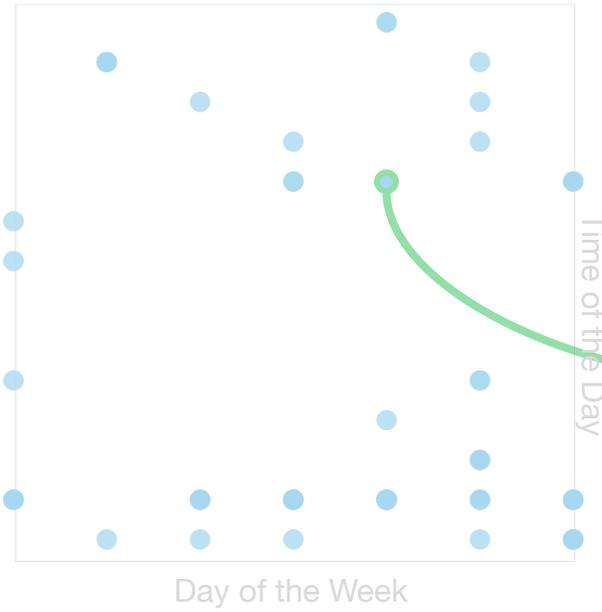


What is Sensitive About (Sensitive) Data?

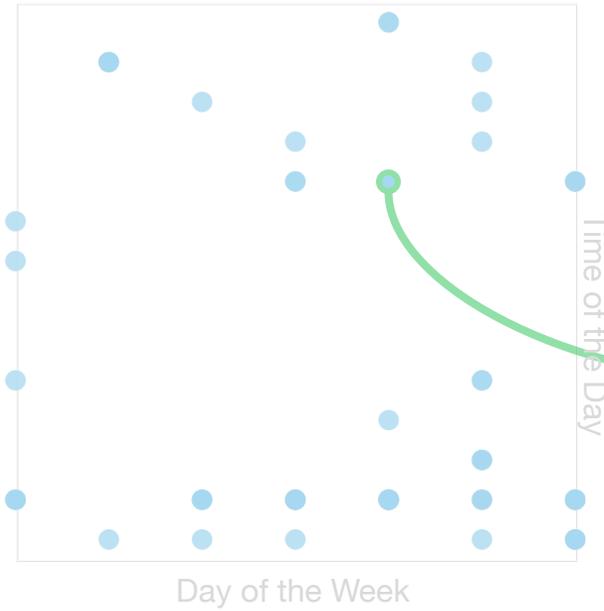
Characterizing Sensitivity and Intimacy with Google Assistant Users

