The goal of the Privacy and Civil Liberties (PCL) team is to design, build, and deploy privacy-protective technologies, and to foster a culture of responsibility around their development and use.
Why is this problem important to solve?

Organizations struggle with data classification.
Detecting & subsequently managing sensitive information is **difficult**.

- The term "sensitive" can take very many definitions, making it difficult to generalize automation across industries
- Sensitivity of data is context specific
- Access Controlling sensitive data en masse can be too restrictive
- Translating a data protection policy to a technical constraint requires significant collaboration between legal & engineering teams
**Solution Skeleton**

**Detection:** Provide a simple and customizable way of defining “Sensitive Data”.

**Management:** Provide an easy and effective interface for specifying what the system should do when “Sensitive Data” is found.
Data Protection Units can use **regexes** to establish organization-specific definitions of sensitive data.
Data Protection Units can use **value overlap** to find exact matches in unstructured data. This is great for names, or even ID numbers.

“My name is Robert **Quinn** and I’ve been an employee for 5 years. I’m commenting here to ask …”
Management – ACCESS CONTROL –

Data Protection Units lock down sensitive data to only select users.
Data Protection Units can **minimize** sensitive data by encrypting, hashing or redacting it.
The solution still needs to do more.

- Accuracy and cost of false positives
- Scale and compute considerations
- How about automation by leveraging the power of ML and NLP techniques?
- Regulatory Restrictions
- When sensitive data entails protected attributes, it can lead to fairness implications in data science pipelines
Key Take-aways

• Most technical systems ingest and store sensitive information; it’s challenging to identify such data and appropriately tackle it via a generalized product solution

• B2B PETs needs to be built to serve more dynamic use-cases than B2C

• A thoughtful software engineering approach is required to build robust and flexible PETs

• Data management solutions should be built with Data Protection users in mind while facilitating collaboration between Data Engineering Teams & DPUs