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

Problem 1 (9 points)
(a) The goal of Interoperable Private Attribution (IPA) is to match certain “source” and “trigger” events that happen on users’ devices. Explain what the source and trigger events are, and why they are sensitive from the privacy perspective.

(b) Apple’s SKAN performs event matching on the device. By contrast, IPA performs event matching on the server. Which approach is superior from the advertisers’ perspective? Why? Which is superior from the privacy perspective?

(c) IPA uses both secure multi-party computation (SMC) and differential privacy (DP). Explain what information in IPA is protected by SMC, and what is protected by DP.

Problem 2 (9 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 2(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 2(a) and 2(b).

Problem 3 (12 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 4 (4 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 (10 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 an argument in favor of one of these positions. (Hint: all the parameters matter! 3 layered analysis.)

Problem 6 (4 points)
Nick Doty shared the following quotation in the context of internet-standard setting: “We reject: kings, presidents and voting. We believe in: rough consensus and running code.” Define multi-stakeholder governance, and provide one argument for why it is critical for standard setting?

Problem 7 (8 points)
Pop-ups, warning of possible security or privacy threats, don’t seem to alter users behavior. You are a researcher seeking to understand this phenomenon by applying empirical research methods.

Select two of the methods listed below, as described in Nathan Malkin’s lecture.
(a) Explain what these methods involves
(b) What you might be able to learn from them and what might not be learned using these methods.

Cognitive walkthrough
Lab study
Interview
survey,
Field observation
Field intervention

Problem 8 (4 points)
What is most distinctive about DNA as a biometric (compared with, e.g. fingerprints)?

BONUS (+2 points) What definition of privacy does this feature challenge? (Explain)
Problem CS1 (30 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.

INFO ONLY

Problem INFO1 (total 8 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 (total 22 points)

(a) What is an internet standard? (2 points)
(b) How are standards developed? (2 points)
(c) Pick one of the following internet standards. Based on your chosen standard, outline 3 privacy considerations the standard outlines. (6 points)

1. https://www.w3.org/TR/webrtc/
2. https://www.w3.org/TR/geolocation/
3. https://www.unicode.org/versions/Unicode14.0.0/
4. https://www.w3.org/TR/activypub/
6. https://www.w3.org/TR/CSS2/
8. https://262.ecma-international.org/12.0/

(d) Do these three privacy considerations comply with the CI framework? Explain your answers carefully for each case. (12 points)