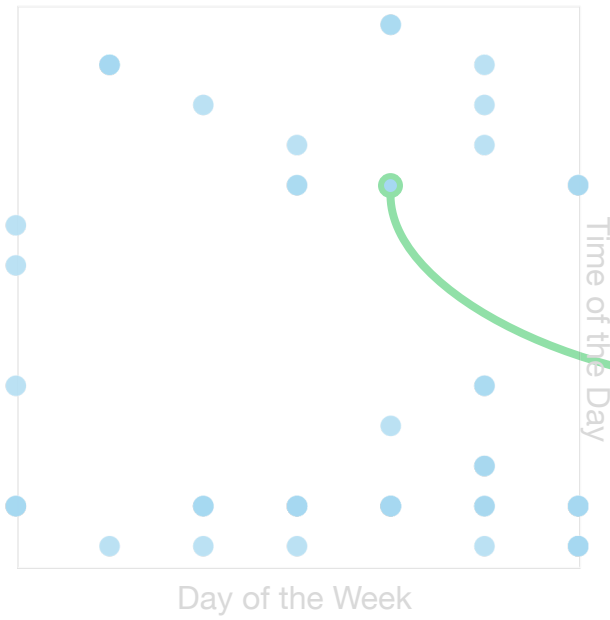


What is Sensitive About (Sensitive) Data?

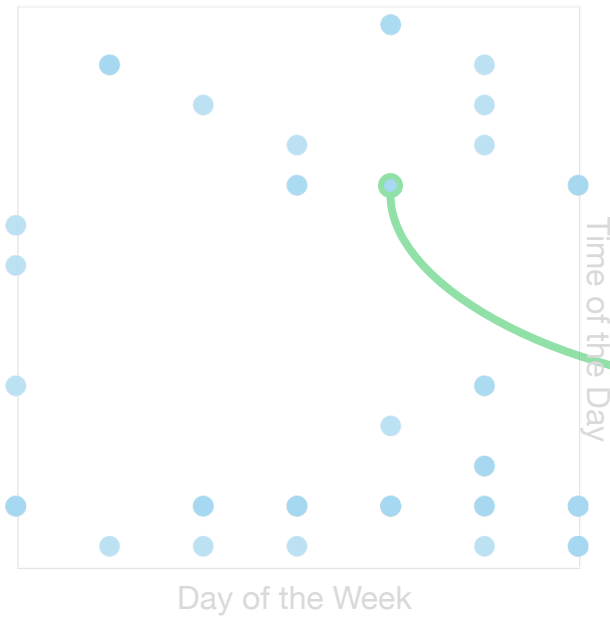
Characterizing Sensitivity and Intimacy with Google Assistant Users

