Eliciting Data Subjects' Privacy-Accuracy Preferences for Differentially-Private Deployments

Priyanka Nanayakkara
Harvard University

Jayshree Sarathy
Northeastern University

Mary Anne Smart
Purdue University

Rachel Cummings*
Columbia University

Gabriel Kaptchuk*
University of Maryland College Park

Elissa M. Redmiles*
Georgetown University

** equal advising*


differential privacy
Image: Unsplash

(Desfontaines 2023; Tezapsidis 2017; Triedman et al. 2023)

Differential Privacy (Dwork et al. 2006)

$$Pr\,[A(D) = o] \leq e^{\varepsilon}\, Pr\,[A(D') = o]$$

more noise
privacy

less noise
accuracy

Differential Privacy (Dwork et al. 2006)

$$Pr\,[A(D) = o] \leq e^{\varepsilon}\, Pr\,[A(D') = o]$$

more noise
privacy

less noise
accuracy

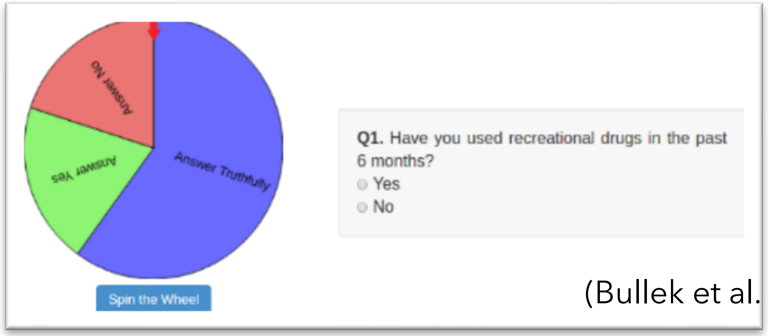
probabilistic **non-linear** **value-laden**

Prior Work on Explaining DP to Data Subjects

“To respect your personal information privacy and ensure best user experience, the data shared with the app will be processed via the differential privacy (DP) technique. That is, the app company will store your data but only use the aggregated statistics with modification so that your personal information cannot be learned. However, your personal information may be leaked if the company’s database is compromised.” (Xiong et al. 2020)

Text descriptions

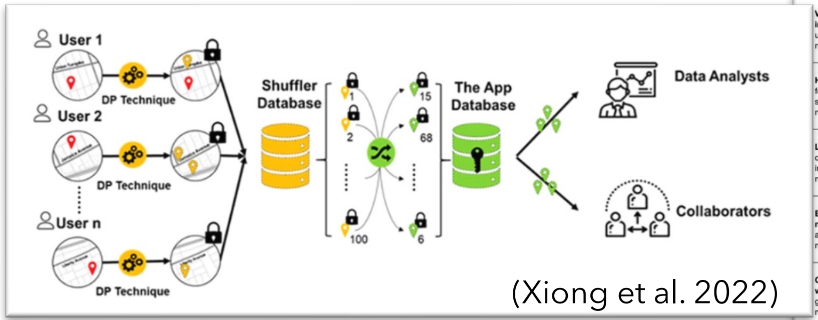
(e.g., Xiong et al. 2020, Cummings et al. 2021, Smart et al. 2023, Franzen et al. 2023)



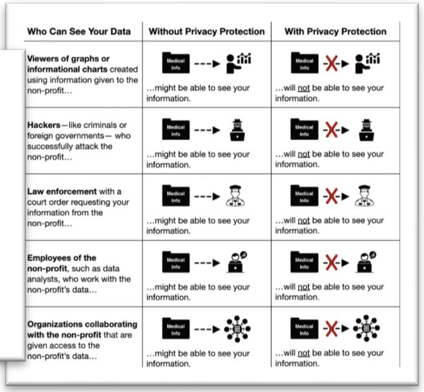
(Bullek et al. 2017)

Metaphors

(e.g., Bullek et al. 2017, Karegar et al. 2022)



(Xiong et al. 2022)

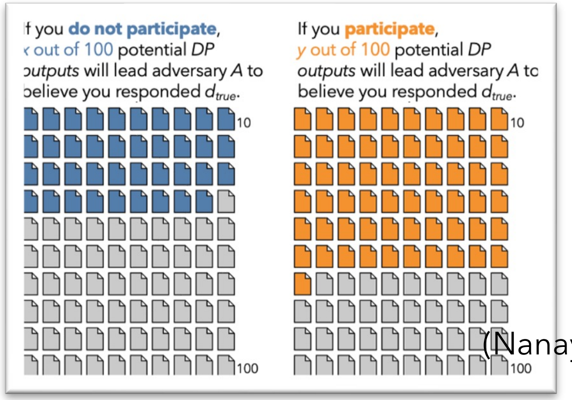


Who Can See Your Data	Without Privacy Protection	With Privacy Protection
Viewers of graphs or informational charts created using information given to the non-profit…	…might be able to see your information.	…will not be able to see your information.
Hackers—like criminals or foreign governments— who successfully attack the non-profit…	…might be able to see your information.	…will not be able to see your information.
Law enforcement with a court order requesting your information from the non-profit…	…might be able to see your information.	…will not be able to see your information.
Employees of the non-profit, such as data analysts, who work with the non-profit’s data…	…might be able to see your information.	…will not be able to see your information.
Organizations collaborating with the non-profit that are given access to the non-profit’s data…	…might be able to see your information.	…will not be able to see your information.

