AI & USEC

INFR11158/11230 Usable Security and Privacy

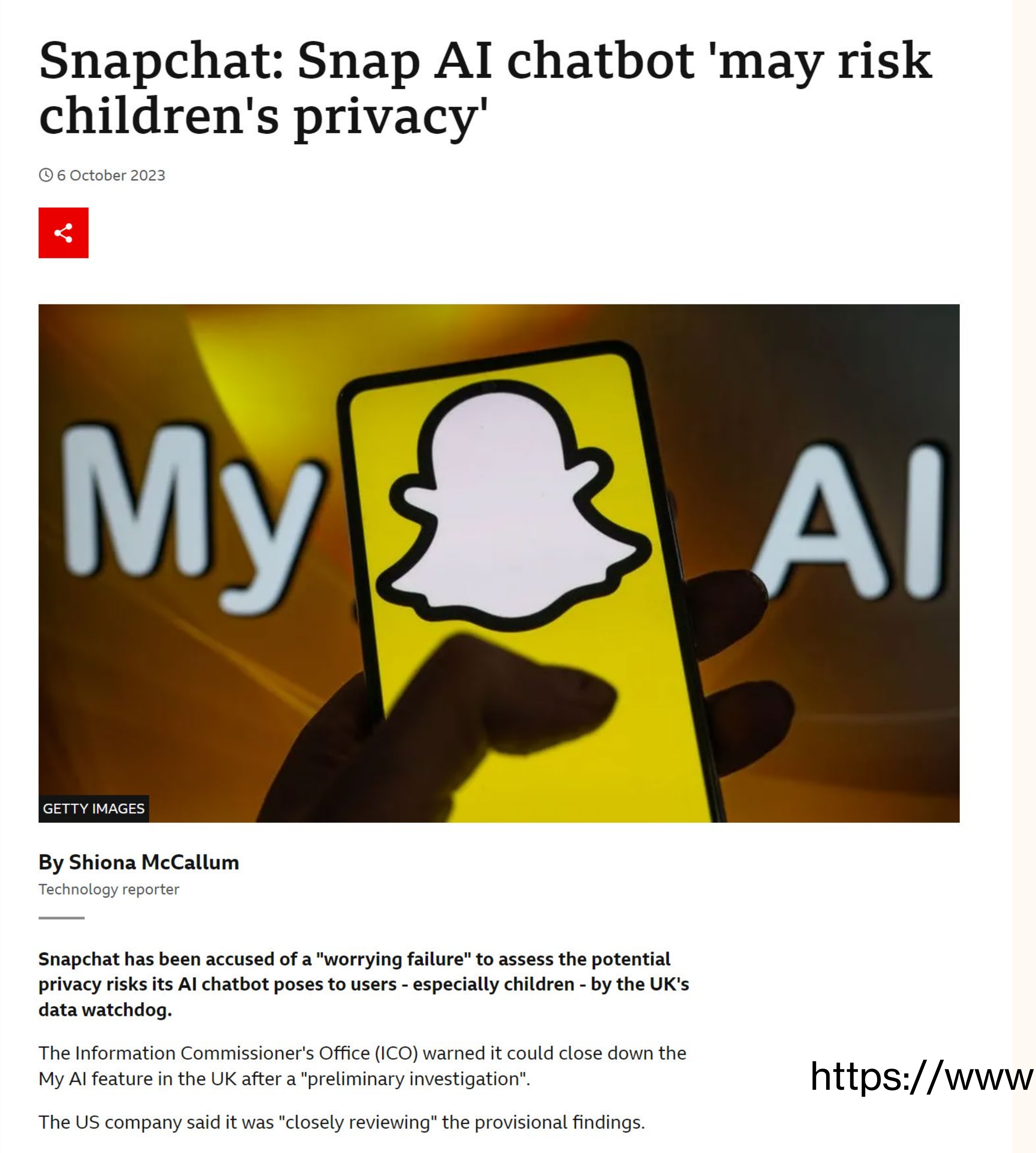
Dr. Jingjie Li

26/03/2024

THE UNIVERSITY *of* EDINBURGH

Overview

- Finally week**s**
- Privacy issues of AI
- Recap – privacy policy
- Guest lecture



Snapchat: Snap AI chatbot 'may risk children's privacy'

6 October 2023

GETTY IMAGES

By Shiona McCallum

Technology reporter

Snapchat has been accused of a "worrying failure" to assess the potential privacy risks its AI chatbot poses to users - especially children - by the UK's data watchdog.