Alejandra Gómez Ortega, Jacky Bourgeois, Gerd Kortuem



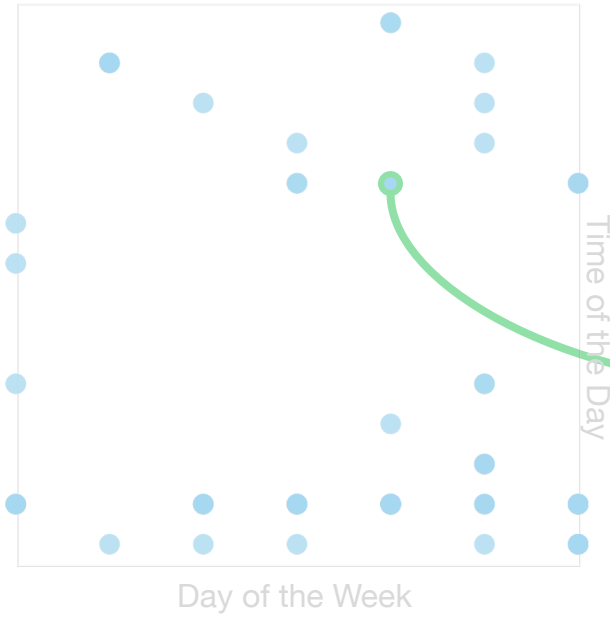


OK Google, set an alarm



Speech Record

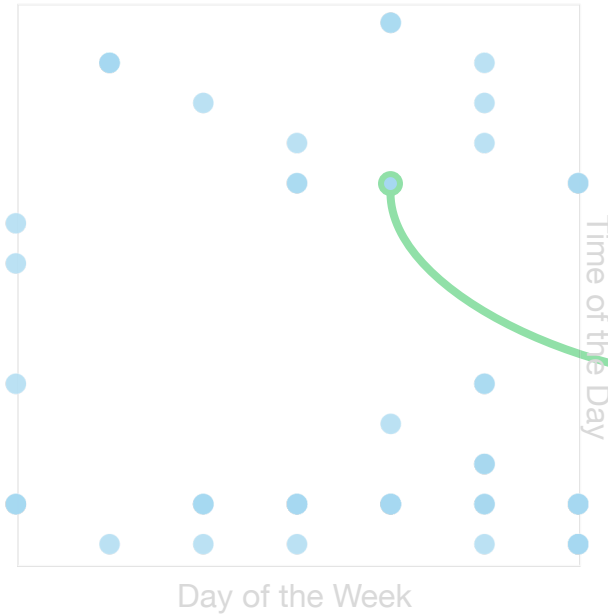
- *Timestamp*
- *Transcript*
- *Audio Recording*



Speech Record

Personal Data¹: Data through which a person can be directly or indirectly identified.

¹European General Data Protection Regulation (GDPR)



Speech Record

Sensitive Data¹: Special category of personal data.

Intimate Data²: Data from intimate contexts and activities.

¹European General Data Protection Regulation (GDPR)

²Kwon et al., 2018; Almeida et al., 2016

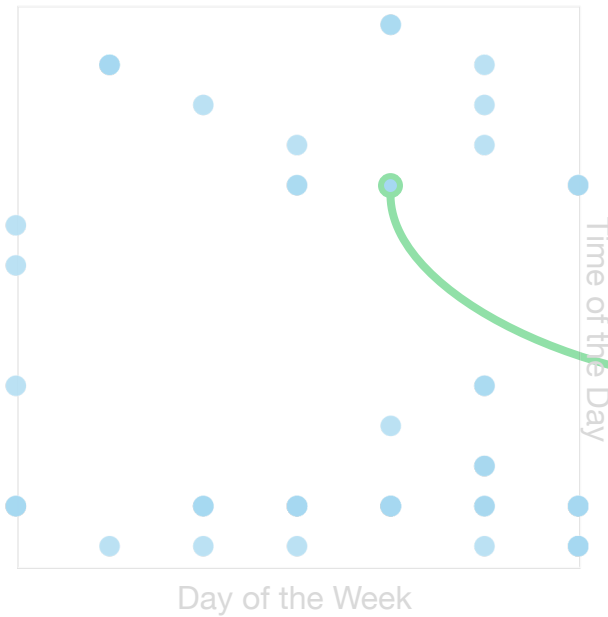


Research Question

How do voice assistant users perceive **sensitivity** and **intimacy** when faced with **a comprehensive overview** of their speech records?

Speech Records: Sensitive or Intimate Data?

“On the whole, data currently stored with voice assistants is not considered sensitive”¹

