Tuesday, November 5, 8:30 AM - 9:30 AM *Data Privacy*

Elwha

John Abowd, Chief Scientist, U.S. Census Bureau
John Kahan, Chief Data Analytics Officer, Microsoft
Shiva Kasiviswanathan, Computer Scientist, Amazon
Aleksandra Korolova, WiSE Gabilan Assistant Professor of Computer Science, USC
Paul Liu, Principal Economist, Google

Moderator: **John Abowd**

John M. Abowd
Chief Scientist, U.S. Census Bureau

United States™ Census Bureau
U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

The views expressed in this talk are my own and not those of the U.S. Census Bureau.

Estimate: 15GB of final data from 2020 Census ($1/byte!)

Less than 1% of worldwide mobile data use/second

(Source: Cisco VNI Mobile, February 2019 estimate: 11.8TB/second, 29EB/month, mobile data traffic worldwide https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/white-paper-c11-738429.html#_Toc953327.)

The Census Bureau's data stewardship problem looks very different from the one at Amazon, Apple, Facebook, Google, Microsoft, Netflix, Uber, …

… but appearances are deceiving.

Privacy protection is an economic problem.

Not a technical problem in computer science or statistics.

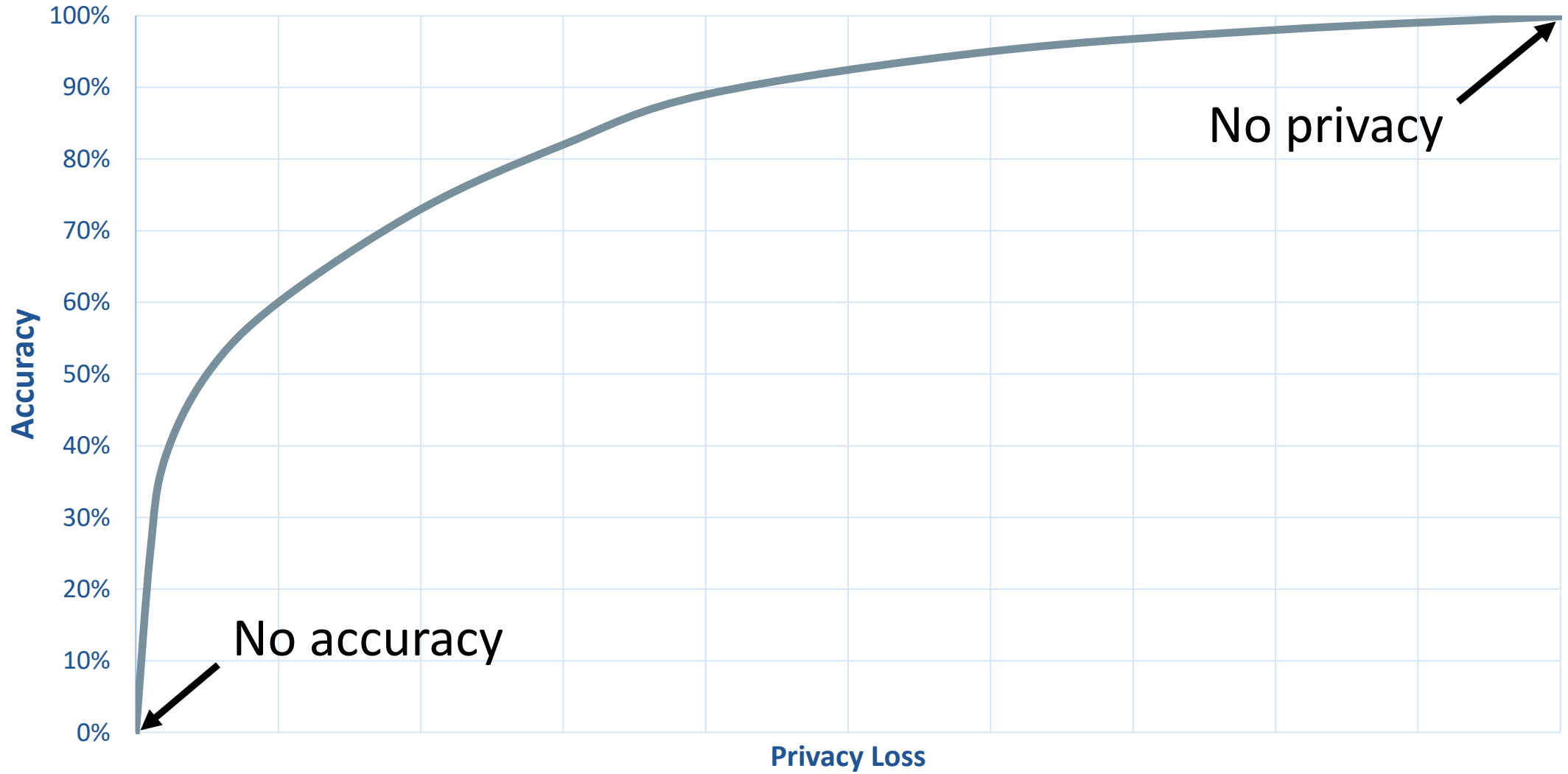
Allocation of a scarce resource (data in the confidential database) between competing uses:

