20211129 Data Privacy

Monday, November 29, 2021 1:59 PM

- Disclaimer:
  - Opinion
  - Philosophy
- Why Prof cares
  - Freedom, equality, accountability
- User - Product Relationships today
  - Centered around trust
    - Only works because you trust the platform
      - When you go to gas station, you trust the dial is accurate, but you could verify it if you really wanted
    - Online a lot is not verifiable
      - You don't know what the code is doing
      - Even if open source, could be extremely complicated
        - Even people who made it may not understand it
    - You may not see the full picture
      - You may not realize you have been "burned by the fire"
    - Your window is at the whim of the user interface
    - To make reasonable decisions, you have to trust the platform is doing what you think its doing
    - (follow link in sides)

        - New York times article about buttons in elevators/crosswalks, etc that don't do anything
    - Black box platforms
- Data privacy
    - Platform uses data to customize the interface and adapt to you
        - Nothing in the physical world is like that
    - Online world kind of customizes itself to you
    - Personalized recommendations
        - Is this really best?
        - Google search customize
            - Polarize groups and opinions
        - What optimizes what the platform desires?
            - May not be in your interest
                - There are a couple of links in the slides
                    - From Stanford: biasing the search for information based on your history
                    - Polarizing societies by only exposing you to what you like or react to in certain ways
        - Customizing you to the platform
        - The mechanism driving the recommendations of what data is being used
            - Totally undisclosed
    - Two sides:

- Information may be accurate and providing a better experience
- Could be totally wrong or bad for you
  - Is it targeted towards you so that you will buy the thing
- Distinction is difficult to make
- Examples:
  - New York times article "How Companies Learn Your Secrets"
    - Targeted ads at Target customers
    - Algorithm was so powerful that is was freaking people out, so they had to tone it down and send some random ads so people would actually buy things
    - Woman sent a pamphlet that implied they were pregnant
      - Before her family knew she was pregnant
    - There are a lot people out there, so there are a lot examples
- Giving the data to these platforms
  - If you disclosed to users what you do with data and what you know about the clients, then the customers could have some free will
    - Steals some agency
    - (His perspective)
- Very accurate and very scary

- Bias
    - It could be terribly wrong, and that is also bad
    - Patterns are learned on historical data
        - Historically we haven't treated each other well
            - Racist, sexist, etc
    - Just because algorithm is written on computer, does not mean it is objective
    - You want representation in your team of coders
    - The data itself could be biased
    - Very hard to remove bias from the algorithm
        - It's important to be aware
            - Don't just perpetuate negative societal impacts
    - Garbage in, garbage out
        - Consequences could be severe
            - If you have a huge negative impact on a group of people
    - Youtube video:
        - "Gender Shades" from MIT Media Lab
        - Facial detection
            - Frequently not detecting African American
                - The ones that did, misgendered
        - Was this unique to her face?
        - Dataset >1000 faces from Parliament members around the world that had a high female representation
            - For both African & European
        - Used another dataset for Asian
        - All companies performed better on males and

lighter subjects
    - Worst darker females
    - IBM had the largest gap in accuracy
- One issue is the training datasets
    - There may be other challenges
- Data centric technologies are prone to bias and can effect employment opportunities, loans, etc
- Any thoughts on the video? (from classmates)
    - (I didn't hear everything)
- Professors comments:
    - Facial recognition is being more and more widely used
    - Backlash:
        - Please don't include us
        - There is bias in these algorithms, but getting more images of genders & skin colors, is that really the best approach?
        - Do we need more and better facial recognition?
            - Will this have a positive societal impact?
                - Policing
                - Potential for furthering discrimination, over policing,
        - Every problem you solve with technology creates more problems
- Washington post:
- Another youtube video:

- HireVue Platform Overview
- Student:
  - Thought it was very impersonal
  - Can better prepare and have notes in front of you about interviews
- Prof:
  - What do you think may be some issues with a platform that ranks about interview videos?
    - Students:
      - One got in trouble because the algorithm got trained on who was wearing suits, but then women weren't getting hired
      - Not everyone has equal access to webcams, internet access
  - Could a system like this be racist? Yes