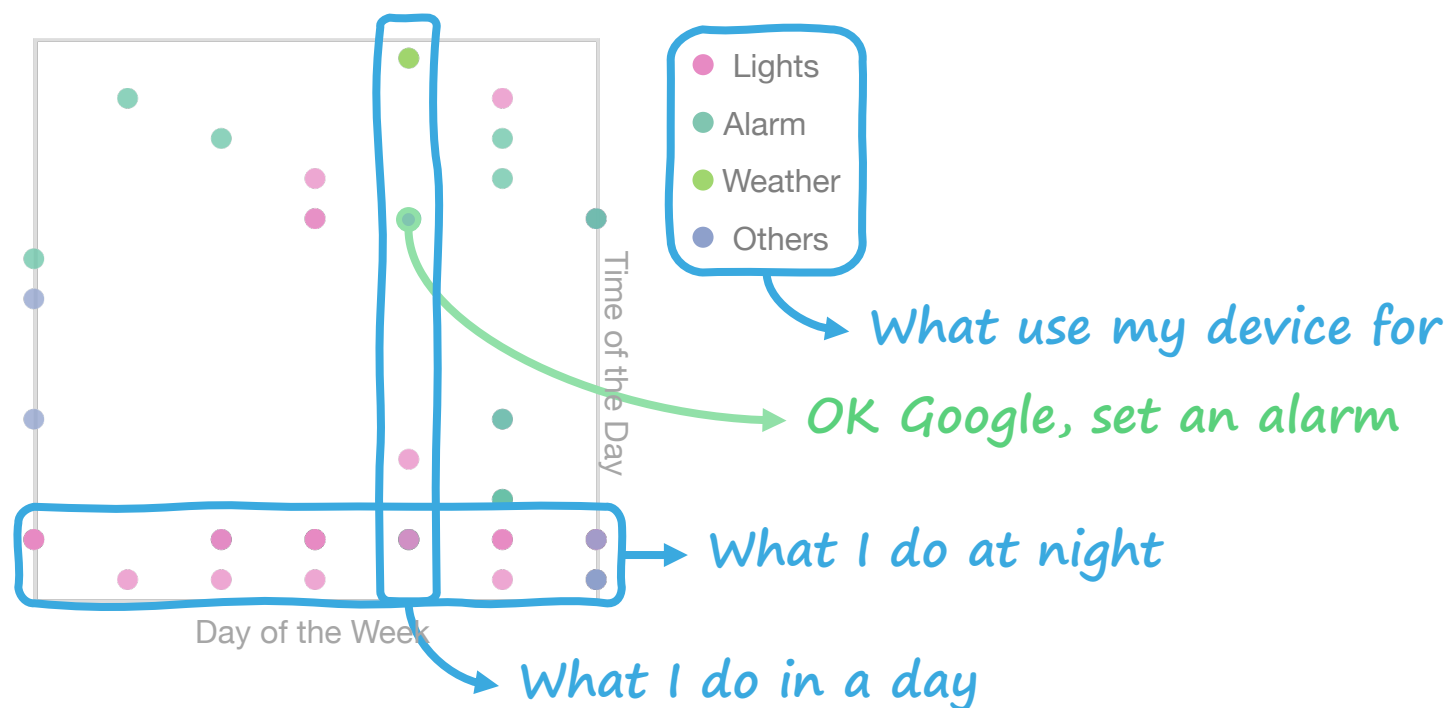


OK Google, set an alarm

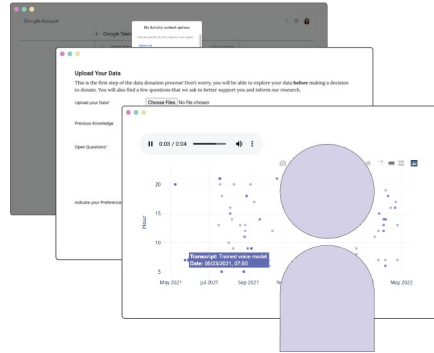
¹Malkin et al., 2019



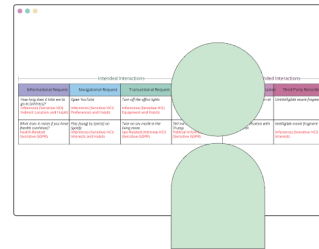
Speech Records: Sensitive or Intimate Data?