information products

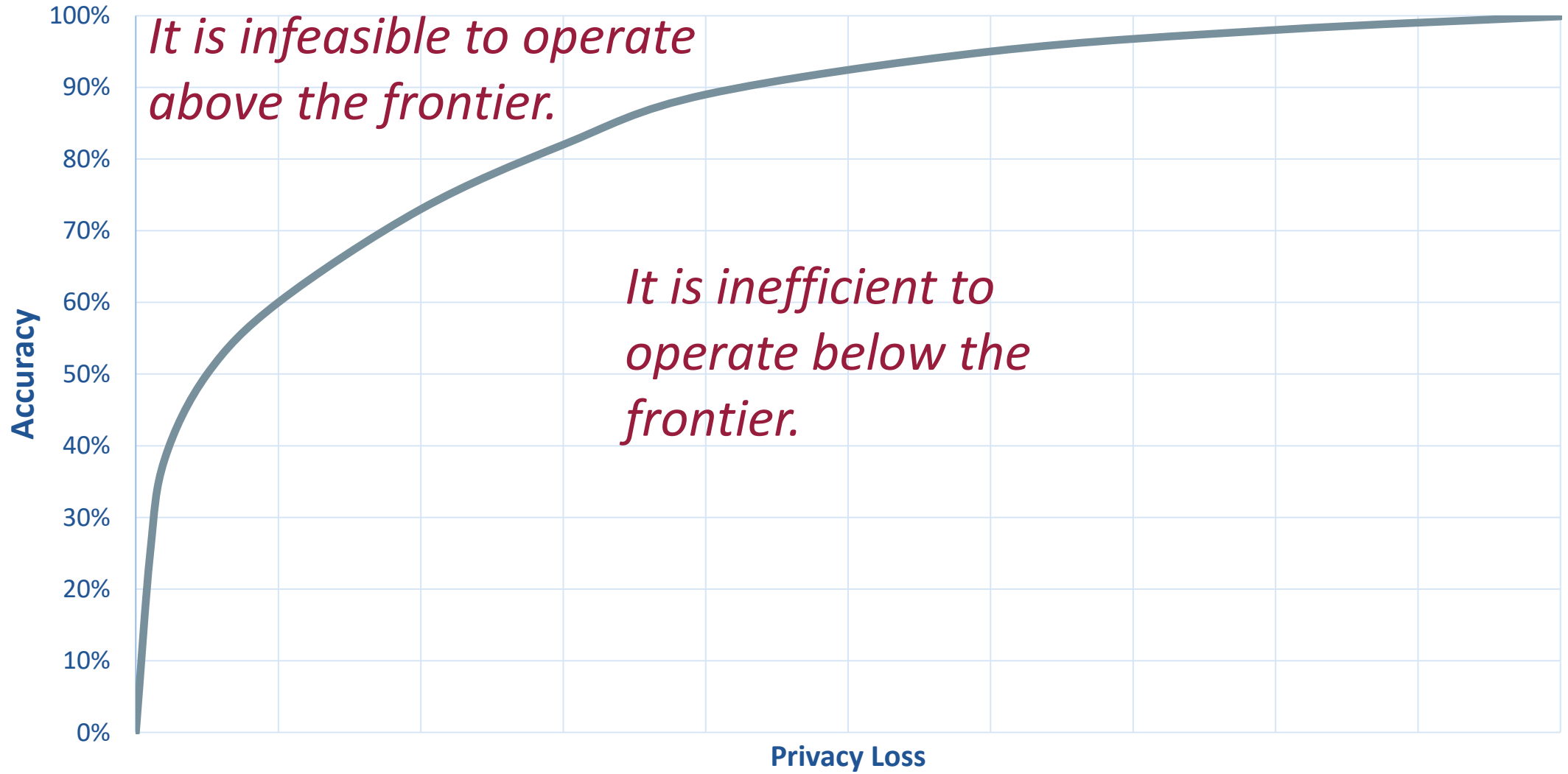
and

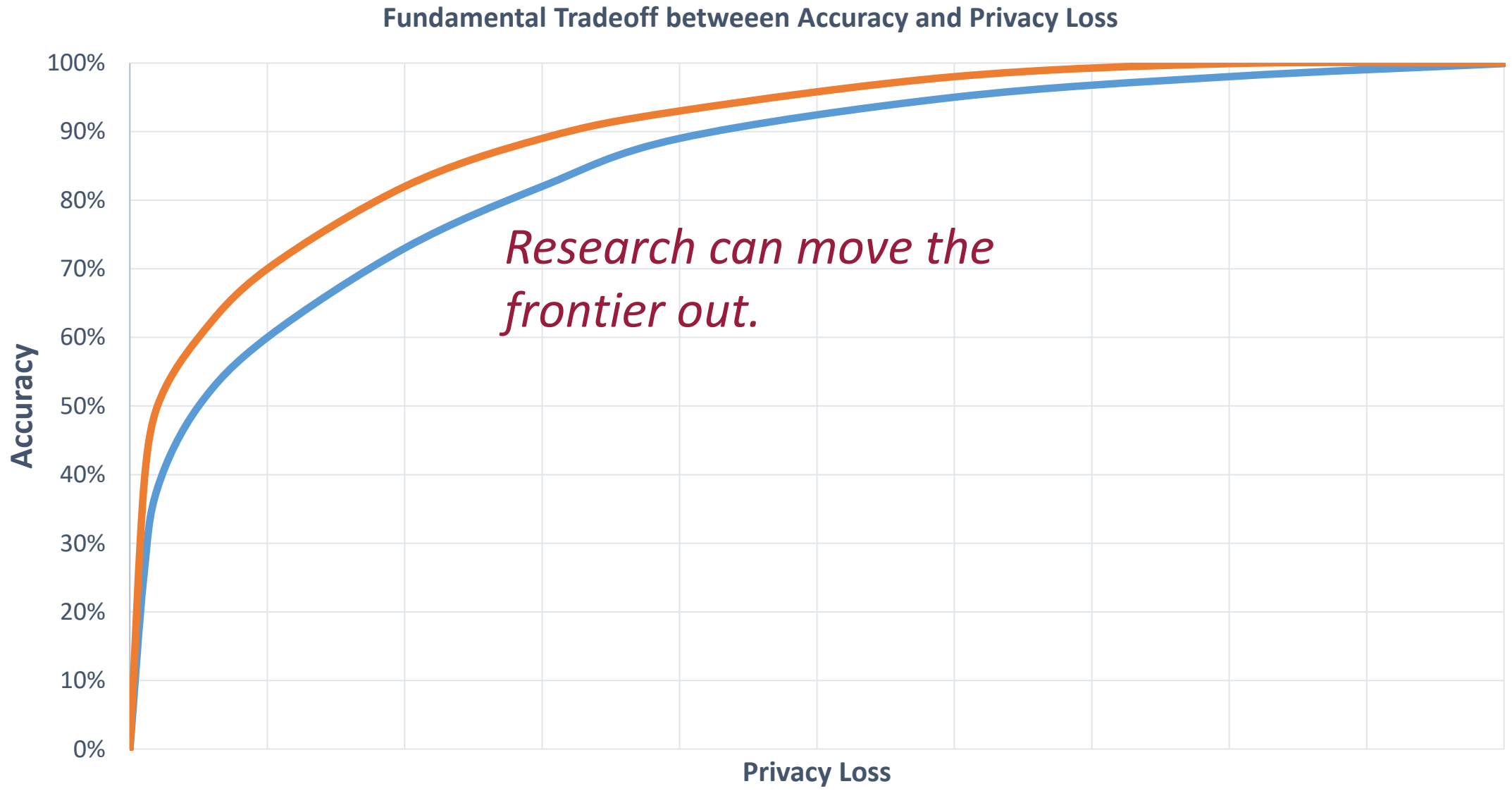
privacy protection.