The Information Commissioner's Office (ICO) warned it could close down the My AI feature in the UK after a "preliminary investigation".

The US company said it was "closely reviewing" the provisional findings.

https://www.bbc.co.uk/news/technology-67027282

Privacy risks of AI

A TAXONOMY OF PRIVACY

INFORMATION PROCESSING

AGGREGATION
Combining of various pieces of personal information

A credit bureau combining an individual's payment history from multiple creditors.

SECONDARY USE
Using personal information for a purpose other than the purpose for which it was collected

The U.S. Government using census data collected for the purpose of apportioning Congressional districts to identify and intern those of Japanese descent in WWII.

EXCLUSION
Failing to let an individual know about the information that others have about them and participate in its handling or use

A company using customer call history, without the customer's knowledge, to shift their order in a queue (i.e. "Your call will be answered in the order [NOT] received")

INSECURITY
Failing to protect information

An ecommerce website allowing others to view an individual's purchase history by changing the URL (e.g. enterprivacy.com?id=123)

IDENTIFICATION
Linking of information to an individual. [Sometimes called 'singling out']

A researcher linking medical files to the Governor of a state using only date of birth, zip code and gender.

COLLECTION

SURVEILLANCE
Watching, listening to, or recording of a person's activities

A website monitoring cursor movements of a visitor while visiting the website.

INTERROGATION
Questioning or probing for personal information

An interviewer asking an inappropriate question, such as marital status, during an employment interview.

INVASION

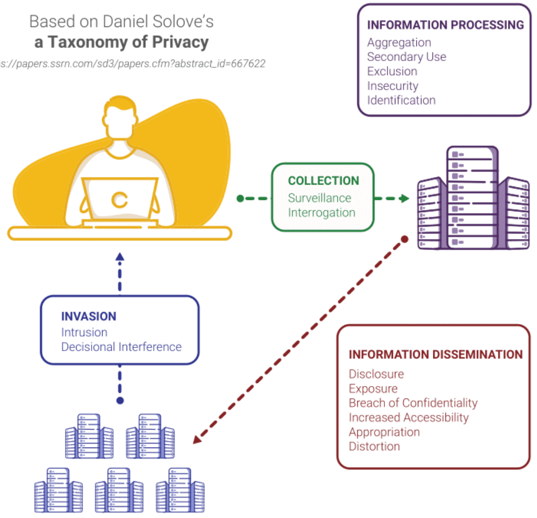
INTRUSION
Disturbing a person's tranquility or solitude

An augmented reality game directing players onto private residential property.

DECISIONAL INTERFERENCE
Intruding into a person's decision making regarding their private affairs

A payment processor declining transactions for contraceptives.

Based on Daniel Solove's
a Taxonomy of Privacy
https://papers.ssrn.com/sd3/papers.cfm?abstract_id=667622



INFORMATION PROCESSING
Aggregation
Secondary Use
Exclusion
Insecurity
Identification

COLLECTION
Surveillance
Interrogation

INVASION
Intrusion
Decisional Interference

INFORMATION DISSEMINATION
Disclosure
Exposure
Breach of Confidentiality
Increased Accessibility
Appropriation
Distortion

INFORMATION DISSEMINATION

DISCLOSURE
Revealing truthful information about a person that impacts their security or the way others judge their character

A government agency revealing an individual's address to a stalker, resulting in the individual's murder.

EXPOSURE
Revealing a person's nudity, grief, or bodily functions

A store forcing a customer to remove clothing revealing a colostomy bag.

BREACH OF CONFIDENTIALITY
Breaking a promise to keep a person's information confidential.

A doctor revealing patient information to friends on a social media website.

INCREASED ACCESSIBILITY
Amplifying the accessibility of personal information

A court making proceeding searchable on the Internet without redacting personal information.

APPROPRIATION
Using an individual's identity to serve the aims and interests of another

A social media site using customer's images in advertising.