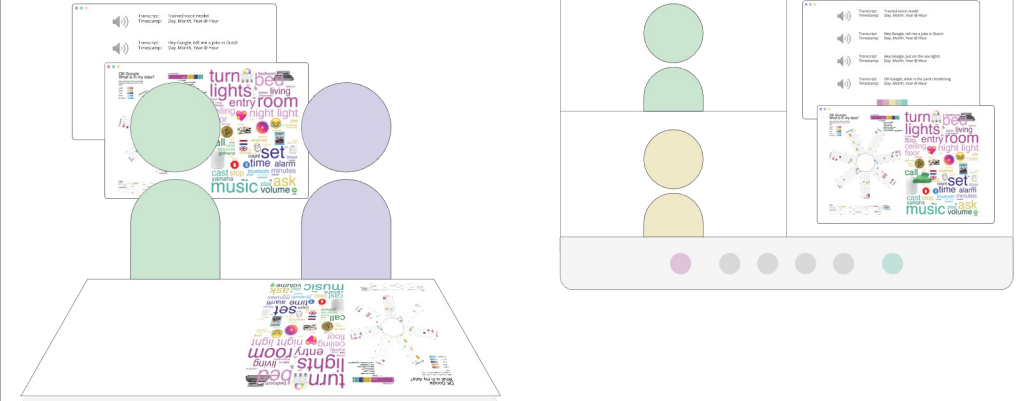
Methodology



Collection of Speech Records
through Data Donation



Familiarization and Classification
of Speech Records



Interviews

The Personal Data Canvas

Interactive data representation used during the interviews.



Transcript: Trained voice model
Timestamp: Day, Month, Year @ Hour
(Intended: training)



Transcript: Hey Google, tell me a joke in Dutch
Timestamp: Day, Month, Year @ Hour
(Intended: playful)

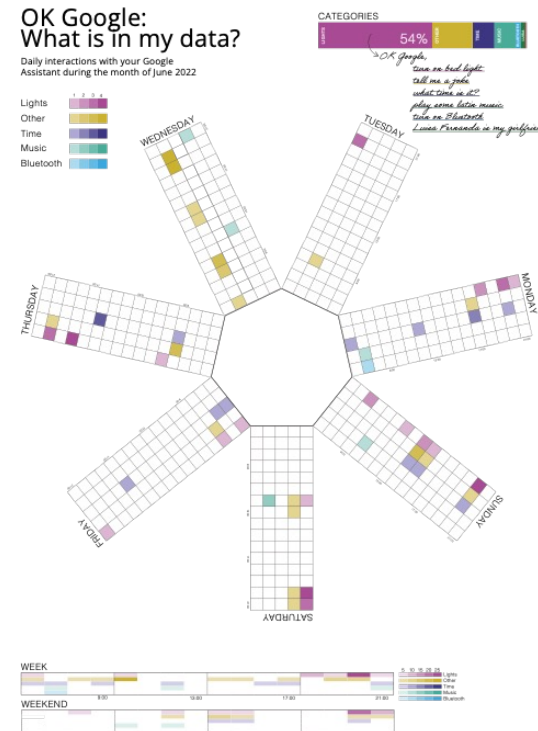


Transcript: Hey Google, put on the sex lights
Timestamp: Day, Month, Year @ Hour
(Intended: common request)



Transcript: OK Google, what is the yard clendening
Timestamp: Day, Month, Year @ Hour
(unintended: background conversation)