Fundamental Tradeoff betweeen Accuracy and Privacy Loss

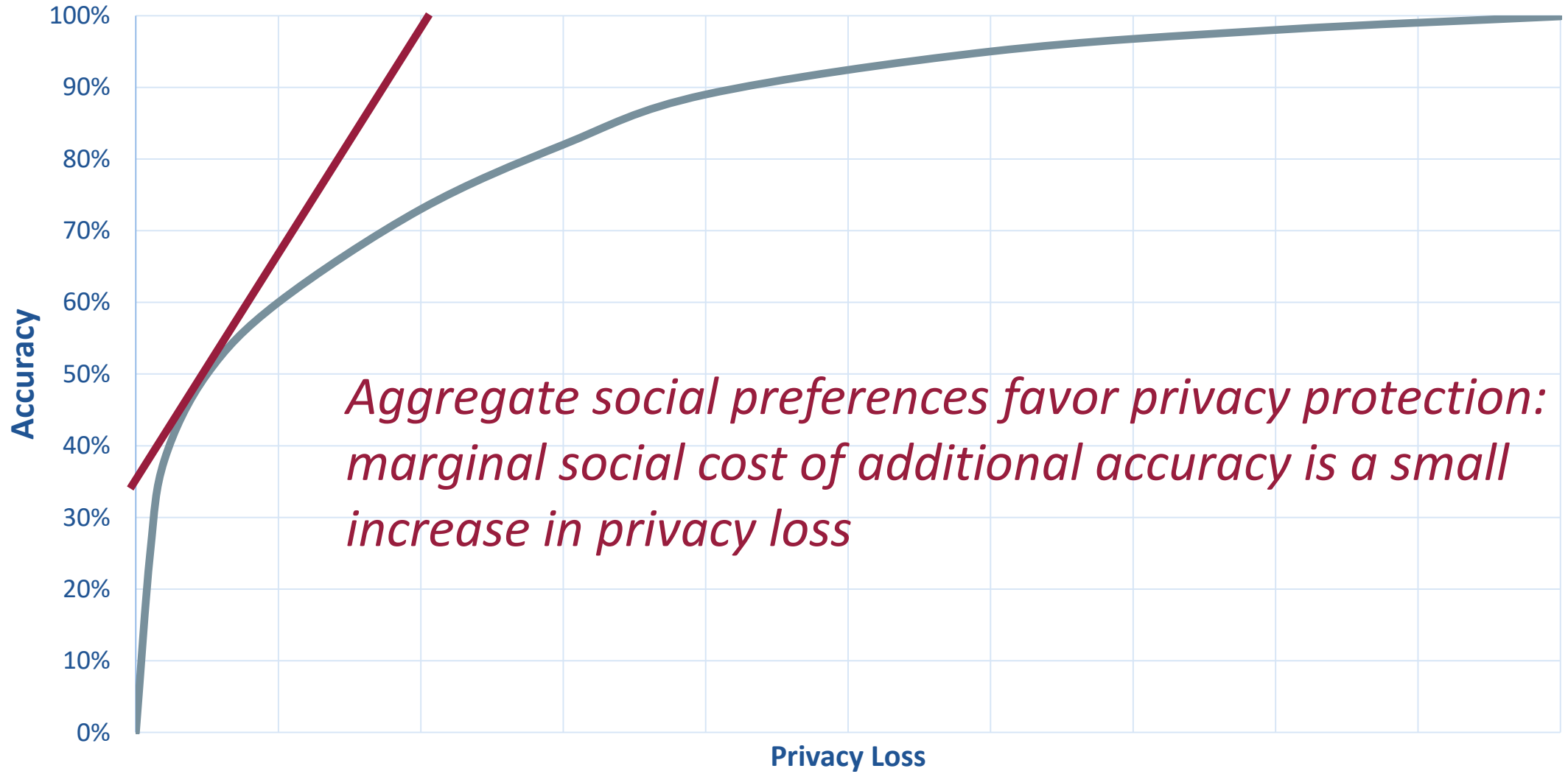


Fundamental Tradeoff betweeen Accuracy and Privacy Loss



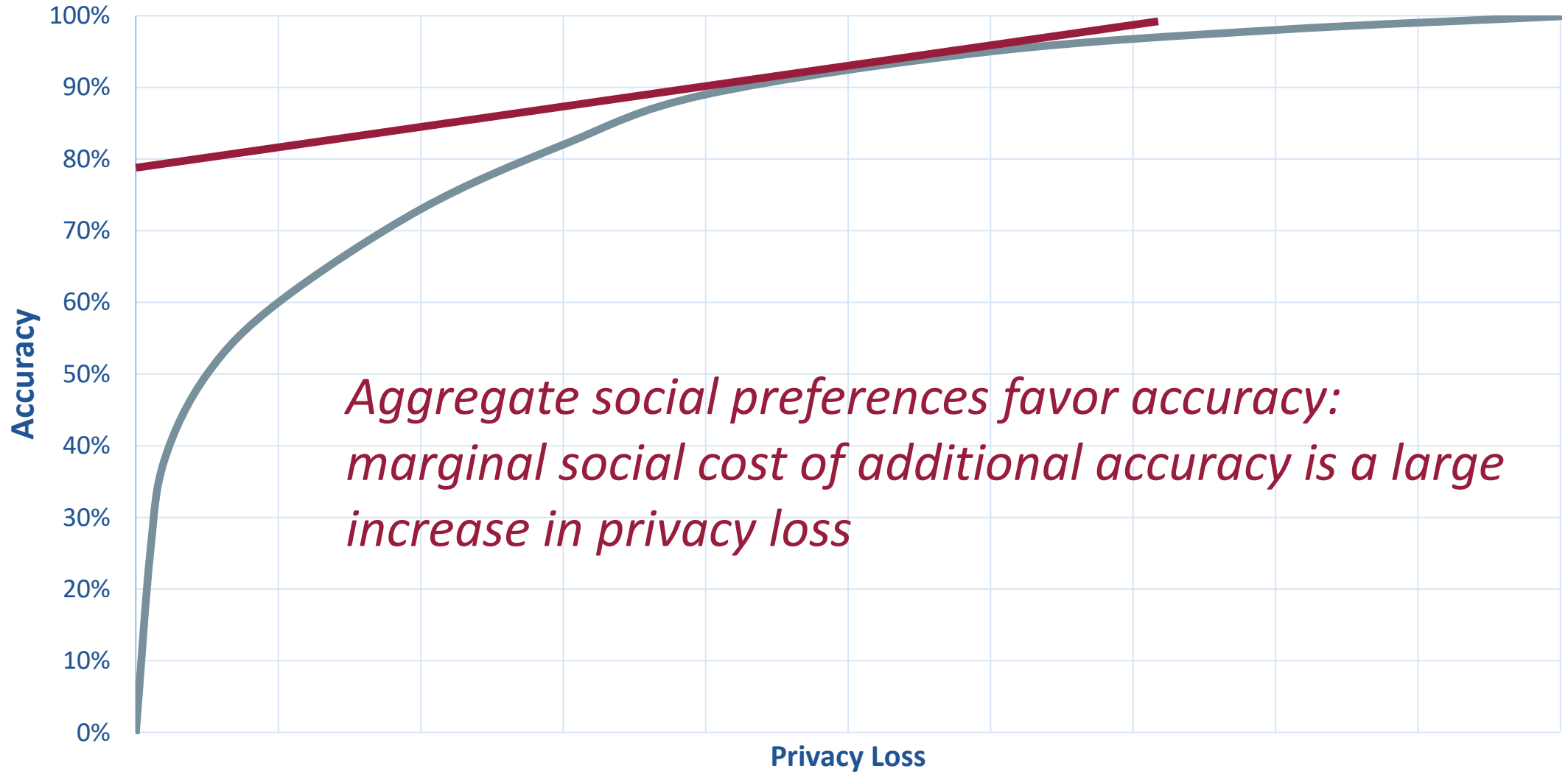

Fundamental Tradeoff betweeen Accuracy and Privacy Loss
100%
90%
80%
70%
60%
50%
40%
30%
20%
10%
0%
Accuracy
Privacy Loss
Research can move the frontier out.