(Smart et al. 2024)

Diagrams & Tables

(e.g., Bullek et al. 2017, Karegar et al. 2022, Smart et al. 2024, Xiong et al. 2022, Wen et al. 2023)

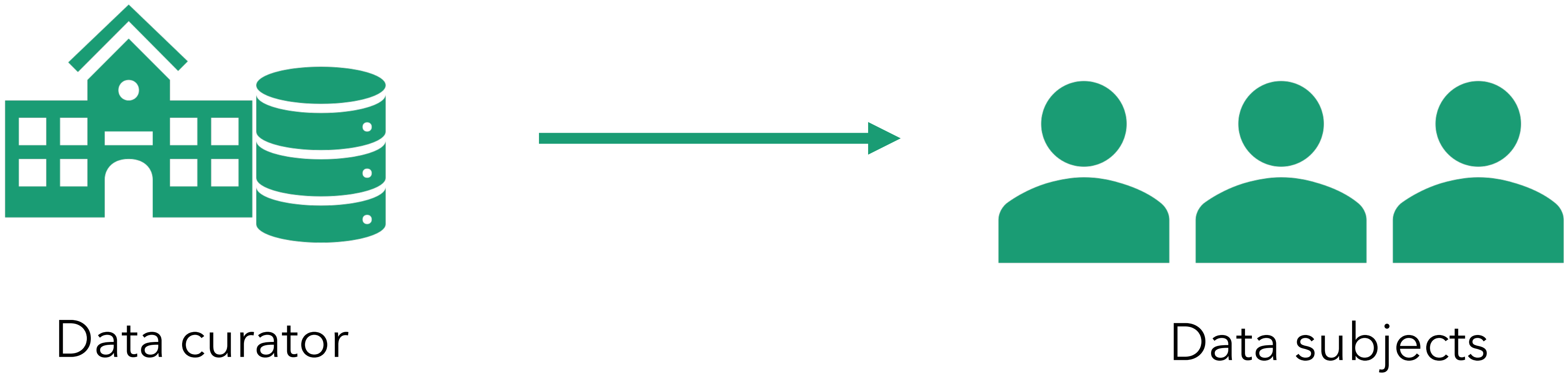


(Nanayakkara et al. 2023)

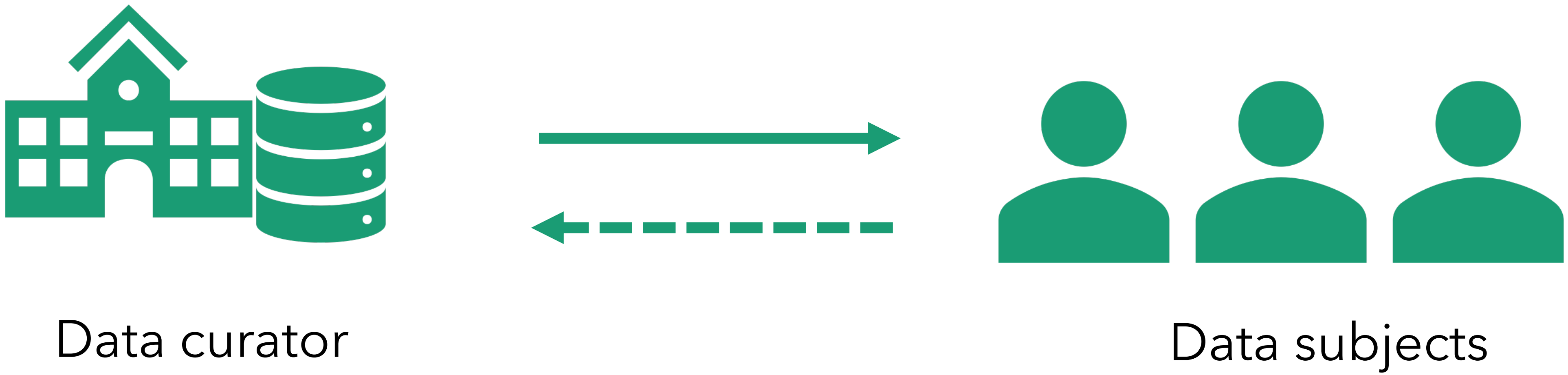
Visualizations

(e.g., Smart et al. 2023, Nanayakkara et al. 2023, Franzen et al. 2024, Ashena et al. 2024)

Prior Work on Explaining DP to Data Subjects



Prior Work on Explaining DP to Data Subjects



Prior Work on Explaining DP to Data Subjects

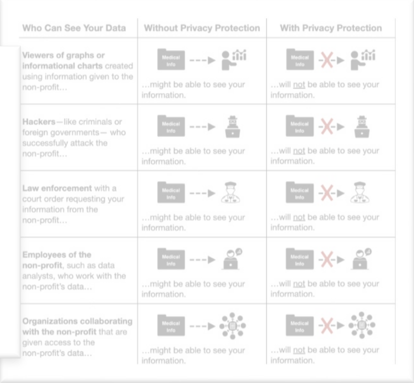
"To respect your personal information privacy and ensure best user experience, the data shared with the app will be processed via the differential privacy (DP) technique. That is, the app company will store your data but only use the aggregated statistics with modification so that your personal information cannot be learned. However, your personal information may be leaked if the company's database is compromised." (Xiong et al. 2020)