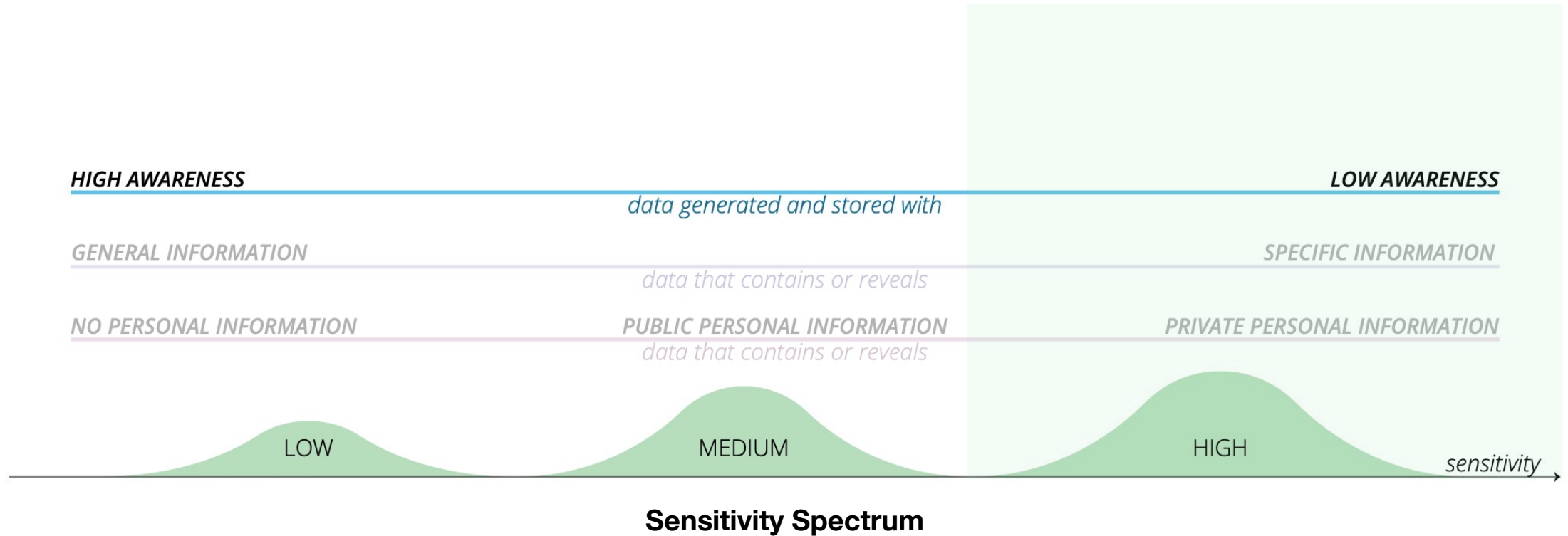
Acceptability and Sensitivity Score



Results

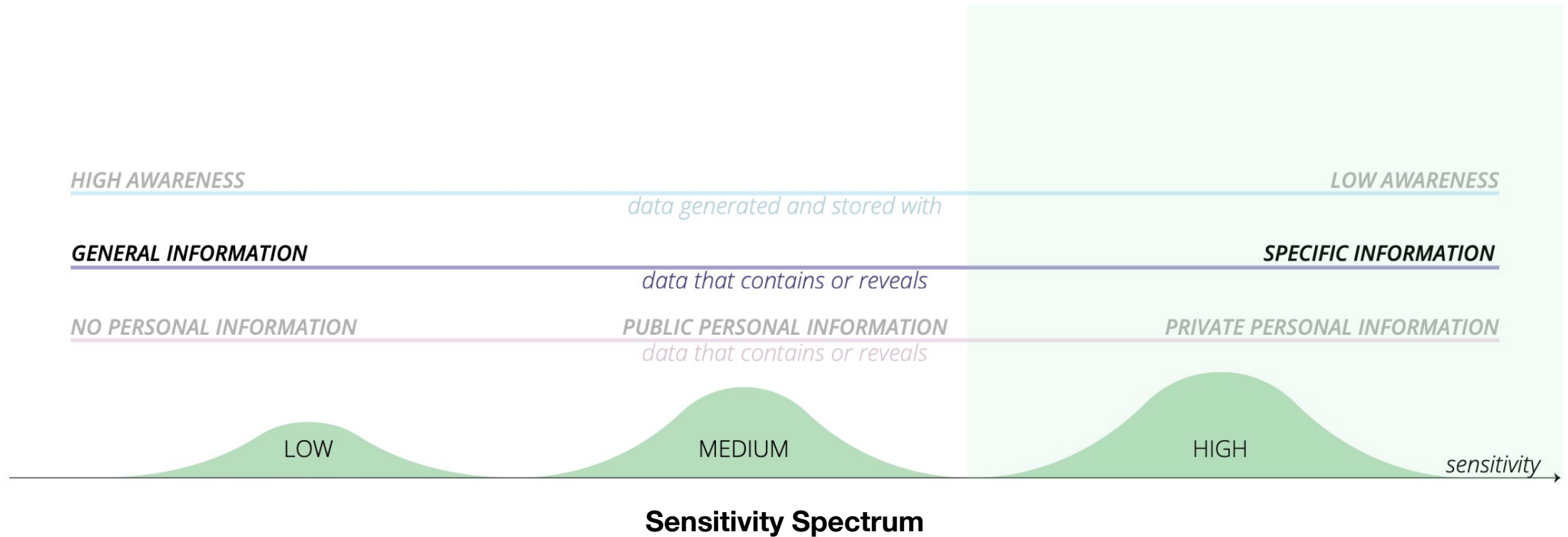
Results: Sensitive Speech Records

Intrusive, specific, and (un)available.



Results: Sensitive Speech Records

Intrusive, **specific**, and (un)available.



Intrusive, specific, and (un)available.



Results: Intimate Speech Records

Close physical proximity through which intimate moments or activities are captured.