Fundamental Tradeoff betweeen Accuracy and Privacy Loss

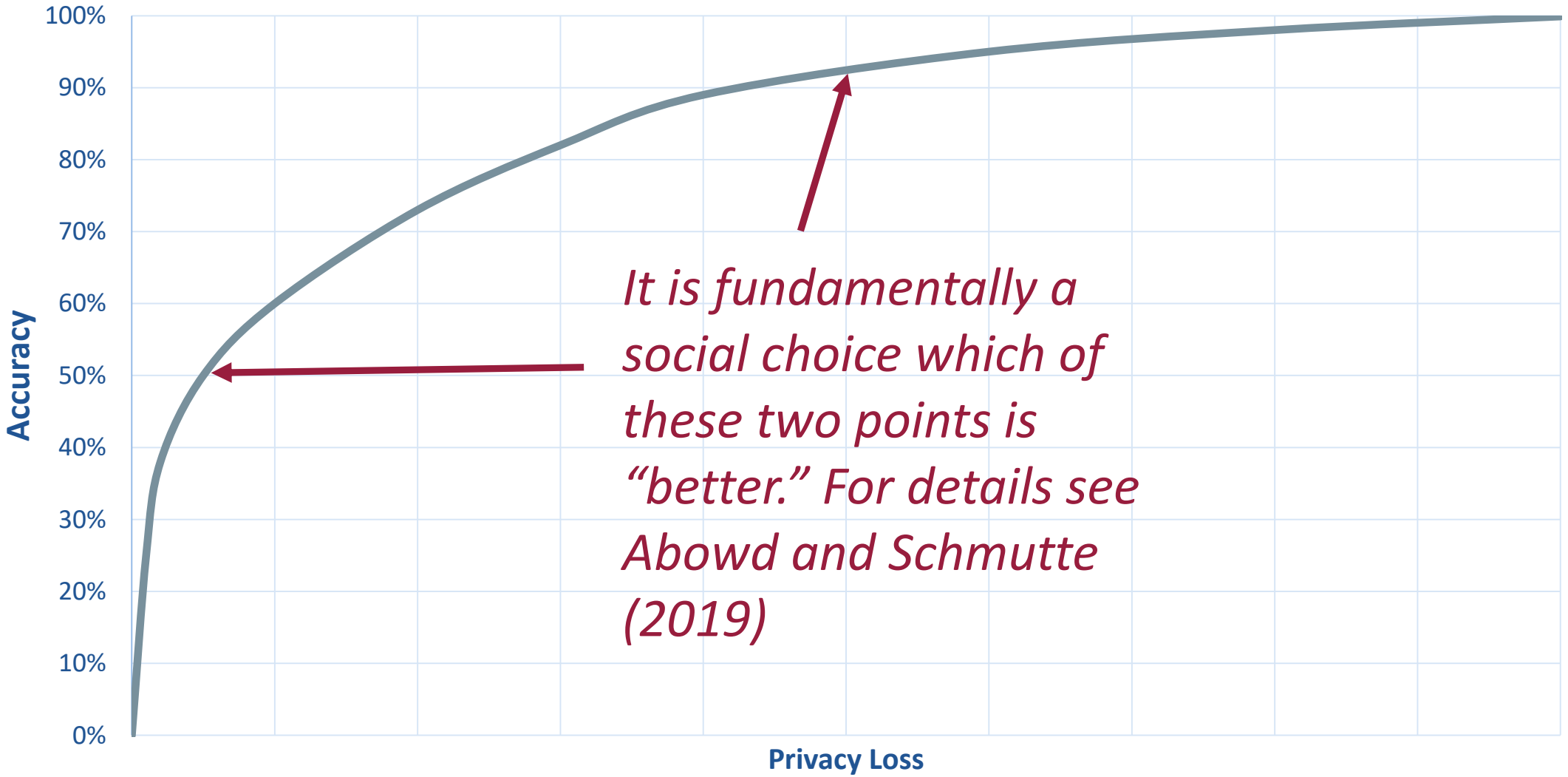


Aggregate social preferences favor privacy protection: marginal social cost of additional accuracy is a small increase in privacy loss

Fundamental Tradeoff betweeen Accuracy and Privacy Loss



Fundamental Tradeoff betweeen Accuracy and Privacy Loss



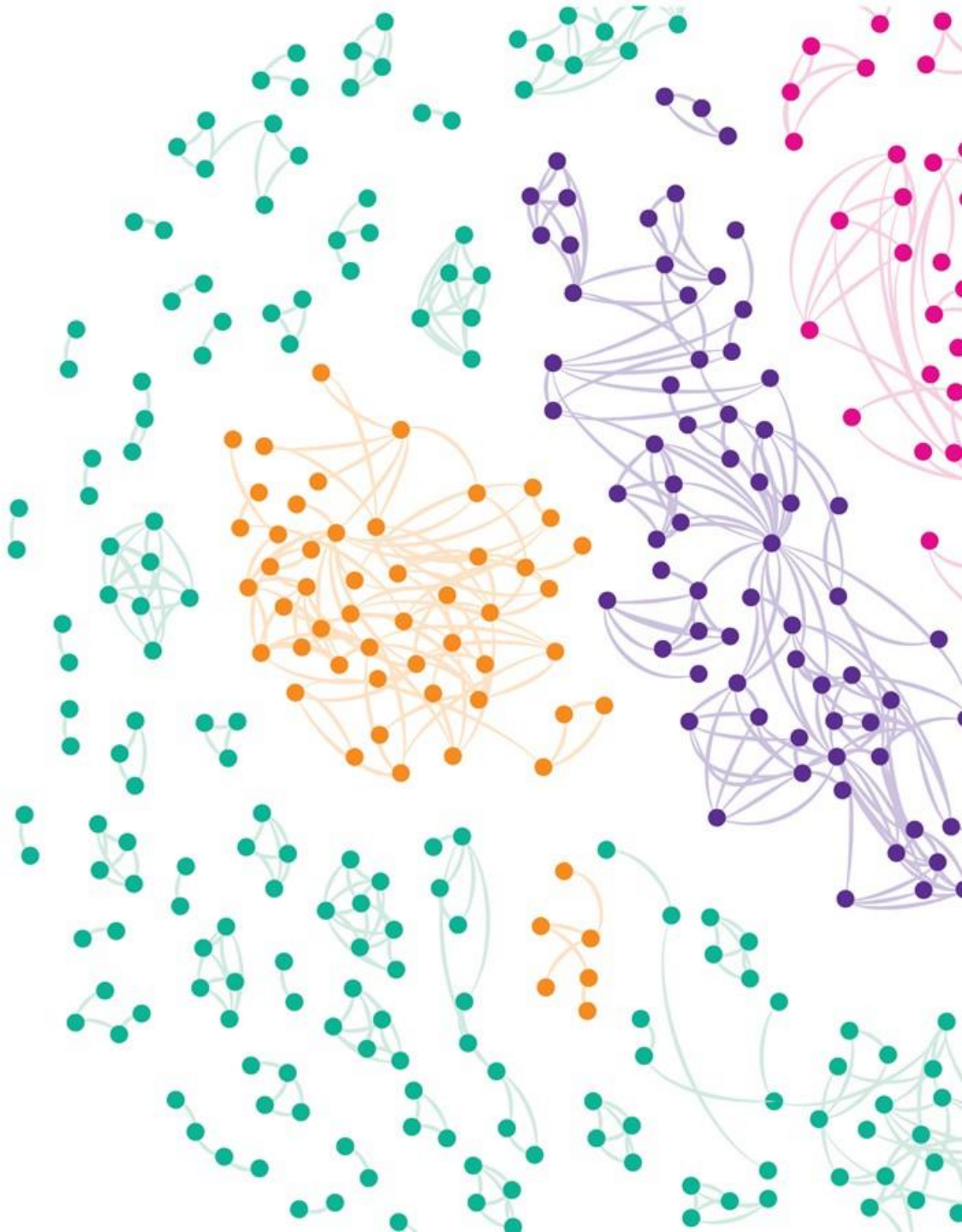
All 2020 Census Publications

- Will all be processed by a collection of differentially private algorithms
- Using a total privacy-loss budget set as policy, not hard-wired
- Code base, technical documents, and extensive demonstration products based on the 2010 Census confidential data have all been released to the public
- More information: https://www.census.gov/newsroom/blogs/research-matters/2019/10/balancing_privacyan.html

Microsoft

Infusing Data Science and AI to address the world's great challenges

John Kahan
Chief Data Analytics Officer



United States™ Census Bureau

U.S. Department of Commerce
Economics and Statistics Administration
U.S. CENSUS BUREAU
census.gov

Privacy Research @ Amazon

Shiva Kasiviswanathan

Research Goals

- Deploy data systems that implement privacy by design
- Educate company on privacy technologies

A lot of our work is based on using Differential Privacy….