Text descriptions

(e.g., Xiong et al. 2020, Cummings et al. 2021, Smart et al. 2023, Franzen et al. 2023)

Metaphors

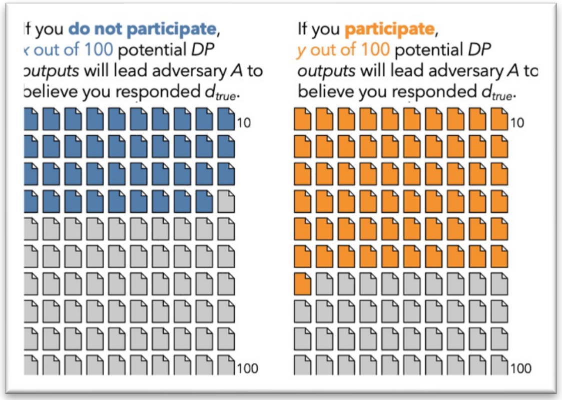
(e.g., Bullek et al. 2017, Karegar et al. 2022)



Who Can See Your Data	Without Privacy Protection	With Privacy Protection
Viewers of graphs or informational charts created using information given to the non-profit…	Medical Info …might be able to see your information.	Medical Info …will not be able to see your information.
Hackers—like criminals or foreign governments— who successfully attack the non-profit…	Medical Info …might be able to see your information.	Medical Info …will not be able to see your information.
Law enforcement with a court order requesting your information from the non-profit…	Medical Info …might be able to see your information.	Medical Info …will not be able to see your information.
Employees of the non-profit, such as data analysts, who work with the non-profit's data…	Medical Info …might be able to see your information.	Medical Info …will not be able to see your information.
Organizations collaborating with the non-profit that are given access to the non-profit's data…	Medical Info …might be able to see your information.	Medical Info …will not be able to see your information.

Diagrams & Tables

(e.g., Bullek et al. 2017, Karegar et al. 2022, Smart et al. 2023, Xiong et al. 2023, Wen et al. 2023)



Visualizations

(e.g., Smart et al. 2023, Nanayakkara et al. 2023, Franzen et al. 2024, Ashena et al. 2024)

If you **do not share data**,
x out of 100 potential *DP outputs* will lead adversary A to believe you responded d_{true}.

If you **share data**,
y out of 100 potential *DP outputs* will lead adversary A to believe you responded d_{true}.

(Nanayakkara et al. 2023)

Probabilities reflect immediate decisions

If you **do not share data**,
x out of 100 potential *DP outputs* will lead adversary A to believe you responded d_{true}.

If you **share data**,
y out of 100 potential *DP outputs* will lead adversary A to believe you responded d_{true}.

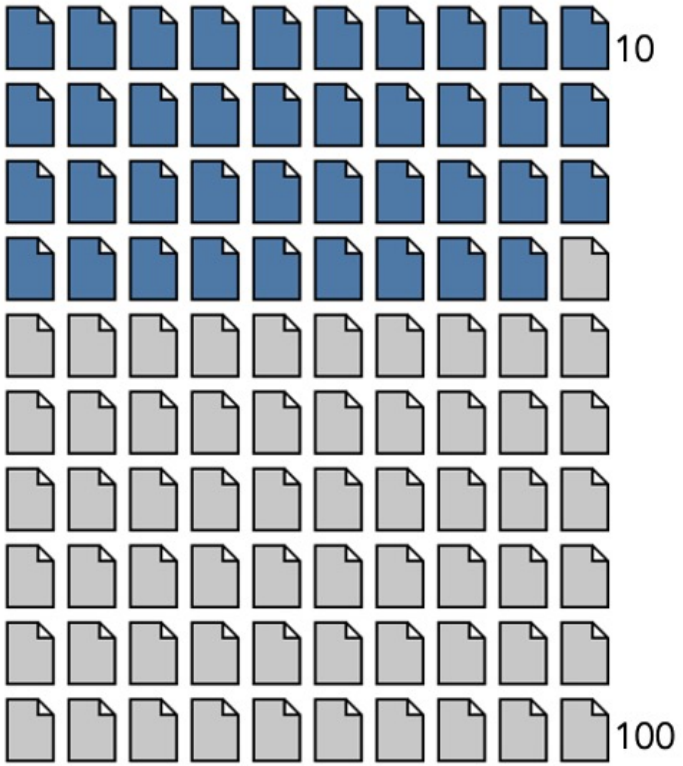
Framing probabilities as frequencies vs. percentages supports statistical reasoning & has been applied in privacy contexts

If you **do not share data**,
x out of 100 potential *DP outputs* will lead adversary *A* to believe you responded d_{true}.

