In a nutshell

• What’s the problem? More, better data and analysis

• Why is that a problem? More inferences, less privacy

• Why is this important? Need trustworthy systems

• What to do about it? PETs, multidisciplinary work
What’s the problem? More, better data and analysis

- More sensors, digitalization, devices, connections, services
- More efficient (cheaper) storage, computation
- Advances in machine learning, autonomous systems
- Concerns: quantity, quality of data and analysis, incentives
Why is that a problem? More inferences, less privacy

- Correct and incorrect classification, both problematic
- More generally, Solove’s taxonomy of privacy harms
ML-ML privacy threats

- Membership inference
- Model inversion

- Silver lining: techniques to avoid overfitting also good for privacy
Why is this important?
Need trustworthy systems

- Privacy-utility trade-off
- Fairness, accountability, transparency (FAT)
- Security: only authorized users/use. Prevention, detection, response.
- Human factor: usability, processes, policies, values
What to do about it? PETs, multidisciplinary work

Privacy-enhancing technologies (PETs). Can’t agree on a definition of privacy, but CS notion of some properties to do with indistinguishability (Pfitzmann, Hansen):

- anonymity
- pseudonymity
- unlinkability
PETs toolkit

- Models: system, properties, adversary
- Advanced cryptography
- Decentralization, local processing
- Adding noise
- Indirection
Examples from our research

- AnonFACES, k-anonymity for image data
- Synthetic data for analysis of structured data
- DataLEASH, multidisciplinary effort for metrics and mechanisms for data learning and sharing under privacy constraints
- Applied crypto: Distance-bounding attribute-based credentials, crowd counting using smartphones, deniable messaging
Education

- More pervasive privacy education in computer science
- Contextualization: ethics, law, society
Take-away

• It’s complicated
• PETs can be comforting, FAT PETs exist
• Let’s talk