Alejandra Gómez Ortega, Jacky Bourgeois, Gerd Kortuem



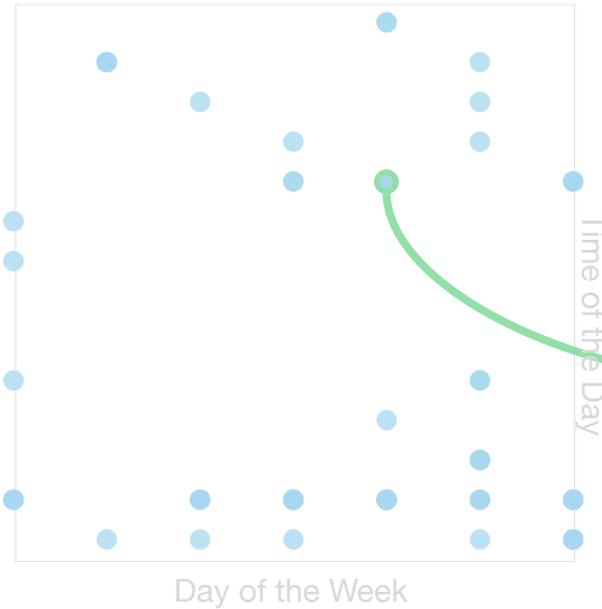


OK Google, set an alarm



Speech Record

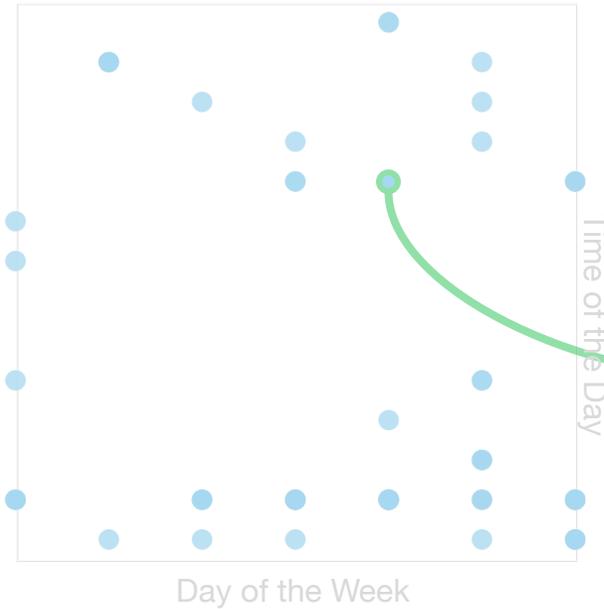
- Timestamp
- Transcript
- Audio Recording



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



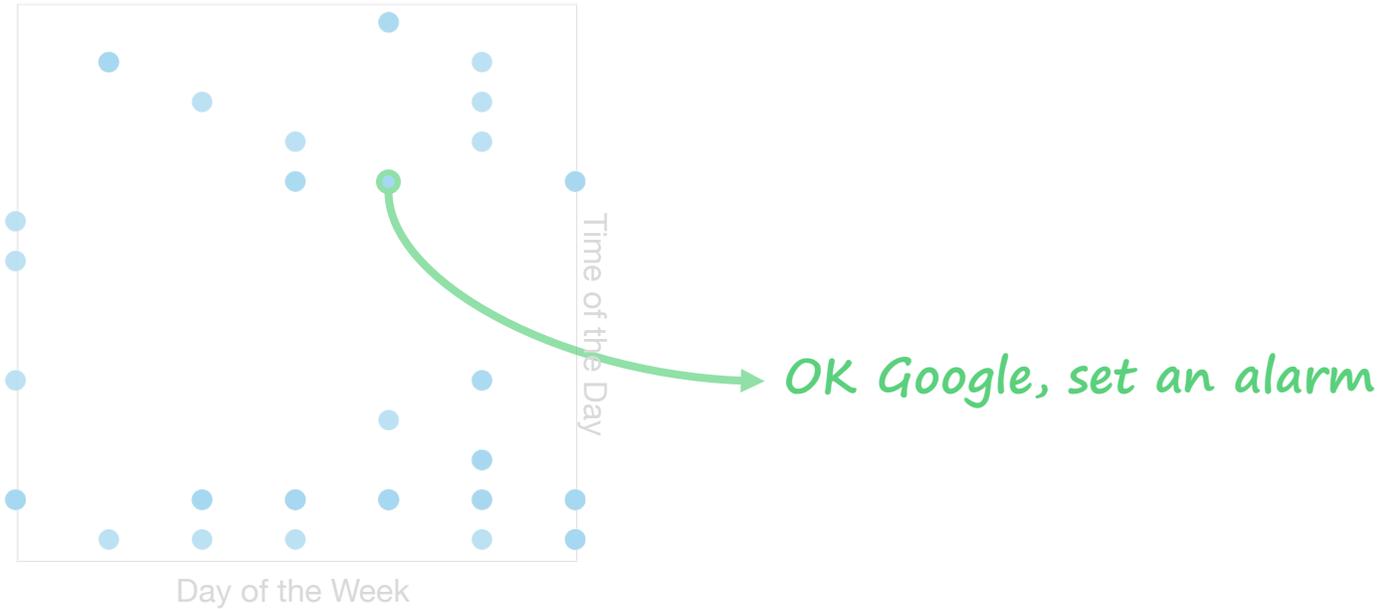
Research Question

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



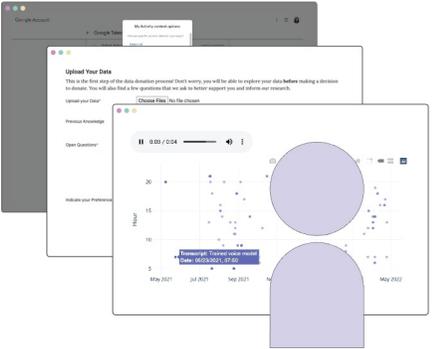
Speech Records: Sensitive or Intimate Data?