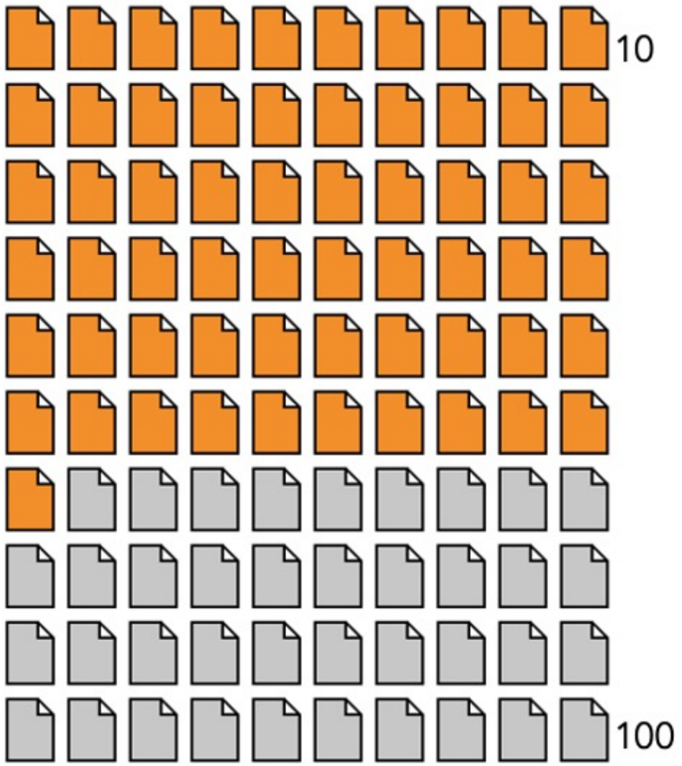
If you **share data**,
y out of 100 potential *DP outputs* will lead adversary *A* to believe you responded d_{true}.

(Gigerenzer and Hoffrage 1995, Hoffrage and Gigerenzer 1998, Slovic 2000, Kaptchuk et al. 2020, Franzen et al. 2022)

If you **do not share data**,
x out of 100 potential *DP outputs* will lead adversary *A* to believe you responded d_{true}.

