

Design Internship

Unconscious Bias in Conversational AI

August 2022

Edosa Leta

Agenda

Introduction

- Introduction to Unconscious Bias
- Problem Statement
- User Goal
- Shape of Data

Solution

- Real World Scenario
- Research Statement
- User Study
- Guidelines

Next Steps

- Summary
- This Week
- The Way Forward



Introduction

What is Unconscious Bias?

What is Unconscious Bias?

Biases towards certain ideas or groups made outside our conscious awareness

What is Unconscious Bias?

Biases towards certain ideas or groups made outside our conscious awareness

It affects the way people perceive other people and make decisions

Introduction

Bias in Conversational AI

Bias in Conversational AI

Good AI model == Good Data

Bias in Conversational AI

Good AI model == Good Data

AI solves the problem of bias since it uses Machine learning algorithms to make predictive decisions and removes the subjective interpretations

Problem Statement

Problem Statement

How do we find biases in a conversational AI and design an inclusive model?

Situation

User Goal



User Goal



As a user of the digital assistant, I need to feel comfortable interacting with the assistant so that I can efficiently accomplish tasks through natural language

Shape of Data



Shape of Data

United States

240M English Speakers
30 Major Dialects



Shape of Data

United States

240M English Speakers
30 Major Dialects

World

1.5B English Speakers
160 Major Dialects

Solution

For a Digital Assistant to be inclusive, it has to be trained with a wide range of utterances collected from a diverse user group

- 📅 Our goal is to create a digital assistant that understands all the users speaking english

Understanding

To understand the situation better, we started locally and looked at the different dialects in the US - i.e. Cajun English, Texas English, Spanglish, Baltimorese, AAVE,

Scope

For this project, we selected users who speak one of the major dialects in the US -
African American Vernacular English (AAVE)

Research Statement

If the DA is able to accurately recognize and respond to written utterances from the selected user group, then it suggests the model is inclusive of that group

*DA - Digital Assistant

Why a User Study?

To understand different trends and patterns of biases in the model

It Involved:

- 1 hour moderated sessions
- 12 AAVE participants
- Interaction with Oracle Digital Assistant

Methodology

Introduction →

Scenarios →

Methodology

Introduction → Uploading a receipt

Scenarios →

Methodology

Introduction → Uploading a receipt

Scenarios → Creating an expense without a receipt

Methodology

Introduction → Uploading a receipt

Scenarios → Creating an expense without a receipt
Modifying previous expense

Methodology

Introduction

→ Uploading a receipt

Scenarios

→ Creating an expense without a receipt

Modifying previous expense

Creating reports

Scenario

Imagine you were asked to buy 3 iPhone cables from the store and you paid \$18 in total. You forgot to collect the receipt. Try to get the assistant to help you get the money back.

Responses

I paid cash for \$18 for 3 iPhone cables for my boss last Wed

I want to file an expense for iPhone cables

I bought an iPhone cable for 18 dollars no receipt

Hello I bought 3 iPhone cables for \$18, I would like to expense this please. I forgot the receipt.

Assessment

How inclusive is the model?



- Not-inclusive
- Inclusive

Inclusiveness

Assessment

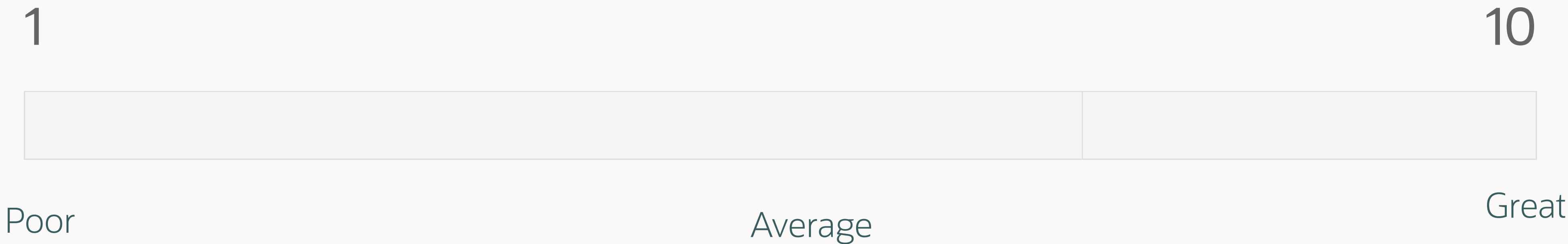
How inclusive is the model?