- Links in slides
- Paper
  - Criminal prediction based on how your face looks
  - Would such a correlation mean anything?
  - Asian male faces
  - Picture of a criminal once they are declared a criminal
    - They may not be happy at the time
  - Some of these are mug shots and some are ID photos
  - High accuracy, but it doesn't mean the face is

representative of whether you committed a crime
        - Who sponsored the research?
    - Old student had a project in their resume
        - Improving facial recognition for a Chinese police department
        - China already has a lot of facial recognition
        - Student: "I invited the knife", it's not my problem what they did with it
        - Prof: you are responsible for the tools and their misuse
            - You can get really deep in the philosophy
        - Could be used to misclassify people, but you could also find missing children
            - Really difficult space
        - Student:
            - Up to creator to look at what something is created to do
        - Prof: build tools for community
    - When Apple released software that scans through photos
        - Looking for child pornography
        - But now everyone is being tracked, and another source of revenue
        - Everyone values different things
            - Prof's opinion: don't know if society has benefited from surveillance
    - Algorithms that give you a score on whether you are likely to commit a crime or not and whether

you should be released from jail
    - Whenever you have a general score that *doesn't really* mean anything, you have to question
    - Maybe a qualitative based system may have been better
        - Behavior
        - A little more transparent
        - More for the person
- Freedom of choice
    - Freewill and your ability to make decisions
    - Who are determinists in this room? Who thinks you are just a product of your upbringing?
    - Who thinks we have free will?
    - How much are you in control of your decisions?
    - Examples?
        - Spontaneous decision to move across the country
        - Research for months what smart phone to buy
        - On the whim decisions
    - We aren't always aware of everything that is going into our decision making process
        - You may be able to list the bigger factors, but you may not be able to list everything
    - Scandal recently:
        - Cambridge analytical
            - Propaganda machine
            - Child company of SCL group
                - London based

        - Purpose is to understand mass behavior and influence it
        - Around 30 or 40 years
    - Example that you can influence people's decisions and perspectives
    - Video:
        - The Guardian: what is the Cambridge analytical scandal?
        - Pull's out entire friend's network data
        - Company only needs to touch 100K people to touch huge numbers of people
        - Spent a million dollars harvesting 10 million Facebook profiles
        - Identified targeted voter groups and targeted
        - Full service propaganda machine
    - Know about this because of the whistle blower in the video
        - Comes down to one person for us to know what is going on
- Wikipedia: Groundhog Day film
    - He hates this movie
    - Repeated same day over and over again
    - He can gain the system
    - Love story
    - But is really social manipulation

- Subject to Future Scrutiny
    - Data is there to stay for tools 10s or 100s of years from now

    - The data you gave 20 years ago, you may not have consented to how its used today
        - Facial recognition is much better
        - Hard to really consent in the first place
    - Data privacy is future proofing
    - Lives may be totally obsolete at some point because our entire lives are captured???? (didn't get this)
    - There have been attempts to regulate this
        - Challenging to do
        - We can barely understand what our own code does. Even two weeks after. (Especially if you didn't comment and you are out of practice)
        - Conflicts of interest with oversight?
        - You can delete your information, but the learned information and the changes in the model can't be taken back
- PII: Personally Identifiable Information
    - Most data privacy laws focused on this
    - Social Security number, phone number
    - But there are other things that uniquely characterize you
    - 2009: mayor released encrypted medical information
        - Using zip, birthday, and sex cross referenced with census information, and someone sent the mayor's medical records to her
    - Netflix
    - IMDB

- Data Brokers
    - Companies making money auctioning your data
    - Who doesn't want to understand their users better?
    - Do you want to make a profit of your data?
    - If you sell your data, how much is your identity worth?
        - Prof: you are your data (perspective)
- What can you do now?
    - Not a lot, as a user you don't have that much power
    - Ask if giving your data is required?
        - If someone asks you to do a survey (awkward if person is in position of authority, so just ask)
    - Reads terms and condition (which is a lot to do)
        - Legalese
            - Spelled out if they sell your data
    - Clear your cookies
    - Reflect on what activity patterns define you (PII)
    - Think about what data you leak across software - try to compartmentalize, make it difficult to join your data from different sources
        - Do everything for task, close browser, clear cookies, start next text
            - Personal tasks, work tasks
                - If all at same time, you are providing the aggregation for free
    - His research
- Incentivizing the use of privacy preserving tools

    - How do we define privacy?
    - Right now, we are asking companies to preserve privacy for the good of humanity or publicity, but we need better incentives
- Differential Privacy - How it works
    - Think about a smoker
        - Flip a coin, and if heads record correct answer
        - If tails, record a random result
        - Can get an aggregate of our data without storing each individual value
            - Know about population as a whole
        - Can tune for more privacy, but then the less granularity you will have in the data
- Privacy tools for the average user
    - Fooling classifiers by creating adversarial examples
        - Cutting edge image recognition
    - Just making a minor perturbation between images can fool classifiers, sometimes as little as a single pixel
    - Self driving car not recognizing a stop sign is a little scary
    - So you can future proof your pictures a bit using these techniques
- Some Advice For Job Searching
    - You are also interviewing the company
        - What drives you, what are you uncompromising about
        - "I believe in this, which is why I would be a good for this company"

- Fit matters, its not just technical stuff
- For your first job, having good mentors help
  - Great to have people around that can help you learn
  - Start up may be awesome, but it may be difficult
- You're not an imposter - no one knows everything. Be transparent about what you do and don't know. Be willing to learn