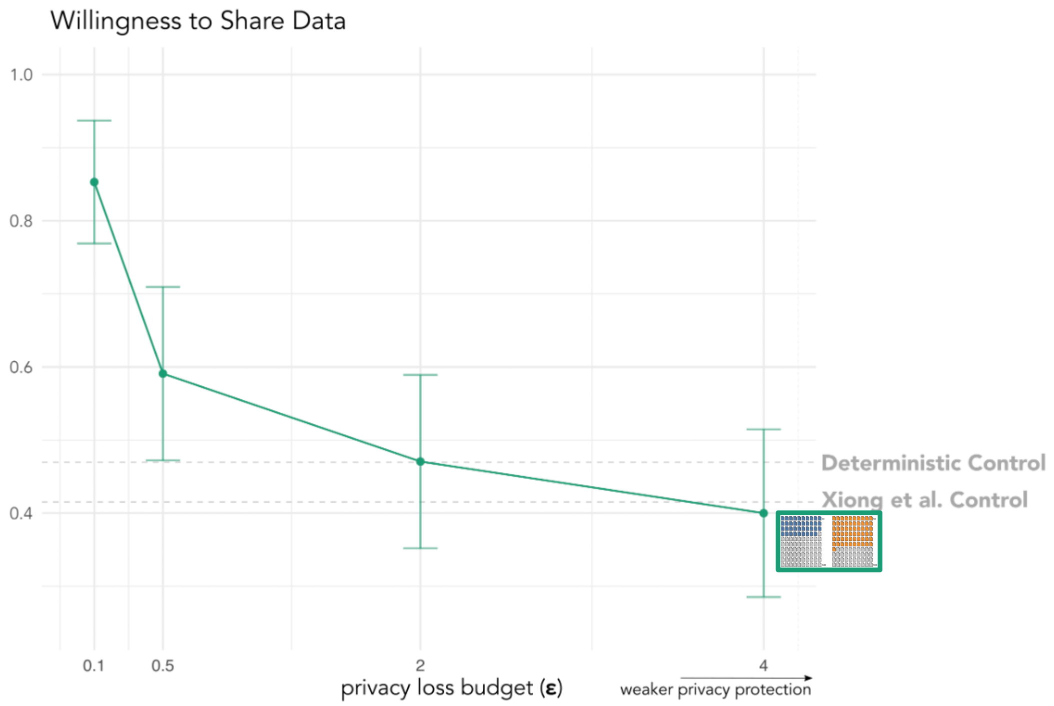


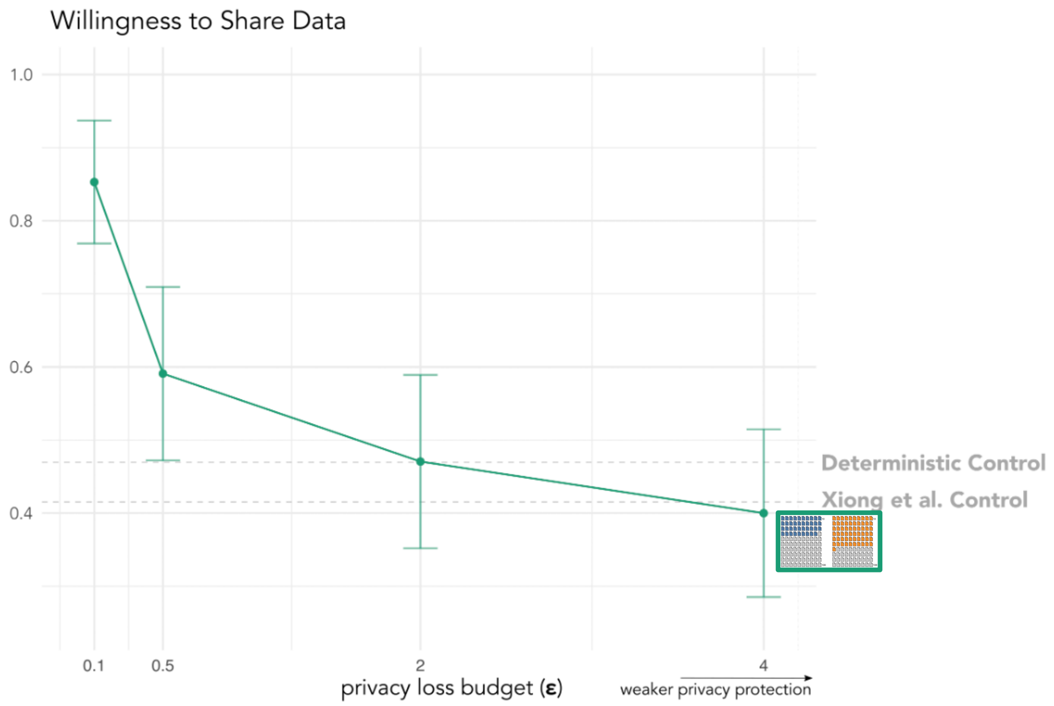
If you **share data**,
y out of 100 potential *DP outputs* will lead adversary *A* to believe you responded d_{true}.



Icon arrays assume x = 39 and y = 61 for illustration purposes.

(Galesic et al. 2009)

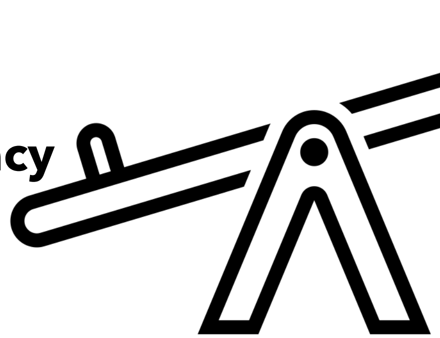

Willingness to Share Data
1.0
0.8
0.6
0.4
0.1
0.5
2
4
privacy loss budget (ε)
weaker privacy protection
Deterministic Control
Xiong et al. Control



> *The random process completely obfuscates the true [data]; that is great for [data subject] anonymity, but is kind of useless for the [data curator].*

CURRENT WORK

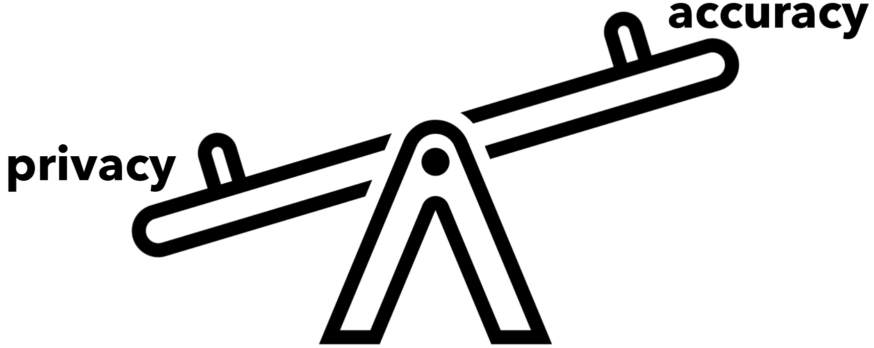
Elicit preferences along the privacy—accuracy tradeoff



CURRENT WORK

Elicit preferences along the privacy—accuracy tradeoff

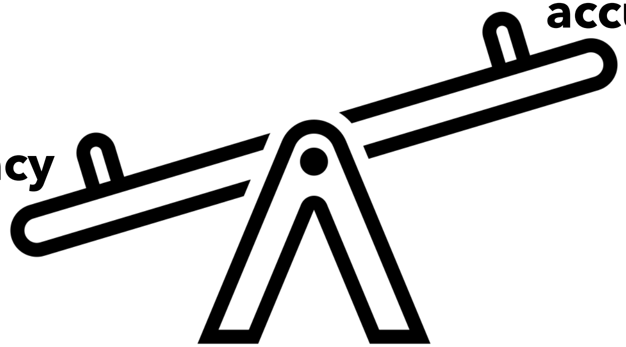
(in a machine learning context)



CURRENT WORK

Elicit preferences along the privacy–accuracy tradeoff

(in a machine learning context)



Step 2: Use these explanations to learn preferences at scale

Step 1: Develop explanations of epsilon's privacy and accuracy implications

Step 2: Use these explanations to learn preferences at scale

Step 1: Develop explanations of epsilon’s privacy and accuracy implications

Privacy

How confident is the adversary in claiming your information was used?

Q1: Does this seem like the right question and why?

Privacy

The data curator will apply privacy protection when creating the program. The privacy protection will limit the adversary's confidence when claiming that your data were used. In particular, they will be at most **X% confident that your data were used**.

