CS 5436 / INFO 5303 Spring 2023
Homework 4

Due: May 9, 11:59p ET

This is an INDIVIDUAL assignment.
You may discuss but each student must submit their own work.

COMMON FOR CS AND INFO

Problem 1 (6 points)
(a) Surname inference for anonymous genomes is similar to re-identifying anonymous datasets by record linkage. There is, however, an additional aspect of anonymous genomic data that makes it especially vulnerable to (re-)identification. What is it?

(b) What is the key type of auxiliary data for identifying anonymous genomes? How does the use of this data fundamentally rely on your answer to Problem 1(a)?

(c) Gymrek et al. paper on surname inference mentions meiosis. What is meiosis, and why is it important for identifying anonymous genomes? In your answer, refer to your answers to Problems 1(a) and 2(b).

Problem 2 (10 points)
A recent paper on data extraction from large language models shows that simply prompting these models with short texts can reveal potentially sensitive information. Explore this idea using Google Colab and come up with 5 different prompts that produce interesting, i.e. privacy sensitive, results. What were your criteria for choosing the prompts that worked for you?

Problem 3 (2 points)
Photography using Daguerreotype technology (circa 1839) required subjects to quietly pose for as long as 15 minutes. In 1890 Warren and Brandeis argued that "instantaneous photography violates privacy, (One of the most famous sentences in their article, *The Right to Privacy,*, "Instantaneous photographs and newspaper enterprise have invaded the sacred precincts of private and domestic life and numerous mechanical devices threaten to make good on the prediction that what is whispered in the closets shall be proclaimed on the rooftops.")

Use CI to pinpoint the source of privacy disruption.

Problem 5 (9 points)

(a) Why is consent considered an important (perhaps essential) aspect of prevailing privacy regulation in the US and Europe?

(b) What role does anonymity play in the prevailing privacy landscape?

(c) Provide two reasons why anonymity is important in free societies?

(d) Explain how the metaphor or "end run" is being applied to anonymity and consent.

(e) Briefly explain how the End Run argument (specifically) exposes a problem:

(i) with treating anonymity as a privacy solution. (Distinguish the end run problem with anonymity with what you've learned in this course about challenges with achieving anonymity in the first place.) Hint: "Your name is noise!"

(ii) with treating consent as sufficient for protecting privacy.

Problem 6 (5 points)

Prof. Knijnenburg discussed various techniques for usable privacy designs, as well as studies of behavior based on some commonly proposed designs.

(a) In this context, explain what we mean by "default" effects and "framing" effects
(b) Explain the Vortrex Newsletter findings using these terms
(c) What does 100% poison-free milk have to do with usable privacy?

Problem 7 (4 points)

(a) What does data minimization mean in privacy engineering?
(b) Why do privacy engineers claim that data minimization is the key principle underlying privacy by design?

Problem 8 (4 points)

(a) What is distinctive about DNA as a biometric (compared with, e.g. fingerprints)?
(b) What definition of privacy does this feature challenge? (Explain)

PICK YOUR OWN – should add up to 20 points
For each problem you pick, do the entire problem (i.e., you cannot choose-and-mix subproblems)

Problem CS1 (20 points)

Implement a simple federated learning framework based on the original paper and the provided Jupyter Notebook. Feel free to consult with the TA regarding implementation details and questions. Submit your modified notebook via Canvas.

Problem INFO1 (4 points)

The IRS is considering using facial recognition to verify your identity requiring individuals to set up and log into their accounts with a selfie. They believe it will be useful for preventing fraud. There have been loud objections to this idea.

Your job is to list conditions (design criteria) that such a system would need to hold in order to justify its use? (Hint: use CI parameters).

Problem INFO2 (2 points)

Gürses et al. remark that the "data minimization" metaphor could be misleading and confusing. Could you explain why?

Problem INFO3 (8 points)

Return to the End Run reading:

(a) Referring to the Target case, explain specifically how a company can make an “end run” around consent.

(b)(i) Briefly describe how an adherent of CI might, in specific cases, counteract an end run around anonymity?

(b)(ii) Barocas & Nissenbaum cite work of Neil Manson and Onora O’Neill to compare and contrast informed consent in biomedicine and consent as embodied in privacy policies for digital services and websites. What approach do they draw from these observations to counteracting the challenges of end run around consent?

Problem INFO4 (6 points)

City Hall is deciding whether to install facial recognition enhanced video cameras to cover all the surrounding areas, which are often used for political protests. Supporters argue that FR has become so commonplace that the norms have changed (E.g. Ring Doorbell, Apple ID). Critics say that just because something is commonplace doesn’t make it OK (morally justifiable.)

Drawing on CI develop and argument in favor of one of these positions. (Hint: (i) all the parameters matter! (ii) 3 layered analysis.)