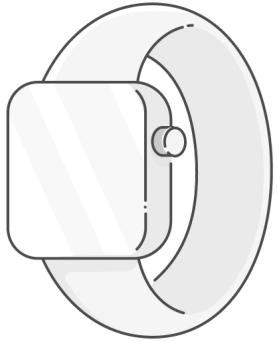
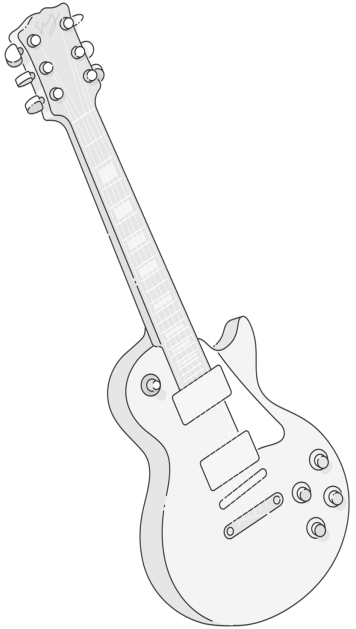
Lemonade

# Data Privacy in Practice

Scott Fischer
December 12, 2022





# One of these things is not like the other

# Open Questions in the Insurance Context

- **What is the best way to handle tradeoffs**
  - For fraud detection, for example, we want to ensure that users can't *manipulate* data. How much transparency can we actually provide?
- **What level of control can users have?**
  - In the traditional consumer context, users are allowed to delete their data for a potentially less personalized experience. In the car insurance context, for example, if we allowed users to delete data, we could end up with a skewed perception of risk. How do we balance this approach?
- **What are the best ways to provide users with context?**
  - Are Terms of Service or Privacy Policies the best way to achieve this? What should or could disclosure look like?

# How should this UNIQUE industry think about privacy?

## Intentional collection & use of a user's data

- What data are we collecting from users and why?
- Who has access to that data, and what can they use it for?
- What do users know about their data, and what control do they have?

# How do insurers collect data?

## Data directly from users

- Ideally sensitive & personal data should come from the user

## Data from 3rd party sources

- Data that relates to user behavior in other domains (e.g. house purchase)

## Data indirectly provided

- User behavior (e.g. scrolling through the product)

## Data generated within the company

- Data about a user that is generated, for example via machine learning (e.g. detection of whether a user has a swimming pool)

# User data ranges in its sensitivity

Less Sensitive

More Sensitive

Not user identifiable
*E.g. Scrolling patterns*

User is identifiable and data is personal
*E.g. Video footage*

Data is considered SPII (significant harm if compromised)
*E.g. SSN, Biometrics*

Less Sensitive

More Sensitive

**Not user identifiable**
*E.g. Scrolling patterns*

**User is identifiable and data is personal**
*E.g. Video footage*

**Data is considered SPII (significant harm if compromised)**
*E.g. SSN, Biometrics*

And depending on how sensitive the data is, we should consider:

**How it is collected**

**What user controls exist**

**Who can access the data internally & externally**

**Less Sensitive** → **More Sensitive**

| Not user identifiable | User is identifiable and data is personal | Data is considered SPII (significant harm if compromised) |
|---|---|---|
| *E.g. Scrolling patterns* | *E.g. Video footage* | *E.g. SSN, Biometrics* |

**The more sensitive the data, the more guarded access should be…**

| | | |
|---|---|---|
| Can provide more broad access and use across use cases & features | Limited access for purposes that are pre-defined for users | Extremely limited access internally, and should be used only for a handful of purposes where no other data can be used |

**Less Sensitive** → **More Sensitive**

| Not user identifiable | User is identifiable and data is personal | Data is considered SPII (significant harm if compromised) |
|---|---|---|
| *E.g. Scrolling patterns* | *E.g. Video footage* | *E.g. SSN, Biometrics* |

## …and the more control and awareness users should have

| | | |
|---|---|---|
| Can be indirectly collected and stored without explicit consent | Provide users appropriate levels of awareness and control depending on the use case | Explicitly seek user permission to collect or generate data<br><br>Allow users controls to be able to access and delete data |

# How do insurers collect data?

## Data from 3rd party sources

- Data that relates to user behavior in other domains (e.g. house purchase)
- Need to hold 3rd party accountable for clarity during collection

## Data directly from users

- Ideally sensitive & personal data should come from the user
- Purpose should be clear to users through the user experience