What does this mean? Suppose the privacy protection were applied 100 separate times and each time, the adversary were to claim that your data were used. They would be **correct for at most X out of 100 claims**.

Privacy

The data curator will apply privacy protection when creating the program. The privacy protection will limit the adversary's confidence when claiming that your data were used. In particular, they will be at most **X% confident that your data were used**.

Bound on membership inference (Thudi et al. 2022)

What does this mean? Suppose the privacy protection were applied 100 separate times and each time, the adversary were to claim that your data were used. They would be **correct for at most X out of 100 claims**.

Q2: Does this explanation make sense to you as a {data subject, computer scientist, policymaker}?

Privacy

Q3: Is a baseline probability in the no-DP setting helpful?

The data curator will apply privacy protection when creating the program. The privacy protection will limit the adversary's confidence when claiming that your data were used. In particular, they will be at most **X% confident that your data were used**.

What does this mean? Suppose the privacy protection were applied 100 separate times and each time, the adversary were to claim that your data were used. They would be **correct for at most X out of 100 claims**.

Accuracy

How often will the program make correct predictions?

Q4: Does this seem like the right question and why?

Accuracy

We expect the program without privacy protection to make correct predictions for every y out of 100 people. With privacy protection, we expect the program to make correct predictions for at least **z out of 100 people**.

Accuracy

We expect the program without privacy protection to make correct predictions for every y out of 100 people. With privacy protection, we expect the program to make correct predictions for at least **z out of 100 people**.

Bound on accuracy
(Mangold et al. 2023)

Step 2: Use these explanations to learn preferences at scale

Step 1: Develop explanations of epsilon's privacy and accuracy implications

Conjoint Analysis Applied to DP Elicitation

Conjoint analysis:

Method commonly used in marketing research

Quantifies how people weight different attributes of a tested product

Output: attributes' relative importance (%), computed through hierarchical Bayesian modeling

Slide adapted from Elissa M. Redmiles, USENIX Sec 23 presentation

(Arning, 2017; Cattin & Wittink, 1982; Ayalon et al. 2023)

Conjoint Analysis Applied to DP Elicitation

Conjoint analysis:

Method commonly used in marketing research

Quantifies how people weight different attributes of a tested product

Output: attributes' relative importance (%), computed through hierarchical Bayesian modeling

Why is this a good choice for the DP setting?

Slide adapted from Elissa M. Redmiles, USENIX Sec 23 presentation

(Arning, 2017; Cattin & Wittink, 1982; Ayalon et al. 2023)

Conjoint Analysis Applied to DP Elicitation

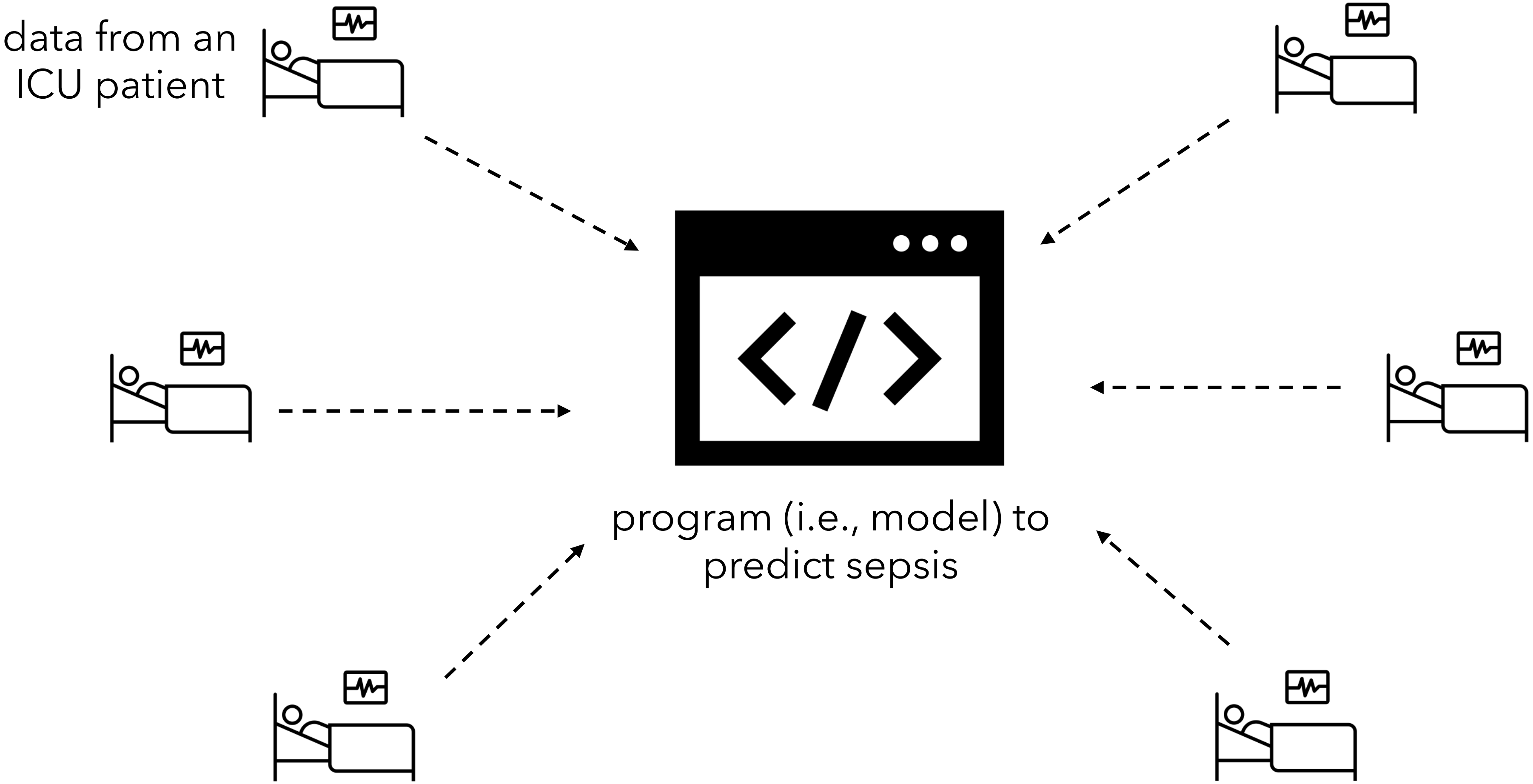
1. DATA COLLECTION SCENARIO presented to data subjects