- Not-inclusive
- Inclusive

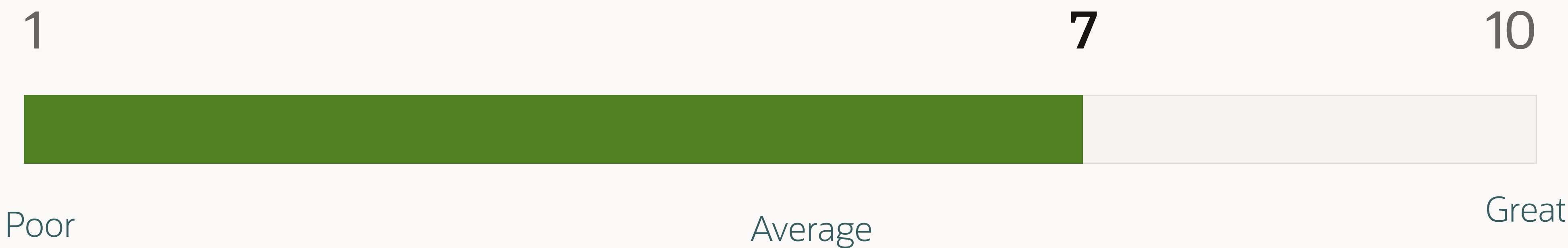
Inclusiveness

How did the DA perform?



How would you rate the experience with the Oracle Digital Assistant?

How did the DA perform?

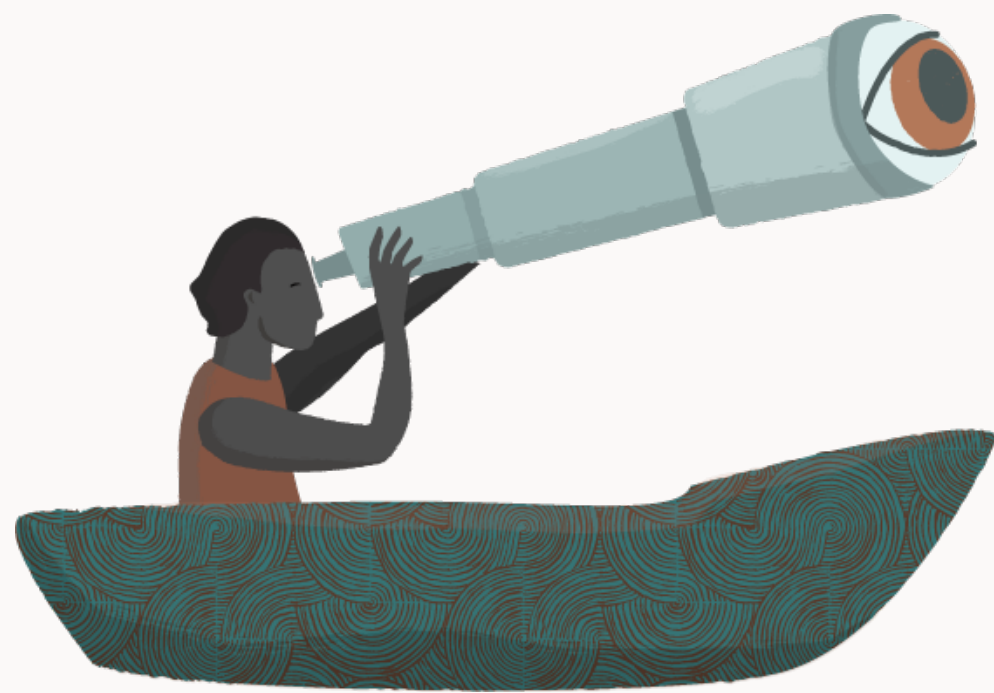


How would you rate the experience with the Oracle Digital Assistant?