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



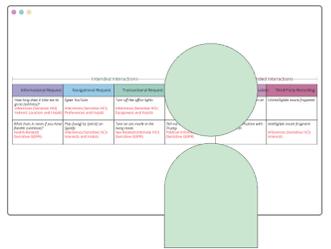
¹Malkin et al., 2019

Methodology



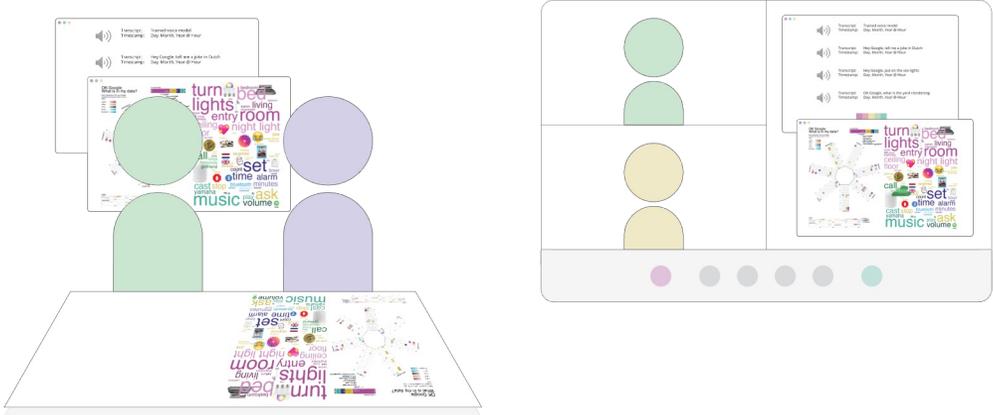
The screenshot shows a web interface for uploading data. It includes a 'Previous Knowledge' section, an 'Open Question?' section, and a scatter plot with a y-axis labeled 'Hours' ranging from 5 to 20 and an x-axis with dates from May 2021 to May 2022. Two stylized human figures, one green and one purple, are overlaid on the plot.

Collection of Speech Records through Data Donation



The screenshot shows a table with columns for 'Intended interpretation', 'Intended interpretation', and 'Intended interpretation'. The table contains several rows of text, likely representing speech records. Two stylized human figures, one green and one purple, are overlaid on the table.

Familiarization and Classification of Speech Records



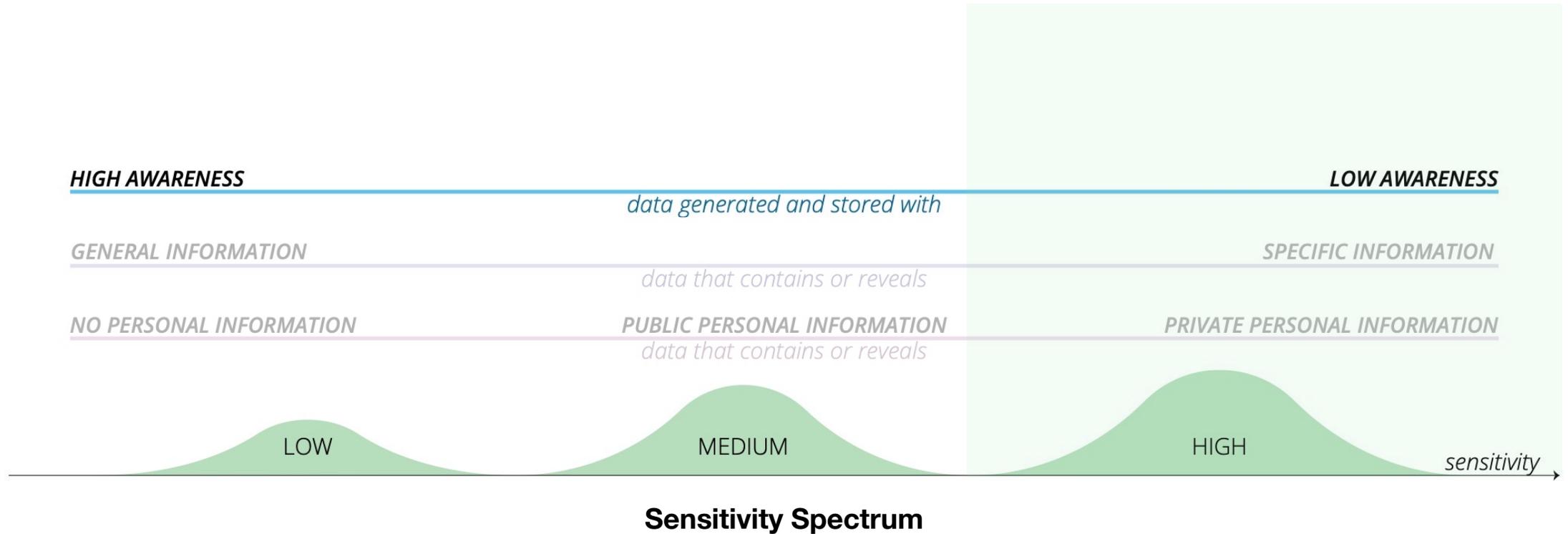
The illustration shows an interview setup. A laptop is in the foreground with a word cloud on its screen. A monitor is behind it, displaying a similar word cloud. Two stylized human figures, one green and one purple, are positioned behind the laptop and monitor. To the right, a separate monitor displays a list of speech records with audio icons.