2. PREFERENCES BETWEEN PROGRAMS COLLECTED among data subjects

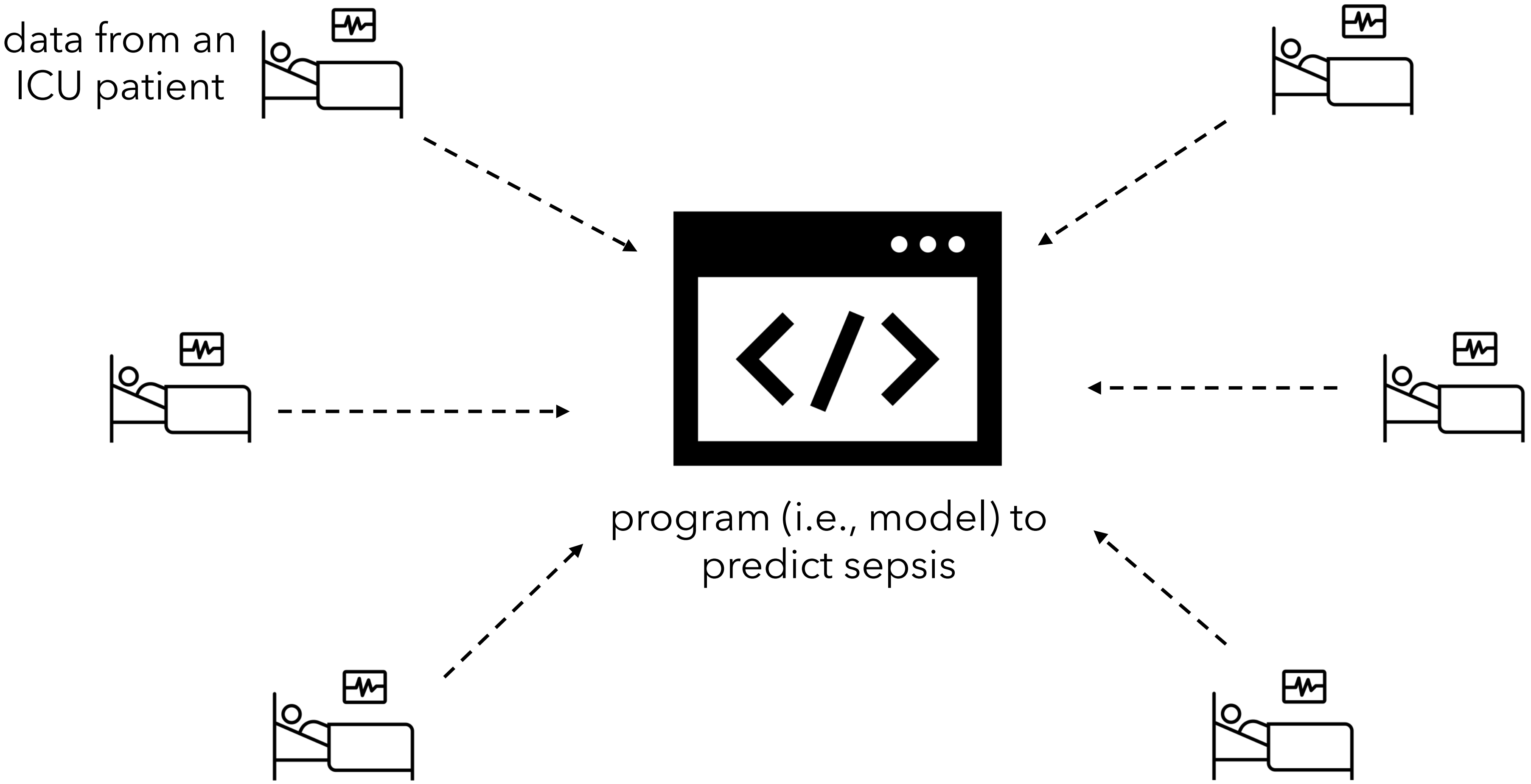
3. RESPONSES ANALYZED to learn relative importance of privacy, accuracy, etc.