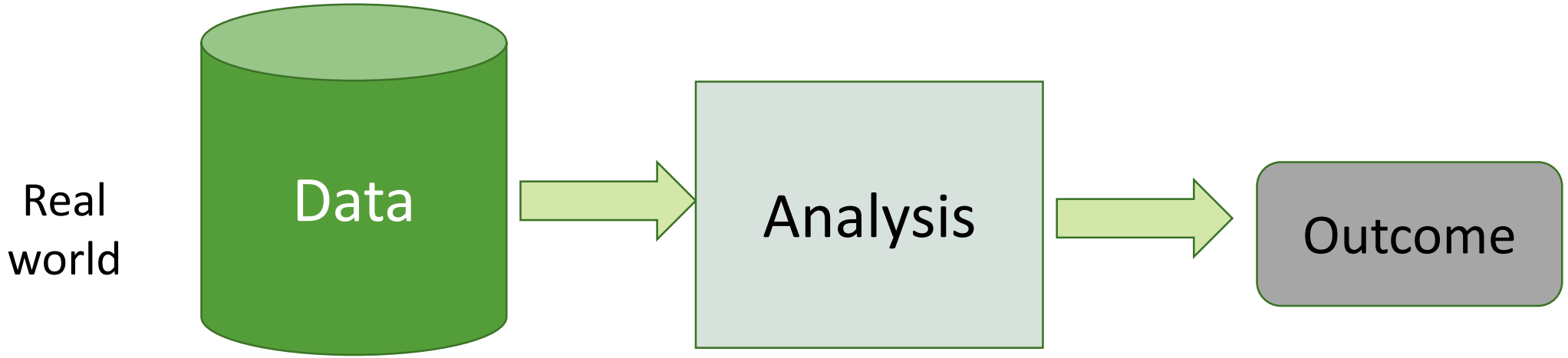
Differential Privacy is…..

... a definition (i.e., a standard) of privacy

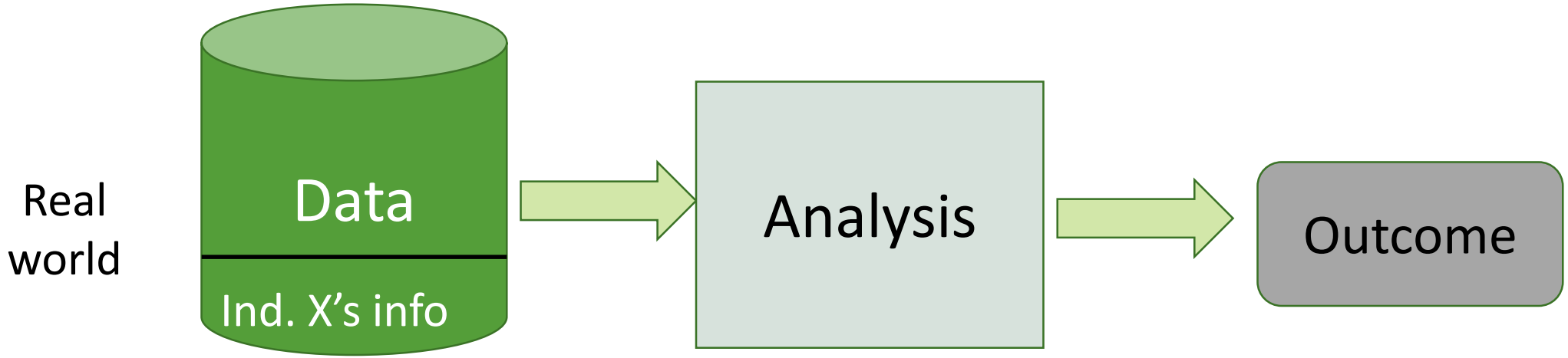
It expresses a specific desiderata of an analysis:

Any information-related risk to a person should not change significantly as a results of that person's information being included,
or not, in the analysis

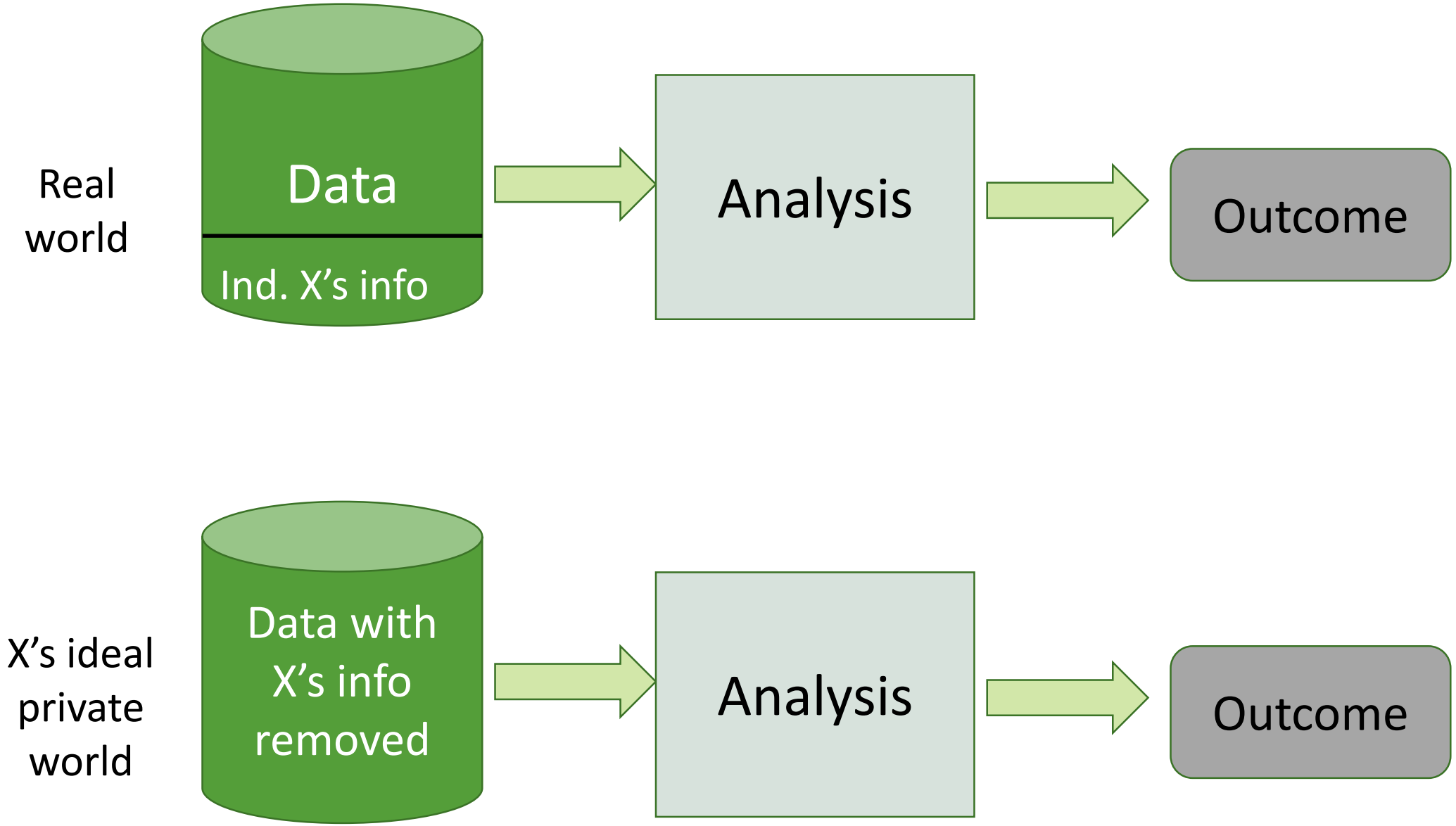
Differential Privacy [Dwork et al. 2006]