Interviews

Results

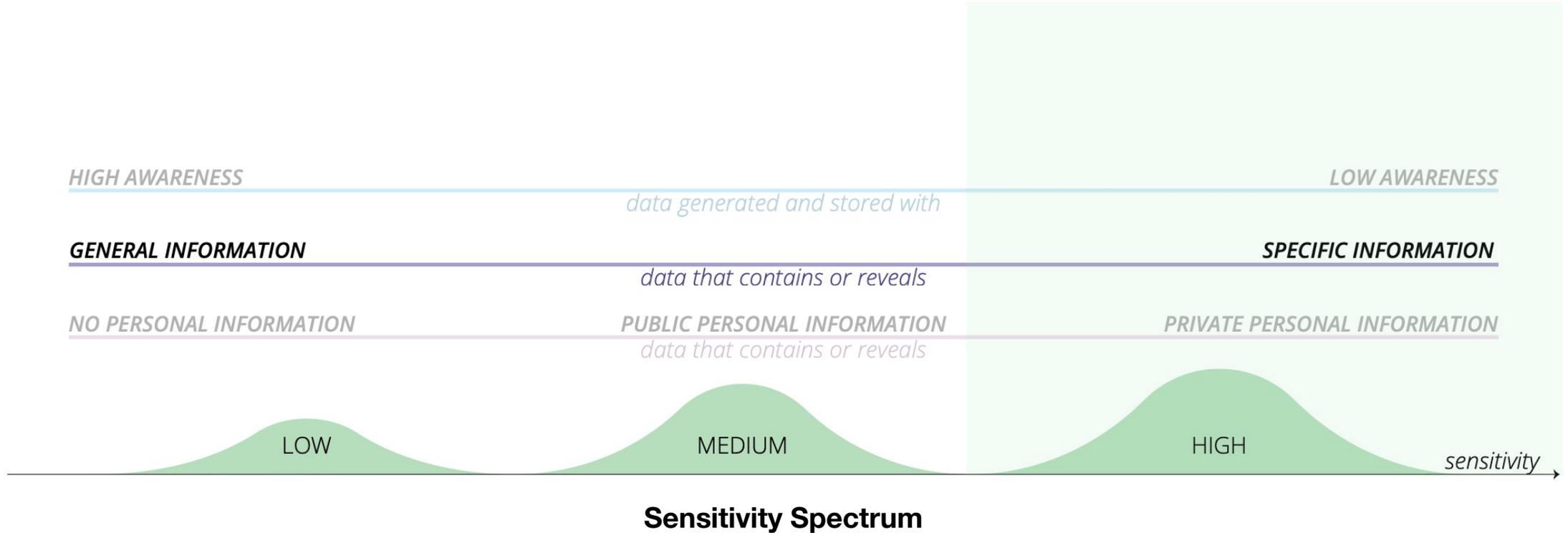
Results: Sensitive Speech Records

Intrusive, specific, and (un)available.