Insights

Most people **code switch** depending on the enviroment

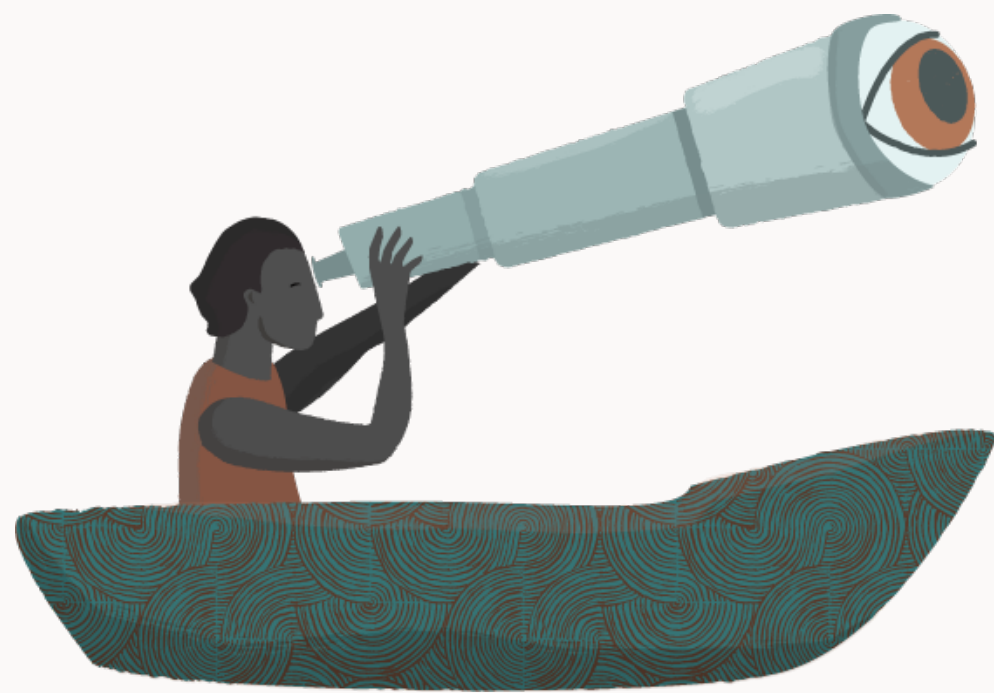
“ I choose to speak in a different way when i am in a work setting vs when I am home or talking to my friends”



Insights

Users response is majorly influenced by the **style and tone** of the digital assistant

“Once the DA responded create after I clicked create I knew the conversation would not be as organic as I would have if the greeting had been more conversational”



Insights

Users...

tend to **speak carefully** when **working with money**

prefer to be **concise** to just get the job done

communicate differently when **speaking** vs **writing**



User Feedback

Did you **feel comfortable enough** with the DA that you think that there was an influence of AAVE in your conversation?



“I wouldn't use AAVE in this type of interaction”

“Yes, it did for the most part. There were some hiccups, like when it recorded descriptions, but it was mostly responsive.”

“Somewhat yes because I could ask in many ways that I normally speak and would either get the correct answer or no answer as it did not know”

User Feedback

Would you have **said things differently** if you were talking to the DA as if it was an AAVE speaker?