DISTORTION
Disseminating false or misleading information about a person

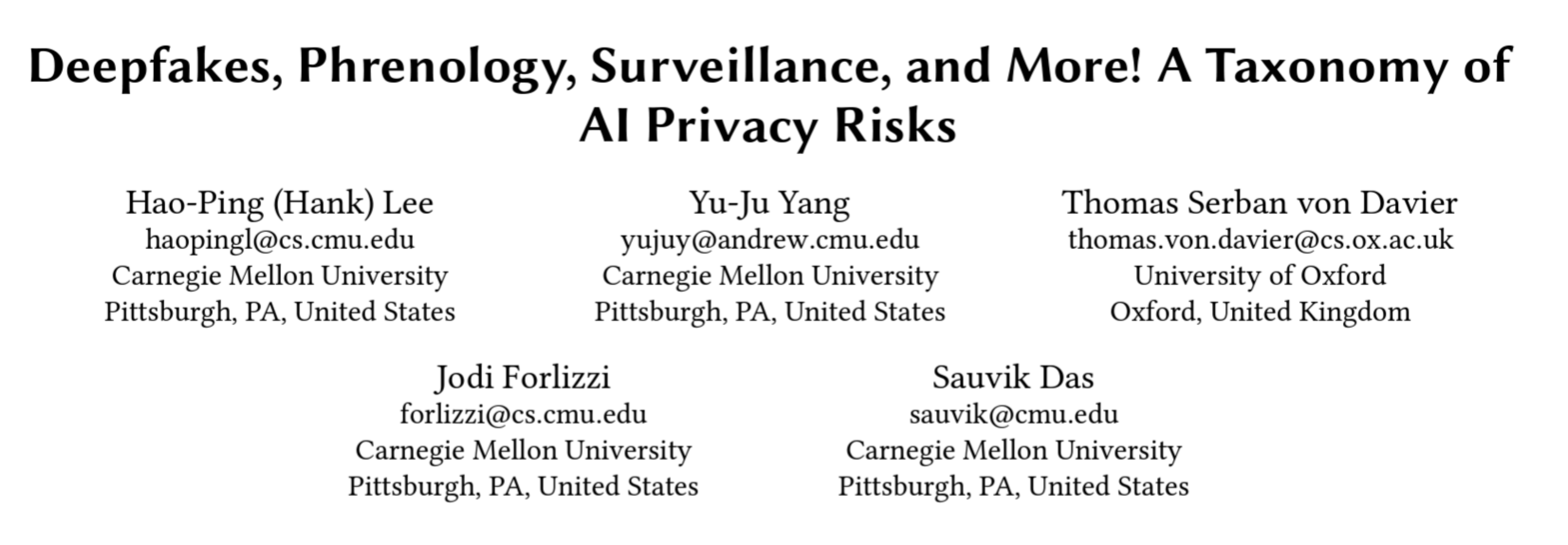
A creditor reporting a paid bill as unpaid to a credit bureau.

cc BY NC ND

Version 6 (2022)

PRIVACY BY DESIGN

https://privacybydesign.training



Deepfakes, Phrenology, Surveillance, and More! A Taxonomy of AI Privacy Risks

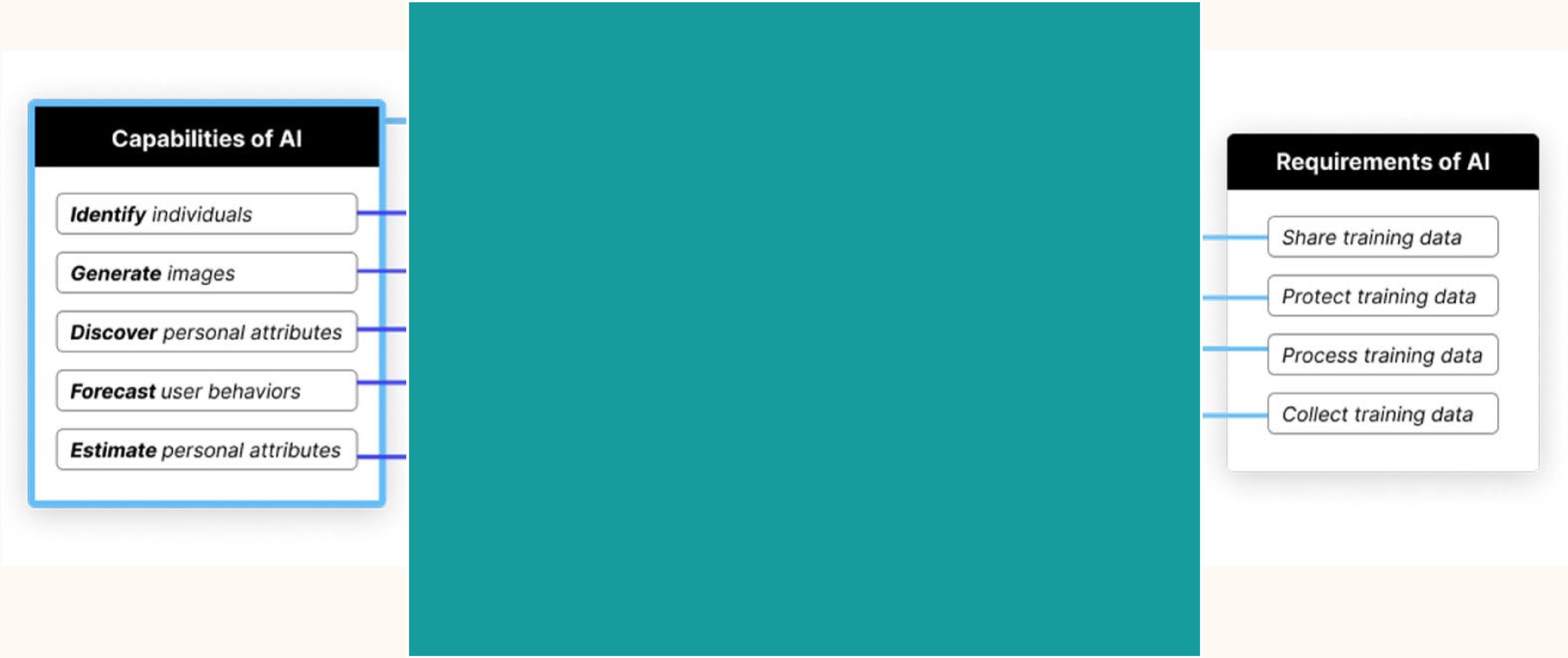
Hao-Ping (Hank) Lee
haopingl@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, PA, United States