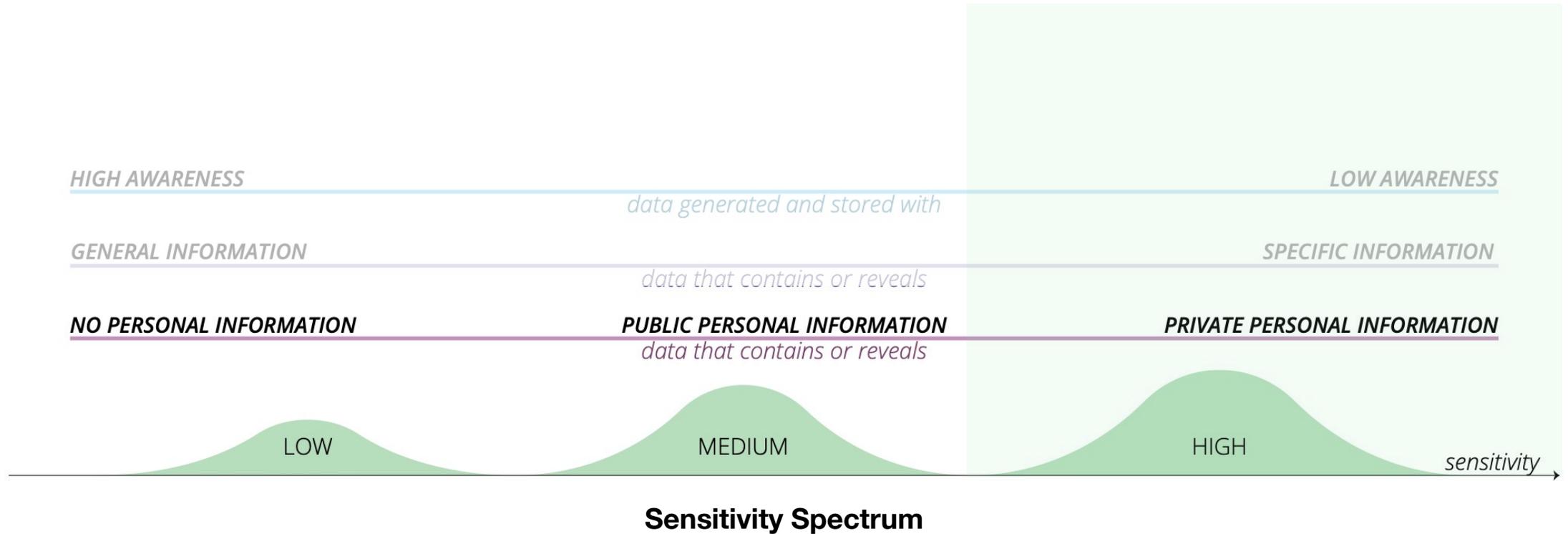
Results: Sensitive Speech Records

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



Results: Sensitive Speech Records

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?

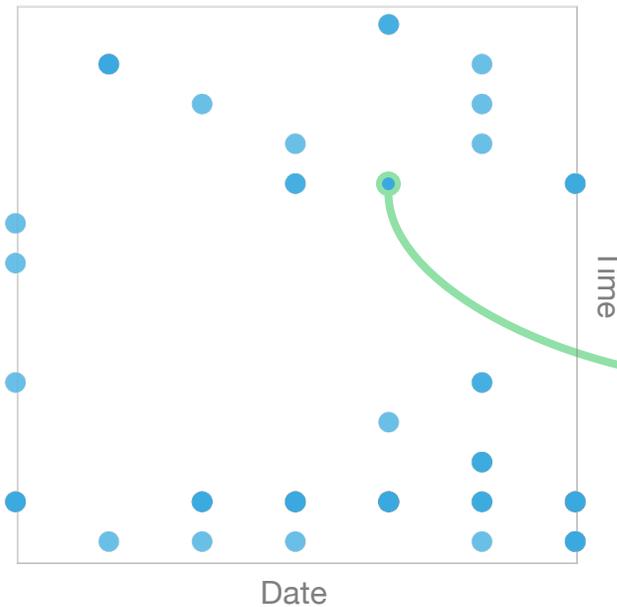
And it feels like **it was in bed with us**, kind of, and so it's not the **public voice** I have when I'm online or with my friends.

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?