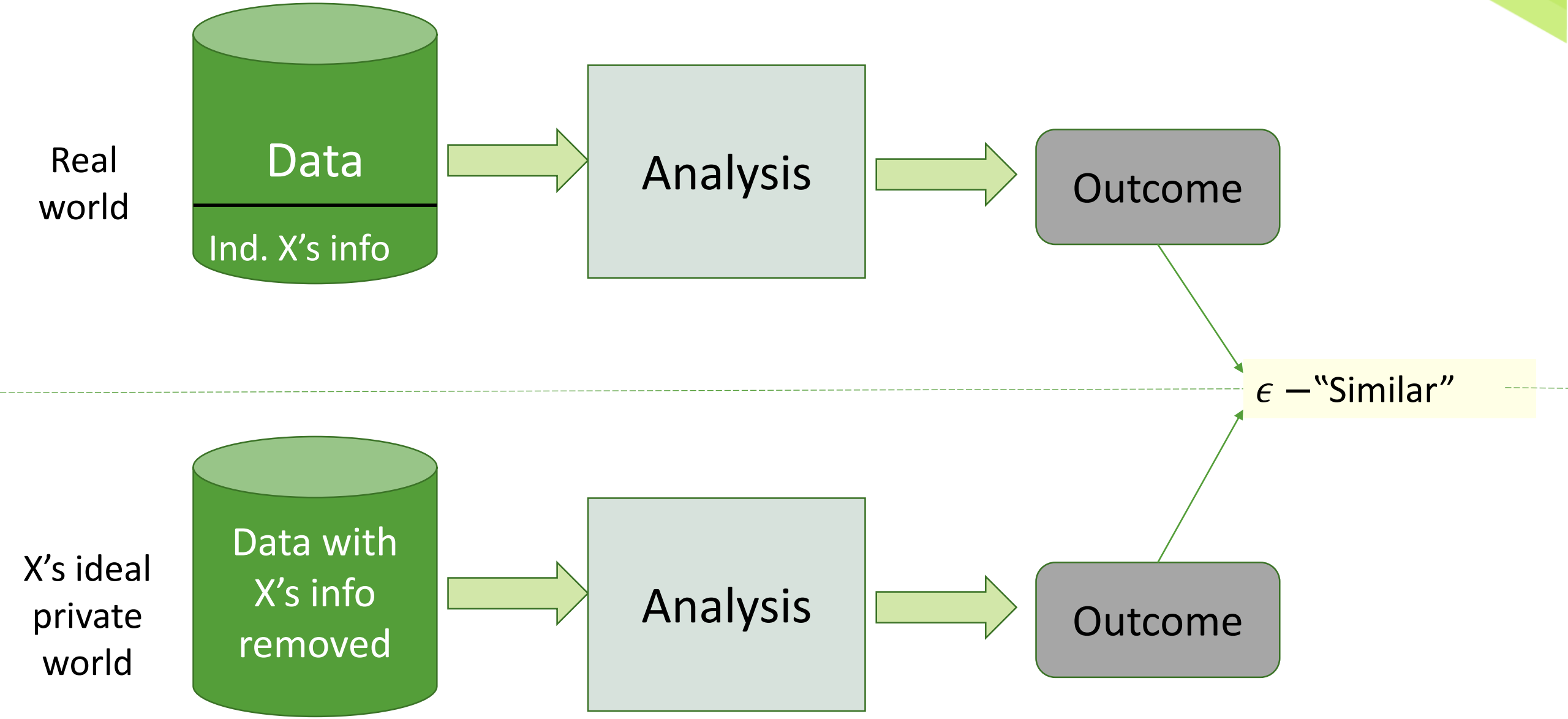
Differential Privacy [Dwork et al. 2006]



Differential Privacy [Dwork et al. 2006]



Differential Privacy [Dwork et al. 2006]



What data analyses can we successfully do under Differential Privacy….

Privacy Projects @ Amazon

1) A Natural Language Processing Application

2) A General Purpose Differential Privacy Package

Differential Private Text Perturbation*

Task: Sanitize sensitive text

Perturb sentences while maintaining its semantic meaning

e.g., “goalie wore a hockey helmet” → “keeper wear the nhl hat”

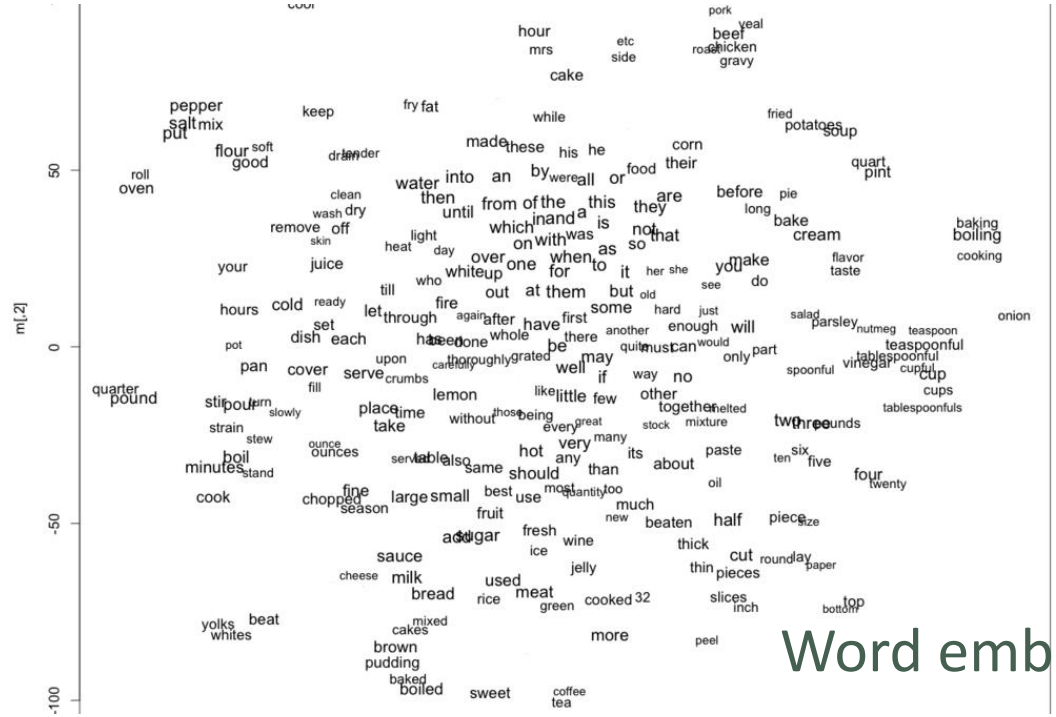
* Feyisetan, Drake, Diethe, Balle , “Privacy- and Utility-Preserving Textual Analysis via Calibrated Multivariate Perturbations”, WSDM 2020