“We were listening to recordings of when **we asked the alarm to turn off**, and you can tell there is like **a couple in bed, very sleepy**, you know?

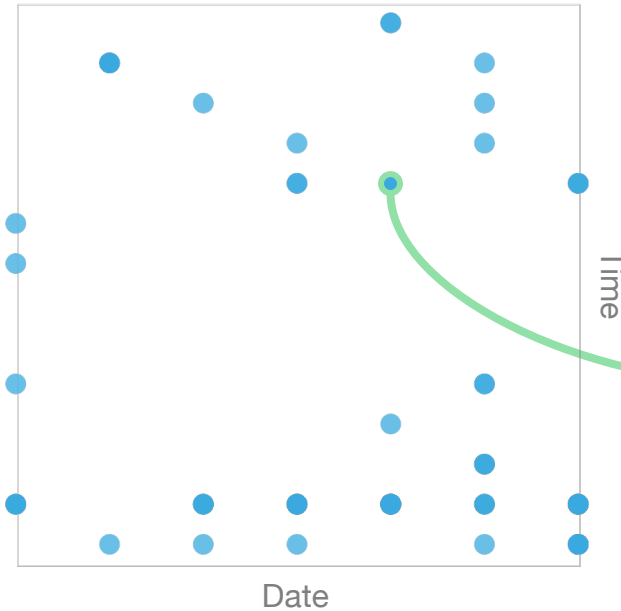
And it feels like **it was in bed with us**, kind of, and so it's not the **public voice** I have when I'm online or with my friends.

It's like a space that is **really, really, intimate**, in the sense that **no one else is usually allowed.**”¹

¹Quote from D9_a



Speech Records: Sensitive or Intimate Data?