Yu-Ju Yang
yujuy@andrew.cmu.edu
Carnegie Mellon University
Pittsburgh, PA, United States

Thomas Serban von Davier
thomas.von.davier@cs.ox.ac.uk
University of Oxford
Oxford, United Kingdom

Jodi Forlizzi
forlizzi@cs.cmu.edu
Carnegie Mellon University
Pittsburgh, PA, United States

Sauvik Das
sauvik@cmu.edu
Carnegie Mellon University
Pittsburgh, PA, United States



Capabilities of AI

- ***Identify*** *individuals*
- ***Generate*** *images*
- ***Discover*** *personal attributes*
- ***Forecast*** *user behaviors*
- ***Estimate*** *personal attributes*

Requirements of AI

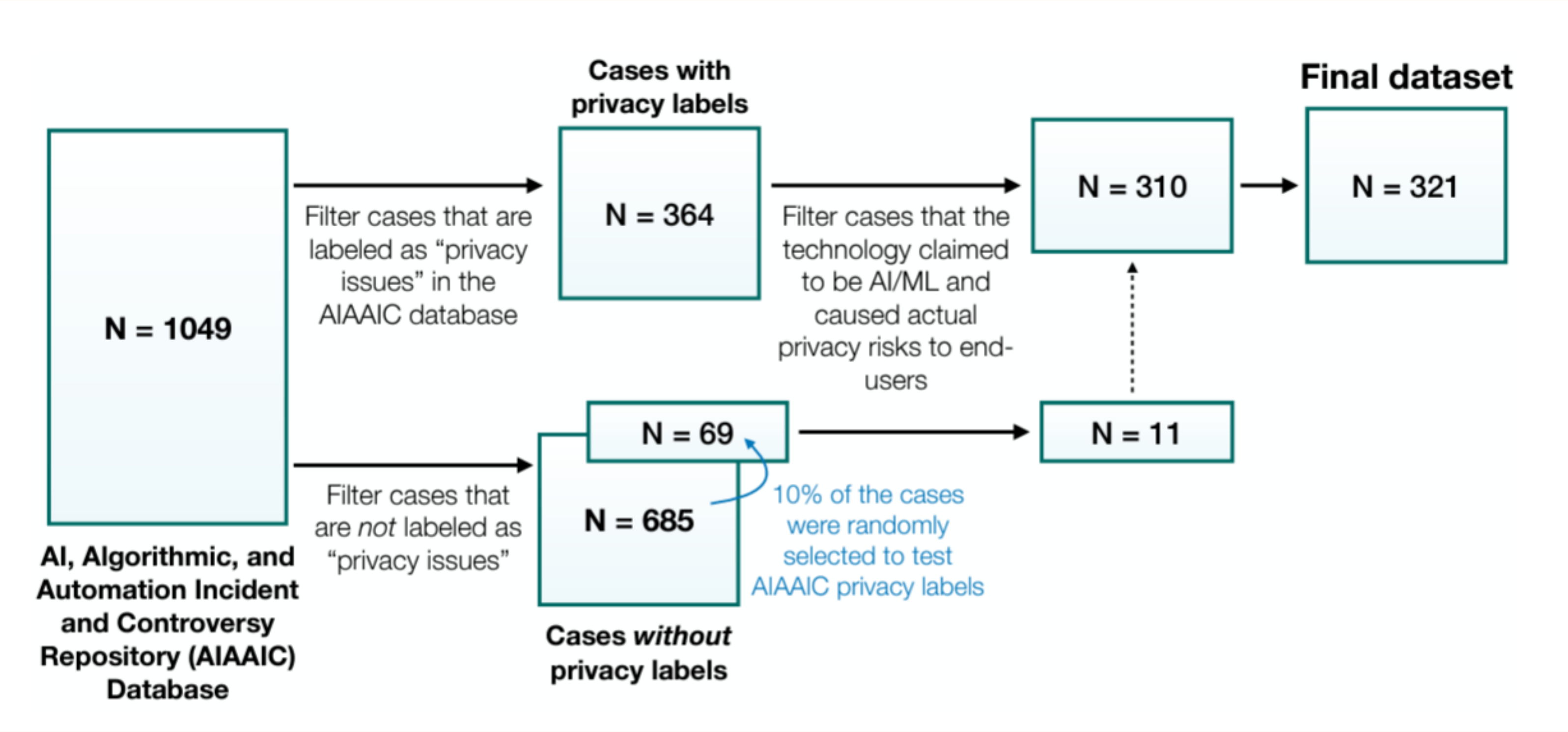
- *Share training data*
- *Protect training data*
- *Process training data*
- *Collect training data*