Differential Private Text Perturbation

Task: Sanitize sensitive text

Perturb sentences while maintaining its semantic meaning

e.g., “goalie wore a hockey helmet” → “keeper wear the nhl hat”



Word embeddings

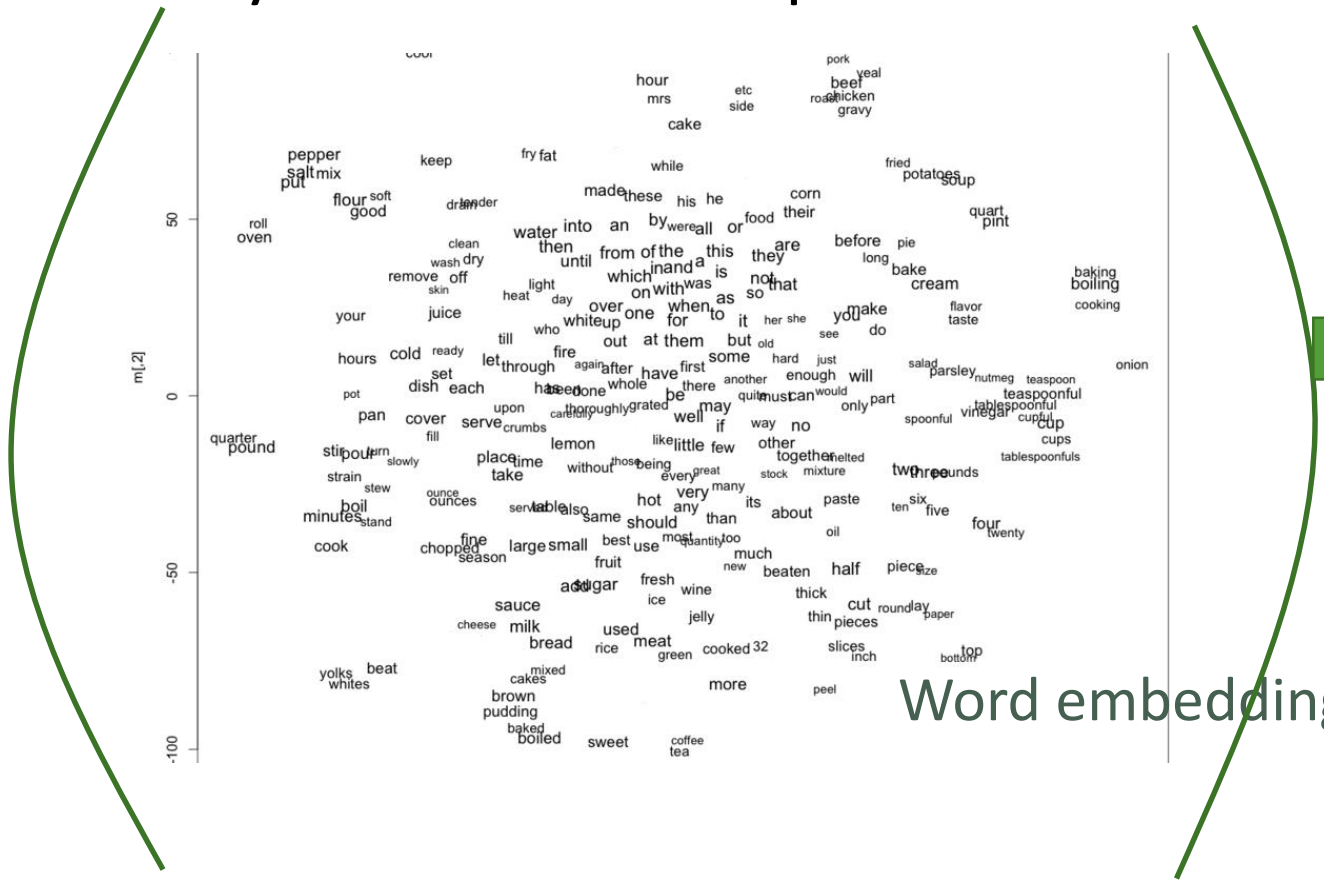
Differential Private Text Perturbation

Task: Sanitize sensitive text

Perturb sentences while maintaining its semantic meaning

e.g., “goalie wore a hockey helmet” → “keeper wear the nhl hat”