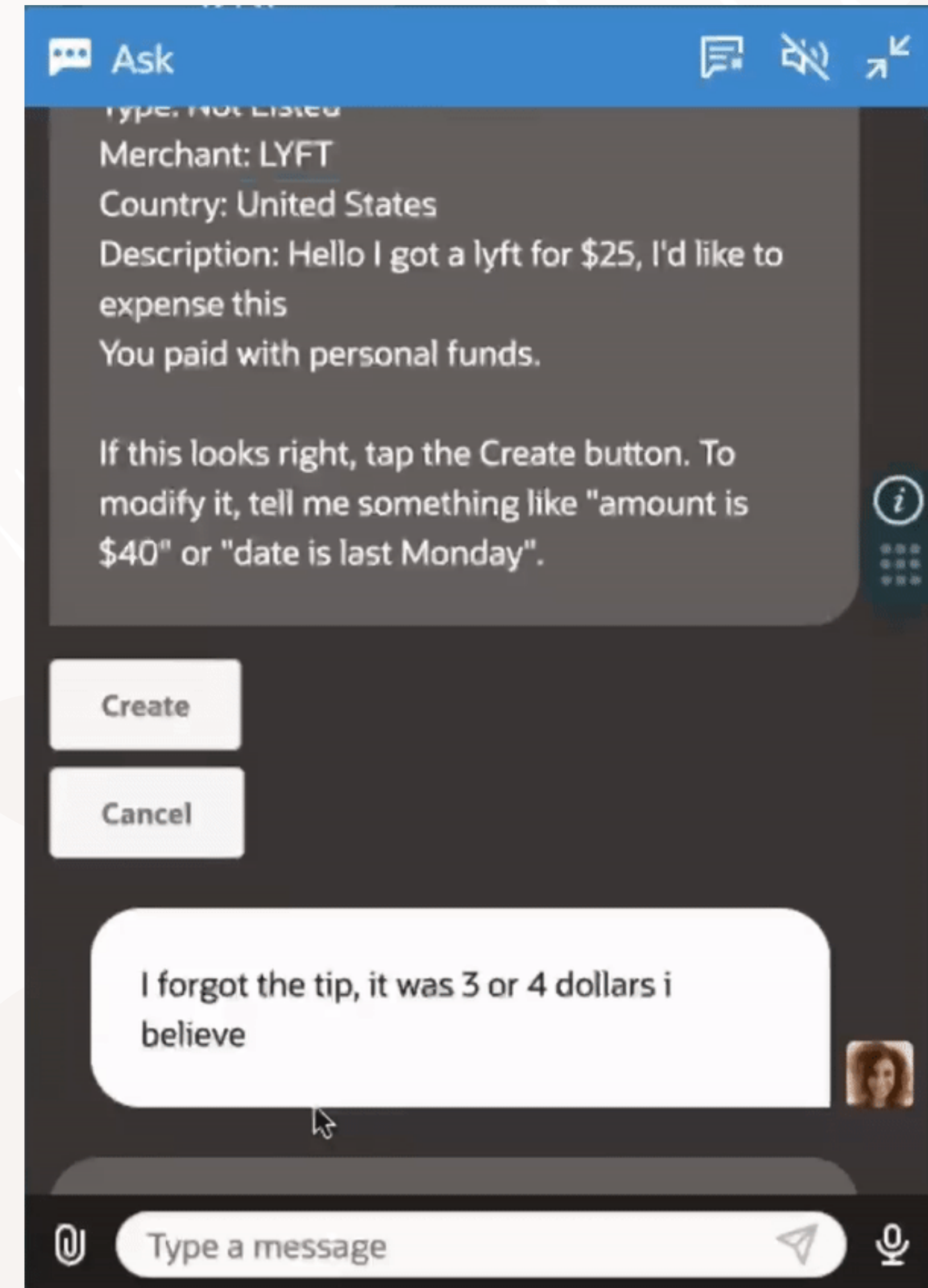
“No. That would be weird. An AAVE digital assistant would be odd.”

“It's possible. The more the DA can appear as if they have some characteristics I'm familiar with the more relaxed I would be”