Objective

- Develop a privacy taxonomy for AI privacy risks
- What's AI?
  - "perform tasks or behaviors that a person could reasonably deem to require intelligence if a human were to do it" – an umbrella definition

Method

- Materials: AI incident database
- Approach: qualitative coding and analysis
  - Top-down/deductive coding: Solove's privacy taxonomy


Cases with privacy labels
Final dataset
N = 1049
Filter cases that are labeled as "privacy issues" in the AIAAIC database
N = 364
Filter cases that the technology claimed to be AI/ML and caused actual privacy risks to end-users
N = 310
N = 321
N = 69
N = 11
Filter cases that are *not* labeled as "privacy issues"
N = 685
10% of the cases were randomly selected to test AIAAIC privacy labels
AI, Algorithmic, and Automation Incident and Controversy Repository (AIAAIC) Database
Cases *without* privacy labels

Data flow

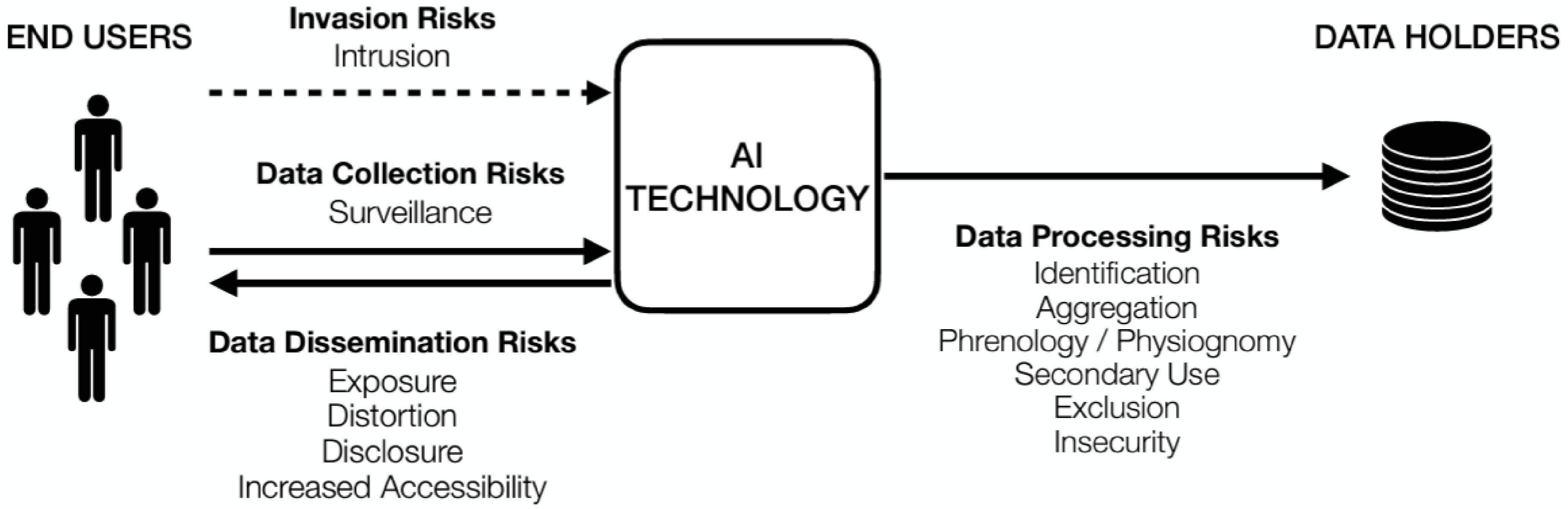