Apply Differential Privacy Transformations

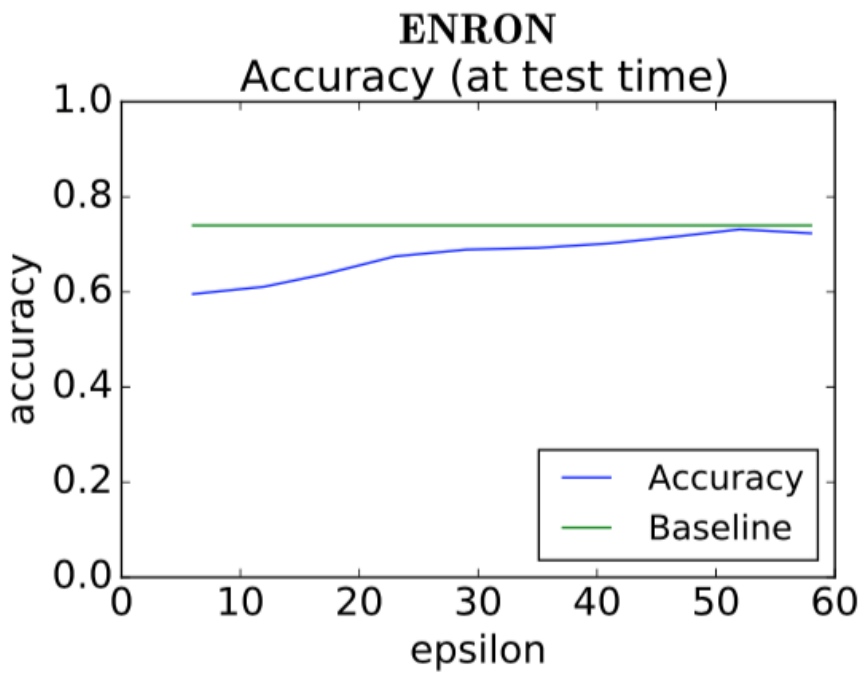


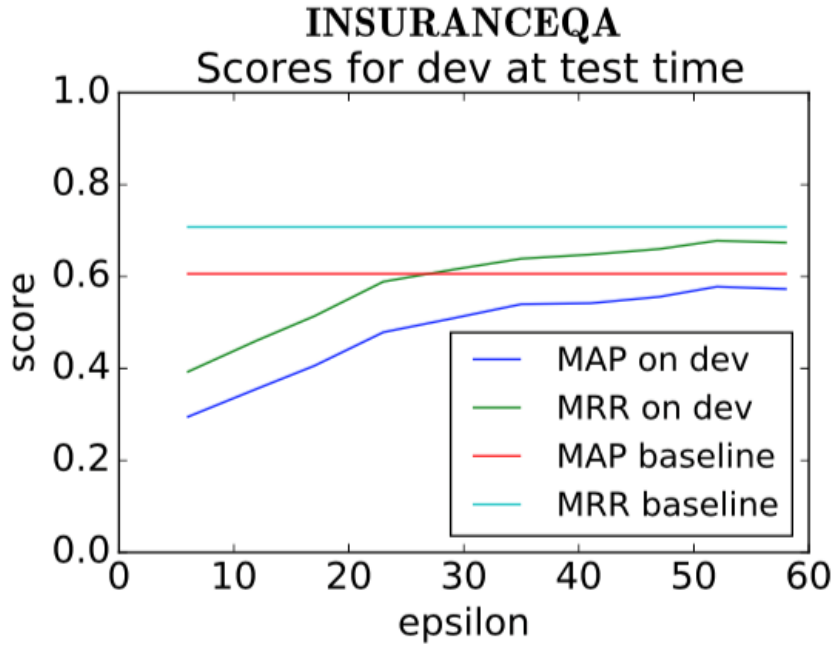
Word embeddings

Perturbed Sentences

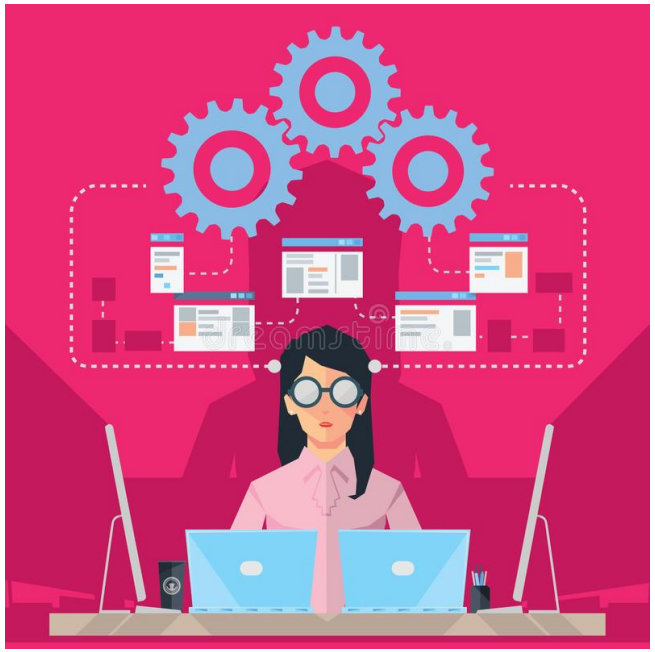
Differential Private Text Perturbation

Impact of accuracy given ϵ on multi-class classification and question answering tasks, respectively:





Code vs. Technical Paper



Theory vs. Practice:

Developers generally want code that they can start with, not technical papers

A General Purpose Differential Privacy Package

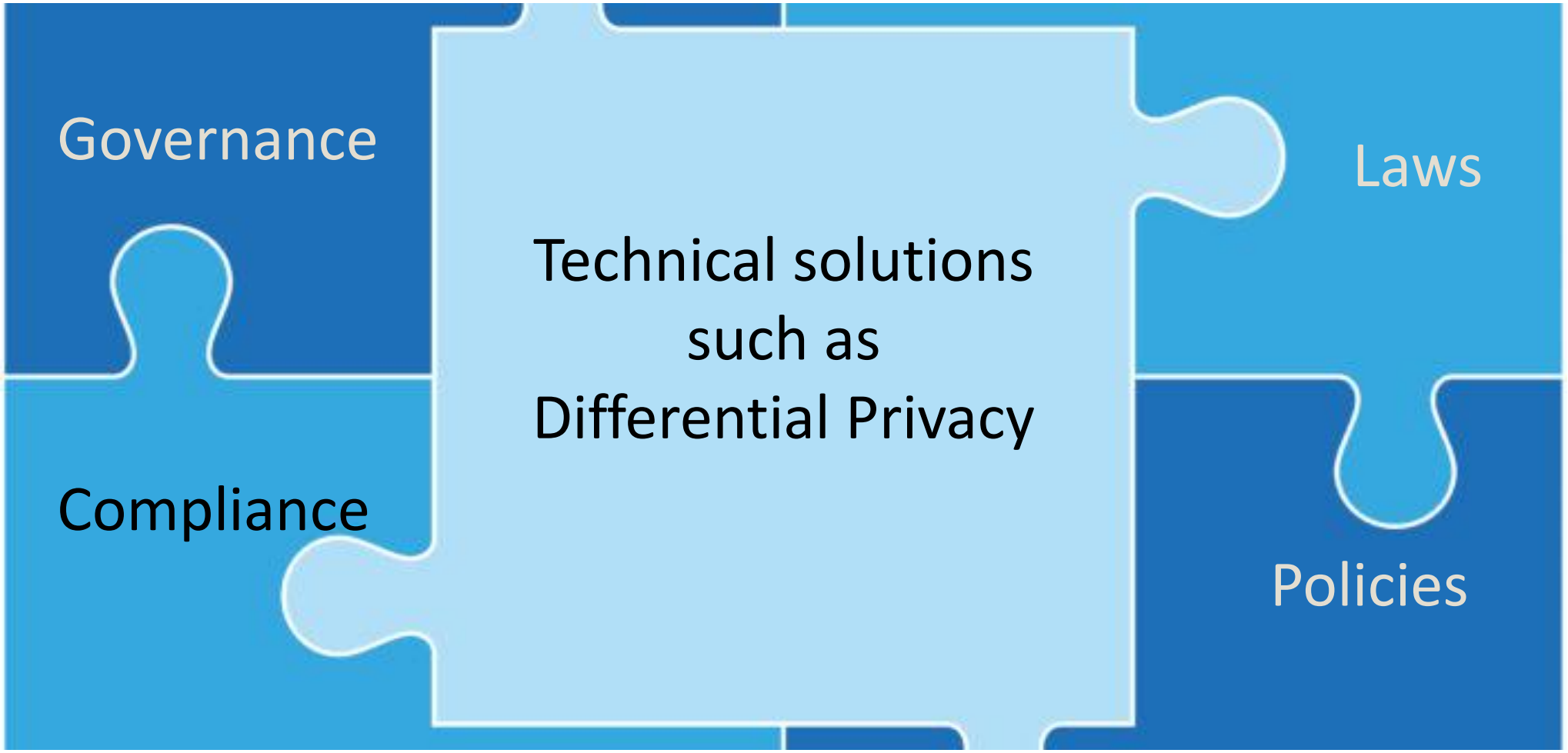
We are building tools to make differential privacy accessible to developers

Python Package:

Implements common differential privacy techniques

Available at: : https://github.com/yuxiangw/autodp

Based on: Wang, Balle, **K.**,Subsampled Renyi Differential Privacy and Analytical Moments Accountant, AISTATS 2019



Thank you for your attention!

Google

Data Privacy

NABE Tech Economics 2019

Paul Liu / Nov. 5, 2019

Time for questions and discussion