

- 40-year study by the US Public Health Service begun in 1932
- Goal: observe natural history of untreated syphilis
- Enrolled 600 poor African American sharecropper men
  - 399 with syphilis, 201 controls
- Told they would be treated for "bad blood"
- Were not treated, merely studied
  - Were not told they had syphilis
  - Sexual partners not informed
  - By 1940s penicillin becomes standard treatment for syphilis
    - Subjects were not told or given penicillin

Doctor taking blood from a Tuskegee Subject
(National Archives via Wikipedia)

United States Public Health Services study in Tuskegee

- 1964 Protest letter from a doctor who reads one of the papers
  - “I am utterly astounded by the fact that physicians allow patients with a potentially fatal disease to remain untreated when effective therapy is available,”
  - “I assume you feel that the information which is extracted from observation of this untreated group is worth their sacrifice. If this is the case, then I suggest the United States Public Health Service and those physicians associated with it in this study need to re-evaluate their moral judgments in this regard.”

- 1965 Memo from authors:
  - “This is the first letter of this type we have received. I do not plan to answer this letter”

United States Public Health Services study in Tuskegee

- 1966 Peter Buxtun, a PHS researcher in San Francisco, sent a letter to the CDC but study was not stopped.
- 1972 Buxtun goes to the press.
- Senator Edward Kennedy calls congressional hearings
- 1974 Congress passes National Research Act

Syphilis Victims in U.S. Study Went Untreated for 40 Years

By JEAN HELLER
The Associated Press

WASHINGTON, July 25—For 40 years the United States Public Health Service has conducted a study in which human beings with syphilis, who were induced to serve as guinea have serious doubts about the morality of the study, also say that it is too late to treat the syphilis in any surviving participants.

Doctors in the service say

NY Times July 26, 1972

Behavioral experiments

Stanford prison experiment

- Conducted by Philip Zimbardo, Stanford University, August 1971
- Goal: test how perceived power affects subjects
- College students were chosen to be either "prisoners" or "guards"
- “Guards” selected uniforms, and defined discipline
- Results as published by Zimbardo:
  - Guards humiliated and abused prisoners
  - Prisoners became depersonalized
  - Evidence for "ugly side of human nature"
- Although scheduled for 2 weeks, experiments had to stop after 6 days

Stanford prison experiment scientific and ethical flaws

- Participants were not random: respondents to an ad for “a psychological study of prison life.”
  - Carnahan and MacFarland 2007: word "prison" selects personalities
- Guards were told the expected results ("conditions which lead to mob behavior, violence")
- Researchers intervened in experiment to instruct guards how to behave ("We can create a sense of frustration. We can create fear")
- Guards not told they were participants
- Researcher refused to allow prisoner participants to leave experiment.

Le Texier, T. (2019). Debunking the Stanford Prison Experiment. *American Psychologist, 74*(7), 823–839. https://doi.org/10.1037/amp0000401

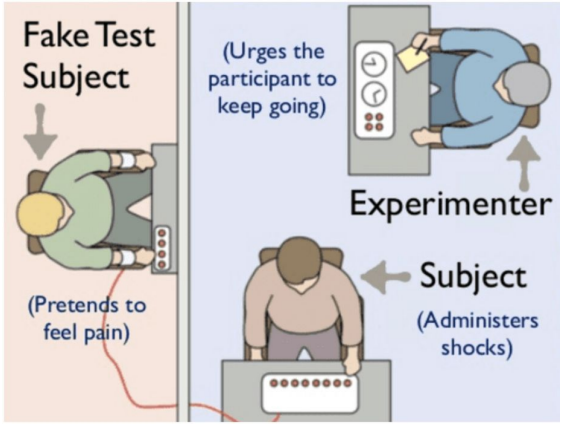
Blue vs brown eye “racism”

- Kids separated by color of eyes
  - Blue eyes are better
  - Brown eyes are worse
- Quickly separate in clans
- Blue given advantages, Brown given disadvantages
- Kids quickly accept the divisions

- Is this experiment ethical?
- Do we learn something?
- Do the participants learn something?