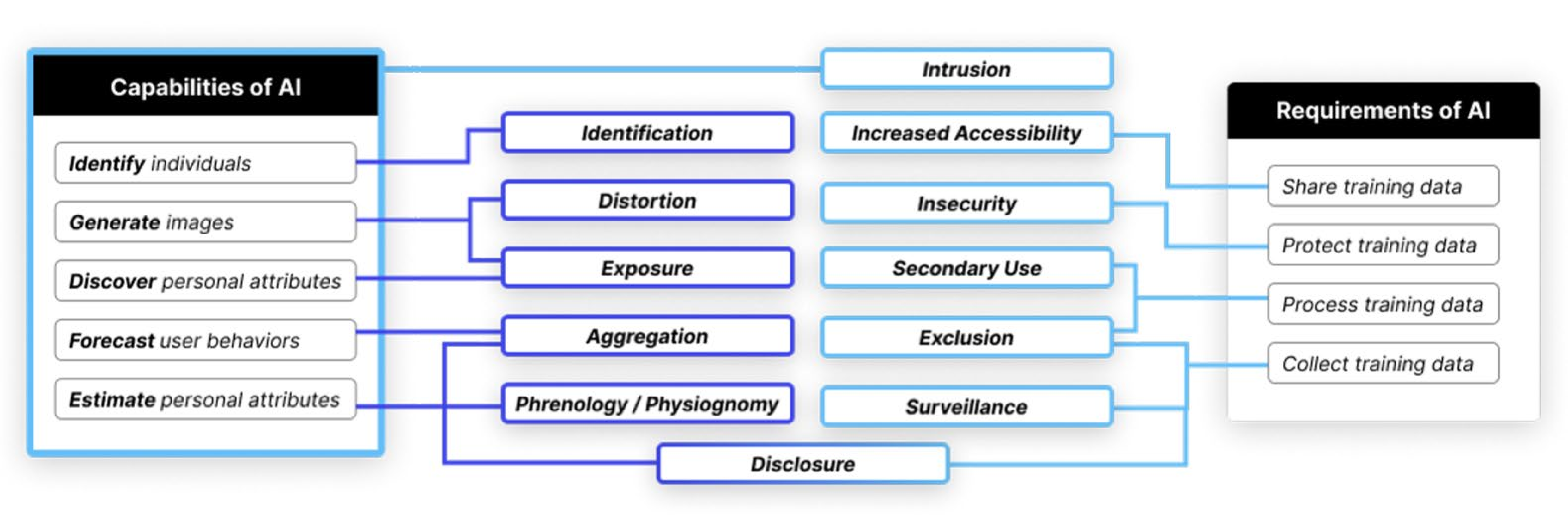


Figure 3: 12 types of privacy risks that AI technologies create and/or exacerbate relate to data collection, data processing, data dissemination, and invasion. The arrows indicate data flow (invasion risks need not involve data, but often do).


Capabilities of AI
Identify individuals
Generate images
Discover personal attributes
Forecast user behaviors
Estimate personal attributes
Intrusion
Identification
Increased Accessibility
Distortion
Insecurity
Exposure
Secondary Use
Aggregation
Exclusion
Phrenology / Physiognomy
Surveillance
Disclosure
Requirements of AI
Share training data
Protect training data
Process training data
Collect training data

Are privacy policies easy to make?

How to Create a
Privacy
Policy

How can we do better?

Guest lecturer: Shidong Pan, PhD student @ Australian National University, (USEC, privacy policy, responsible AI & software engineering)