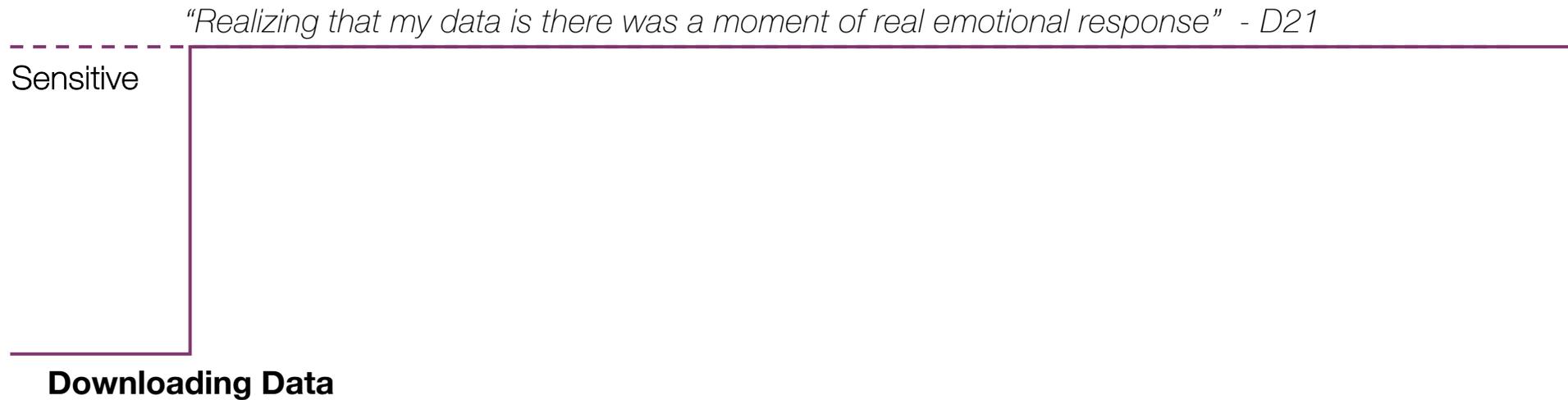


OK Google, set an alarm

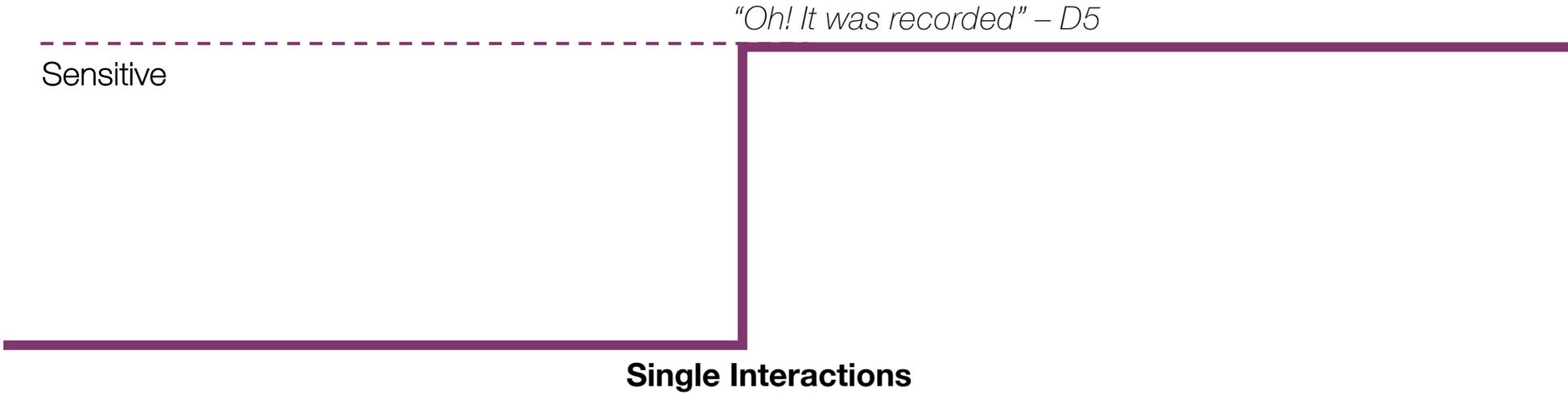
Sensitive data when it comes from a place of **playfulness** or **unconsciousness**.

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