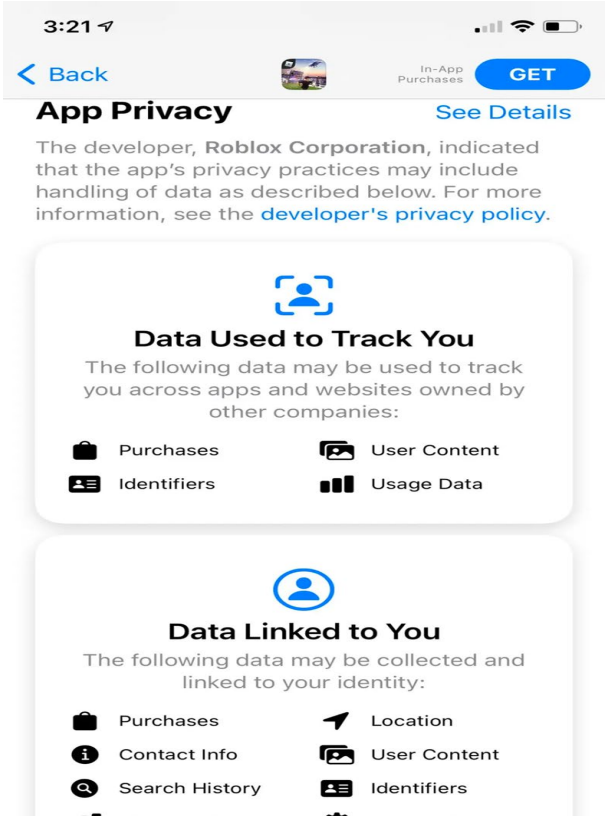
## Data indirectly provided

- User behavior (e.g. scrolling through the product)
- Indirect behavioral data should not contain any sensitive information such as PII

## Data generated within the company

- Data about a user that is generated, for example via machine learning (e.g. detection of whether a user has a swimming pool)
- When users provide information for generation, should be informed of how data may be used, especially for more sensitive cases

# What do users know & control do they have?

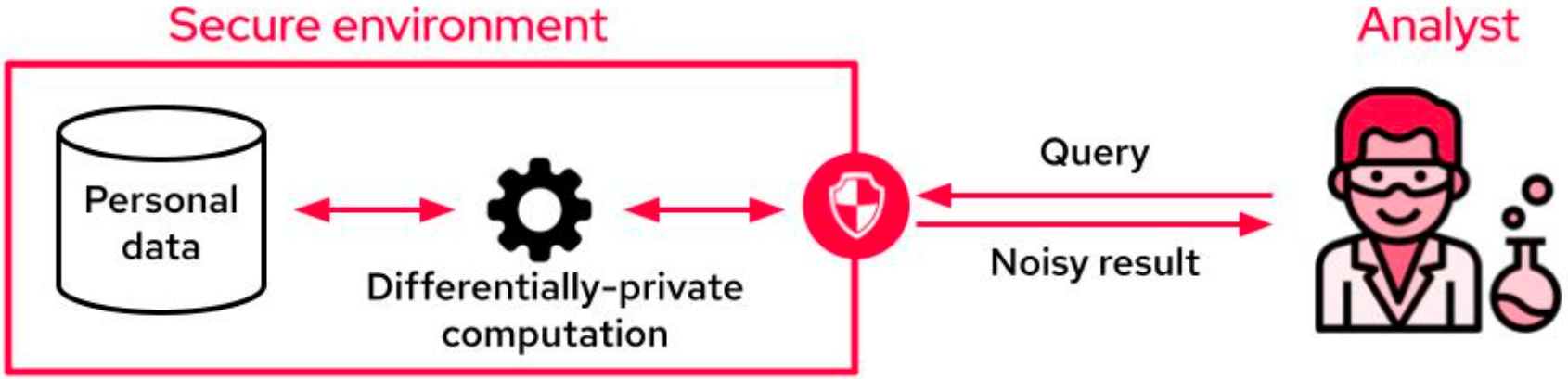



- Provide users context **while** the data is being collected
- Provide a common help center or space for users to understand **what** data is being collected, and for **what purposes**
- Account deletion == Data deletion
- Identify what data makes sense for a user to have control over, in terms of deletion, access, and usage

# Technical innovations in the space

Techniques like **Differential Privacy** and **Federated Learning** are making it possible to add noise to or silos a user's data, while still providing access to train models

*not being used in Insurance yet – are these options?



Lemonade

# Thanks!