our privacy and accuracy explanations

Scenario

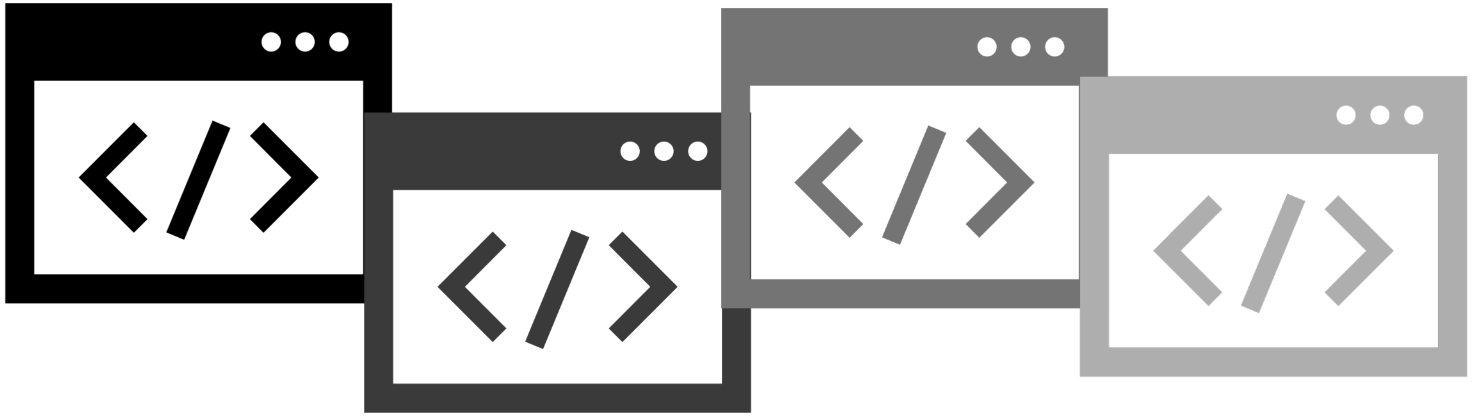


Scenario



sensitive attribute = admission to the ICU

Scenario



Scenario

The hospital wants your opinion about how to implement the program.

In the following screens, you will make a series of choices about which approach you think the hospital should use.

		Option 1	Option 2
privacy	The hospital will apply privacy protection when creating the program. The privacy protection will limit the employer's confidence when claiming that your data were used. In particular, they will be AT MOST…	**X_1% confident that your data were used** *What does this mean?* Suppose the privacy protection were applied 100 times and each time, your employer were to claim that your data were used. They would be correct for X_1 out of 100 claims.	**X_2% confident that your data were used** *What does this mean?* Suppose the privacy protection were applied 100 times and each time, your employer were to claim that your data were used. They would be correct for X_2 out of 100 claims.
accuracy	We expect the program without privacy protection to make correct predictions for every Y out of 100 people. With privacy protection, we expect the program to make correct predictions for AT LEAST…	**Z_1 out of 100 people**	**Z_2 out of 100 people**

Option 3: I prefer that my patient records are not used.

	Option 1	Option 2
privacy		
accuracy		
how collected preferences will be used		
how long the program will be in use		
availability of the program		

Q5: What other factors are we missing, if any?

Option 3: I prefer that my patient records are not used.

Our Plan

1. REFINE SCENARIO TEXT & EXPLANATIONS
   Perform cognitive interviews

2. RUN STUDY ON PROLIFIC IN THE NEXT FEW MONTHS

Our Plan

1. REFINE SCENARIO TEXT & EXPLANATIONS
   Perform cognitive interviews
2. RUN STUDY ON PROLIFIC IN THE NEXT FEW MONTHS

---

3. TEST OUR METHOD IN PRACTICE

Questions

In this work, our goal is to elicit preferences from data subjects. Are there other parties who we should also (or instead) be trying to elicit preferences from?

Would elicitation strategies for policymakers look different? How so?

What might be some real-world challenges to deploying a methodology like ours in practice?

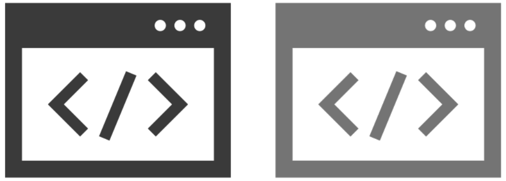
Imagining yourself as a data subject, what other information would you want when making choices between programs in our study?

Thank you!

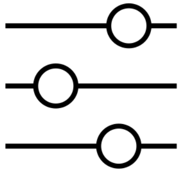
Priyanka Nanayakkara (priyankan@g.harvard.edu | @priyakalot | @priyakalot@hci.social),
Jayshree Sarathy, Mary Anne Smart, Rachel Cummings, Gabriel Kaptchuk, Elissa M. Redmiles

1

2



3



In this work, our goal is to elicit preferences from data subjects. Are there other parties who we should also (or instead) be trying to elicit preferences from?

Would elicitation strategies for policymakers look different? How so?

What might be some real-world challenges to deploying a methodology like ours in practice?

Imagining yourself as a data subject, what other information would you want when making choices between programs in our study?