Milgram obedience experiment

- Stanley Milgram, Yale, 1962
- Three roles in each experiment
  - Experimenter
  - Teacher (actual subject)
  - Learner
- Learner and Experimenter were informed about the experiment
  - Teacher asked to give mild electric shocks to the Learner
  - Learner had to answer questions and got things wrong
  - Experimenter, as the matter of fact, asked Teacher to torture Learner
- Most Teachers obeyed the Experimenter



(Source: moderntherapy.online)

Current Human Participants Rules

National Research Act 1974

- These experiments (especially the Tuskegee experiment) led to the creation of the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research
  - The Common Rule: Title 45, Part 46 of the Code of Federal Regulations: Protection of Human Subjects.
    - Informed consent
  - Required institutional review of all federally funded experiments
    - Institutional Review Boards (IRBs)
  - Issued Belmont Report in 1976/1979

The Belmont Report: three basic ethical principles

1. Respect for Persons

- Individuals should be treated as autonomous agents
  - "Informed Consent"
- Persons with diminished autonomy are entitled to protection

The Belmont Report: three basic ethical principles

2. Beneficence (minimizing harm)

- Do no harm
- Maximize possible benefits and minimize possible harms.

The Belmont Report: three basic ethical principles

3. Justice

Who ought to receive the benefits of research and bear its burdens?

- Fair procedures and outcomes in the selection of research subjects
- Advances should benefit all

IRB – Institutional Review Board

- Institutional Review Board
  - Internal to institution
    - Most universities have at least 2 distinct boards: medical and non-medical
  - Independent of researcher

IRB – Institutional Review Board

- Reviews all human experimentation
  - Assesses instructions
  - Compensation
  - Contribution of research
  - Value to the participant
  - Protection of privacy and confidentiality

IRB – Institutional Review Board

- Different standards for different institutions
  - Medical School vs Engineering School
- Board consists of (primarily) non-expert peers
  - Also helps educate new researches and makes suggestions to find solutions to ethical issues

Human subject

- Human subject: a living individual **about whom** an investigator (professional or student) conducting research:
  - Obtains information through **intervention** or **interaction** with the individual, and uses, studies, or analyzes the information; or
  - Obtains, uses, studies, analyzes, or generates **identifiable private information**

Ethical questions

- Can you lie to a human subject?
- Can you mislead a human subject?

Deceiving participants

Belmont Report:

- "incomplete disclosure" is allowed when:
  - incomplete disclosure is truly necessary to accomplish the goals of the research
  - there are no undisclosed risks to subjects that are more than minimal, and
  - there is an adequate plan for **debriefing** subjects, when appropriate, and for dissemination of research results to them

Ethical questions

- Can you lie to a human subject?
- Can you mislead a human subject?
  - deception, incomplete disclosure, no more than minimal risk, no alternative
  - key concept: **debriefing**

Ethical questions

- Can you lie to a human subject?
- Can you mislead a human subject?
- Can you harm a human subject?

Exempt research

- Your human subjects research qualifies for **exempt status** if
  - Research only includes survey procedures, interview procedures, or observation of public behavior (including visual or auditory recording) and one of:
    - The identity of the human subjects cannot readily be ascertained
    - Any disclosure of the human subjects' responses outside the research would not place the subjects at risk of criminal or civil liability or be damaging to the subjects' financial standing, employability, educational advancement, or reputation; or
    - ...
- Secondary research uses of identifiable private information if the identifiable private information is publicly available
- ...

Ethics of using social media data

Using social media data

- Social media data: posts from Twitter, Reddit, YouTube, etc.
- From IRB perspective this kind of corpus data is exempt if it is public
  - E.g., public twitter data
- But are there still questions?

Possible issues with social media data

- Informed consent
- Privacy
  - Individuals can be searchable by their comments
  - Online disinhibition on the behaviors of users
- Terms of service
- Representativeness of data

Issues with social media data: author

- "Are consent, confidentiality and anonymity required where the research is conducted in a public place where people would reasonably expect to be observed by strangers?"
- What counts as a public vs. private space on/off the web?
  - If people are whispering in a public square is that private?
  - What about religious ceremonies?

Williams, M. L., Burnap, P. 2017. Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation. Sociology, 51(6), 1149–1168.

Issues with social media data: author

- What are the potential harms?
  - Demographic info (age, ethnicity, religion, sexual orientation)
- Associations (membership in groups or associations with particular people)
- Communications that are personal or potentially harmful (extreme options? Illegal activities?)
- Others?