Additional Insights - Usability Feedback

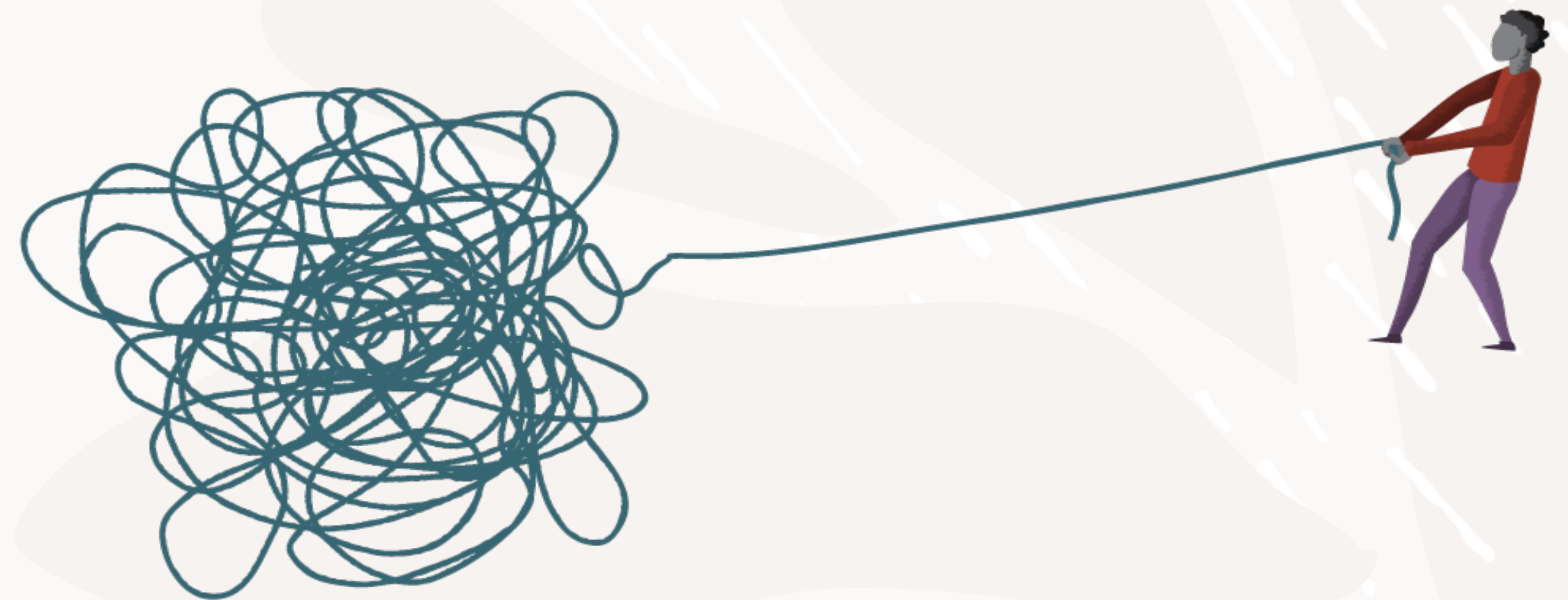
- ✓ Some of the instructions were too wordy
- ✓ Users want to know how long it takes for the expenses to be approved
- ✓ Voice input has short listening time
- ✓ Payment method should include more options (AMEX, credit card, debit card ...)

The DA wasn't able to follow the context of the conversation if the user deviates from the topic



Challenges

- ✓ More than half of the participants didn't show up for the study
- ✓ Some Users chose to speak a certain way with the digital assistant
- ✓ The Expense skill doesn't have a lot of room for conversation



Guideline Highlights

- Include more natural language that people can relate to when creating responses
- Include different responses for the same question - gives a natural feel to the conversation flow
- Test models with diverse range of data with both corporate and cultural informal words
- Diversify data collection, data handling, and processing
- Collect and compare utterances from voice and text to see efficiency

Guideline Highlights

- Include more natural language that people can relate to when creating responses
- Include different responses for the same question - gives a natural feel to the conversation flow
- Test models with diverse range of data with both corporate and informal words
- Diversify data collection, data handling, and processing
- Collect and compare utterances from voice and text to see efficiency

Guideline Highlights

- Include more natural language that people can relate to when creating responses
- Include different responses for the same question - gives a natural feel to the conversation flow
- Test models with diverse range of data with both corporate and informal words
- Diversify data collection, data handling, and processing
- Collect and compare utterances from voice and text to see efficiency

Guideline Highlights

- 📄 Include more natural language that people can relate to when creating responses
- 📄 Include different responses for the same question - gives a natural feel to the conversation flow
- 📄 Test models with diverse range of data with both corporate and informal words
- 📄 Diversify data collection, data handling, and processing
- 📄 Collect and compare utterances from voice and text to see efficiency

Guideline Highlights

- ↳ Include more natural language that people can relate to when creating responses
- ↳ Include different responses for the same question - gives a natural feel to the conversation flow
- ↳ Test models with diverse range of data with both corporate and informal words
- ↳ Diversify data collection, data handling, and processing
- ↳ Collect and compare utterances from voice and text to see efficiency

Summary

Summary

The natural conversation flow encourages users to continue interacting with the DA

Summary

The natural conversation flow encourages users to continue interacting with the DA

There are some common patterns across different dialects the DA can learn and understand

Summary

The natural conversation flow encourages users to continue interacting with the DA

There are some common patterns across different dialects the DA can learn and understand

The functionalities of other Digital assistants (like Alexa, Siri..) is very different

Next Steps

This Week ...

Re-train the model with the utterance collected

Compile and document all the usability feedback to file a bug

Document the guidelines for best practices



The Way Forward

Conduct a bigger user study

- different skills
- diverse user groups

Design a metric to assess the model for inclusivity





Thank You!

Q & A