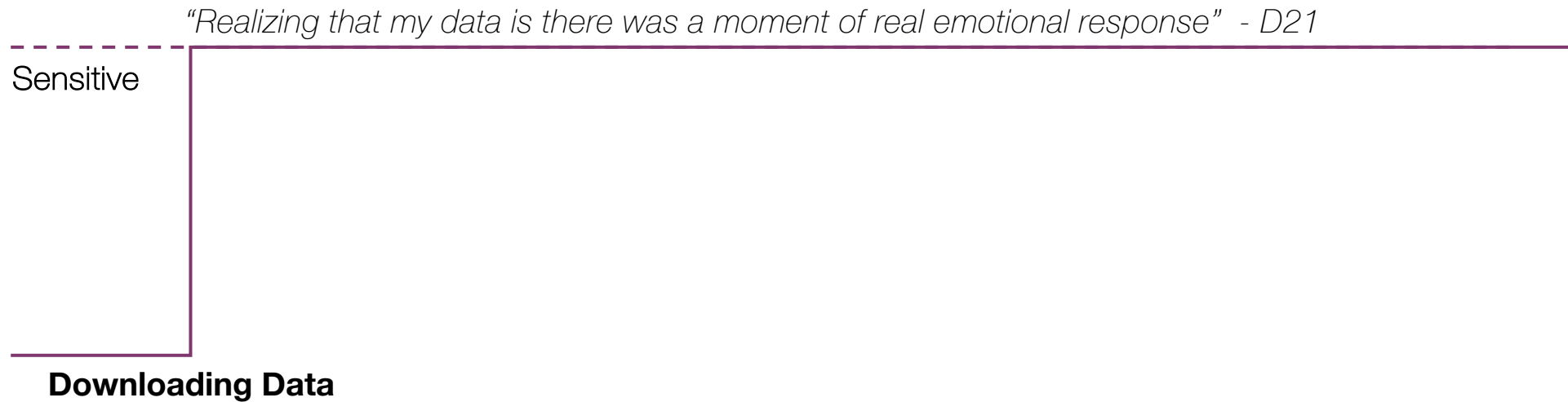


OK Google, set an alarm

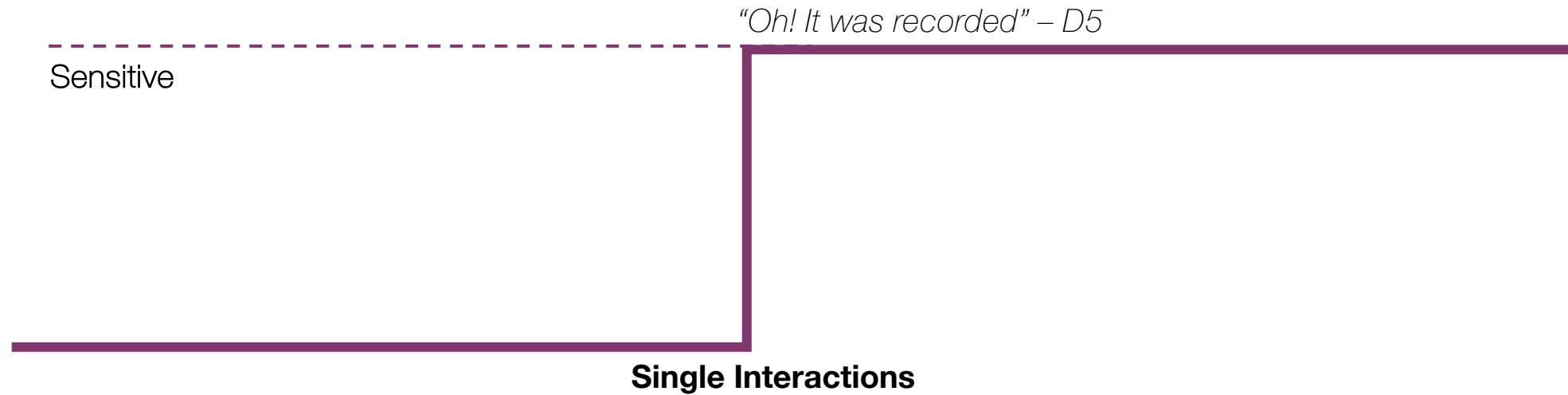
Sensitive data when it comes from a place of **playfulness** or **unconsciousness**.

Intimate data when it comes from a place of **vulnerability**, or it accounts for **intimate contexts and activities**.

Results: Tipping Point



Results: Tipping Point



Results: Tipping Point



Implications

What is Sensitive About (Sensitive) Data?

Characterizing Sensitivity and Intimacy with Google Assistant Users

Alejandra Gómez Ortega, Jacky Bourgeois, Gerd Kortuem