Williams, M. L., Burnap, P. 2017. Towards an Ethical Framework for Publishing Twitter Data in Social Research: Taking into Account Users' Views, Online Context and Algorithmic Estimation. Sociology, 51(6), 1149–1168.

What do Twitter authors think?

Table 2. Comfort Around Tweets Being Used in Research.

Question	Very uncomfortable	Somewhat uncomfortable	Neither uncomfortable nor comfortable	Somewhat comfortable	Very comfortable
How do you feel about the idea of tweets being used in research? ($n = 268$)	3.0%	17.5%	29.1%	35.1%	15.3%
How would you feel if a tweet of yours was used in one of these research studies? ($n = 267$)	4.5%	22.5%	23.6%	33.3%	16.1%
How would you feel if your entire Twitter history was used in one of these research studies? ($n = 268$)	21.3%	27.2%	18.3%	21.6%	11.6%

Casey Fiesler and Nicholas Proferes. 2018. "Participant" Perceptions of Twitter Research Ethics. Social Media + Society, 4(1). 22

Table 4. "How Would You Feel If a Tweet of Yours Was Used in a Research Study and . . ." ($n = 268$).

	Very uncomfortable	Somewhat uncomfortable	Neither uncomfortable nor comfortable	Somewhat comfortable	Very comfortable
. . . you were not informed at all?	35.1%	31.7%	16.4%	13.4%	3.4%
. . . you were informed about the use after the fact?	21.3%	29.1%	20.5%	22.0%	7.1%
. . . it was analyzed along with millions of other tweets?	2.6%	18.7%	25.5%	30.0%	23.2%
. . . it was analyzed along with only a few dozen tweets?	16.5%	30.3%	24.0%	20.2%	9.0%
. . . it was from your "protected" account?	54.9%	20.5%	13.8%	6.0%	4.9%
. . . it was a public tweet you had later deleted?	31.3%	32.5%	20.5%	10.4%	5.2%
. . . no human researchers read it, but it was analyzed by a computer program?	2.6%	14.3%	30.5%	32.3%	20.3%
. . . the human researchers read your tweet to analyze it?	9.7%	27.6%	25.0%	25.4%	12.3%
. . . the researchers also analyzed your public profile information, such as location and username?	32.2%	23.2%	21.0%	13.9%	9.7%
. . . the researchers did not have any of your additional profile information?	4.9%	15.4%	25.1%	34.1%	20.6%
. . . your tweet was quoted in a published research paper, attributed to your Twitter handle?	34.3%	21.6%	21.6%	13.1%	9.3%
. . . your tweet was quoted in a published research paper, attributed anonymously?	9.0%	16.8%	26.5%	28.4%	19.4%

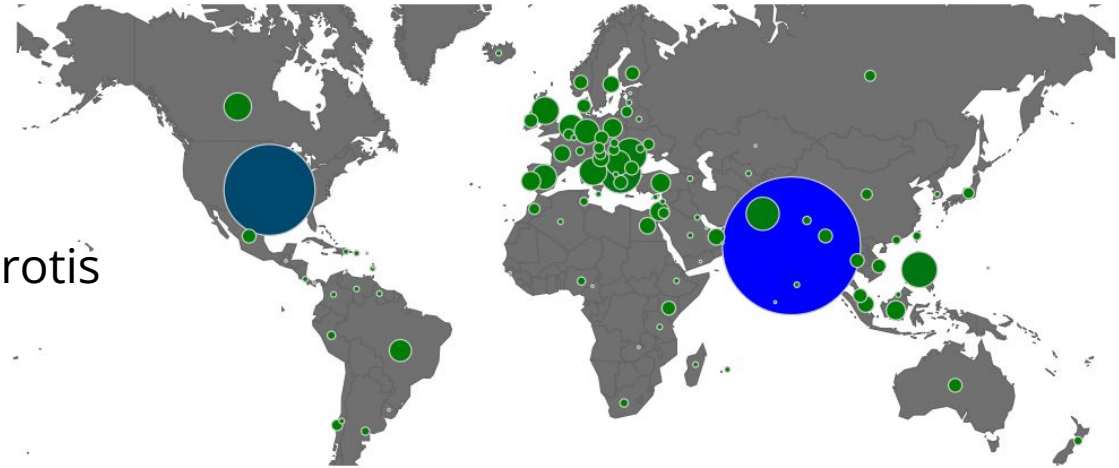
Human subjects selection: crowdwork and crowdworkers

A few questions on crowdsourcing

- Have you ever employed crowdworkers?
  - How much do you pay them?
  - Are you sure? How do you know?
- Have you ever done a crowdsourced task yourself?
- Would you ever do crowdsourced tasks for a prolonged period?
- Would you recommend being a crowd worker to a friend or family member?

Who are crowdworkers?

- Largely from the US and India



- Skewed young, female, and lower income; many underemployed (30%+) (Ipeirotis 2010)

- For multilingual work: Language Demographics (see Pavlick et al. 2014 for details)

Crowdworker incomes

- Multiple studies have found actual wages at or below $2 / hr

  > Our task-level analysis revealed that workers earned a median hourly wage of only ~$2/h, and only 4% earned more than $7.25/h. While the average requester pays more than $11/h, lower-paying requesters post much more work. (Hara et al. 2018)

- Need to consider time finding tasks, lost wages from rejections, etc

- Frequently secondary source of income, sometimes primary
  - 20%+ use crowdwork income to “to make basic ends meet”

Fort, K., Adda, G., & Cohen, K. B. (2011). Amazon Mechanical Turk: Gold mine or coal mine?. Computational Linguistics, 37(2), 413-420.

Turkopticon, TurkerNation, et al.

HINL A372VF8C4PEU98		$3.00	**Respond to stories about others (30 min, computer mouse recommended)** Returned: I'm sorry, but it looks like you are ineligible. You participated in one of our other related studies. Please do not accept this HIT. Pineapples | 2018-05-30 23:21:42 UTC | flag | comment
Ovul Sezer A02083613H11QIJB1H2CW		$0.75	**Survey about social interactions. $0.75(~ 10 minutes)** not recommended unannounced screener Wasted a minute answering questions just to get screened out. no mention of a screener. andr...@u... | 2018-05-30 22:30:41 UTC | flag | comment
Metin Bektas A2U4A2OJ9OQYBG	6m	$0.40 $4.00/hr	**A 6-Minute Childhood, Personality and Time Perception Survey** (5/30/18) Bubbles, and a countdown timer at the bottom of each page; I have no idea why. kimadagem | 2018-05-30 22:01:27 UTC | flag | comment

Crowdworker invisibility

- **Invisibility** facilitated by “distance, anonymity, minimal communication, and electronic exchange”

(Martin et al. 2014)

- Turkers appreciate communication and responsiveness, i.e. less invisibility

The ethics of crowdsourcing

- Many crowdwork platforms' position: we are a neutral facilitator
- One crowdworker's response: "That's like saying someone is running a slave market on my property, and they're paying me, but I have no responsibility."
  - "How Crowdworkers Became Ghosts in the Digital Machine" - The Nation, 2014

- Crowdsourcing is "designed to form a participation→ reward→ participation→ reward cycle that purposefully continues to engage the participant in the research project for an unspecified, and unbracketed, length of time"

(Graber and Graber 2012, Journal of Medical Ethics)

- Great resource: Chris Callison-Burch's class http://crowdsourcing-class.org/

Are crowdworkers…

Employees?

Okay, but they have:

- No job security
- No benefits
- No collective bargaining
- No ability to know their employers
- No recourse in bad situations

or

Human Subjects?

Okay, so then we need to:

- Ensure they benefit
- Allow them to end participation at any time
- Eliminate coercion
- Debrief them

Ethical considerations

- Obviously, fair pay
  (researchers in general are already better at this)
- Give benefit of the doubt
- Consider if the scale is actually necessary
- Semi-automated solutions (and pay more for the annotations you do get)
- Responsiveness and communication - you are dealing with real people
- WeAreDynamo Guidelines for Academic Requesters

Check out:
http://faircrowd.work/

Another perspective: moral philosophy

"Act in such a way that you treat humanity,
whether in your own person or in the person of any other, never merely as a means to an end,
but always at the same time as an end."

- Immanuel Kant, Grounding for the Metaphysics of Morals

Another consideration: data quality

- Paying people to do use your system
  - Not the same as them actually using it
- Spoken Dialog Systems (Ai et al. 2007)
  - Paid, happy to go to wrong place (DARPA Communicator 2000)
  - User: “A flight to San Jose please”
  - System: “Okay, I have a flight to San Diego”
  - User: “Okay”
  - :-(
- All human experimentation includes bias