ORACLE

Taking the Chatbot's Word for it?
Trust with Enterprise Chatbots

Anna Wichansky | anna.wichansky@oracle.com

Sneha Kanneganti | sneha.Kanneganti@oracle.com

February 11, 2020

1

Research Questions

1 What are people's mental models when interacting with the Expenses Chatbot?

2 Are people able to successfully file expenses using the chatbot?

3 How do they feel about their entire experience?

Who

All Oracle employees provisioned on expenses chatbot

20
Participants

17 iPhone

3 Android

12 Female

8 Male

8 New
Never used expenses chatbot

7 Experienced
Submitted 1-10 receipts

5 Expert
Submitted >10 receipts

Study Design

—

1 **Pre-evaluation perception**
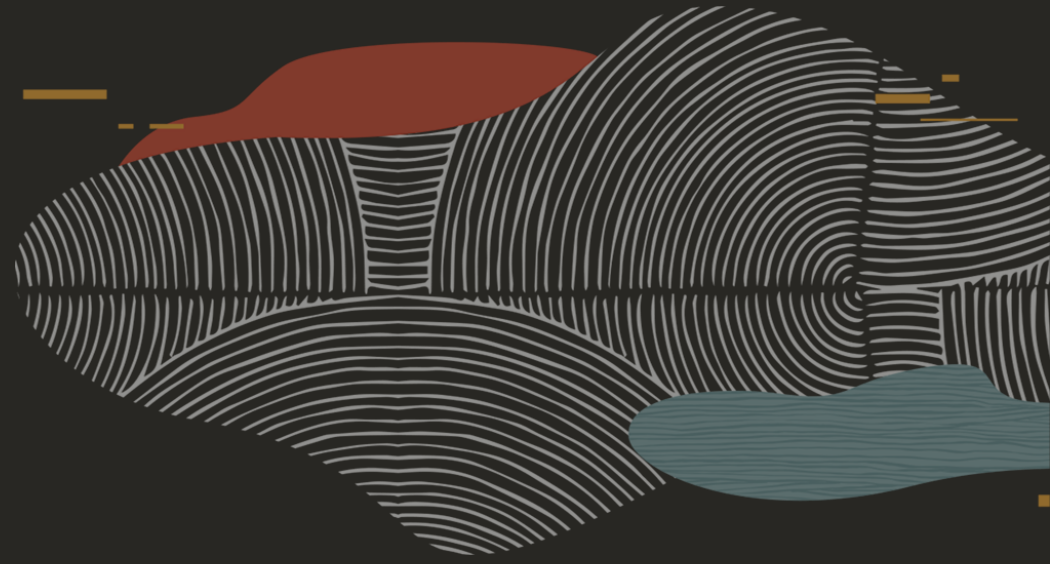Trust, reliance, confidence, and expected ease of use

2 **Evaluate chatbot**
Submit 5 receipts, check receipt in the web application

3 **Post-evaluation perception and**
Trust, reliance, confidence, and ease of use, SUS, NPS
feedback

Results

Users' mental models of the chatbot generally aligned with the way the system worked.

However, there were some areas of disconnect.

Chatbot's Conceptual Model

①　Initiate

②　Enter details

③　Confirm details

④　Submit

Send image or manual entry

Submit expense type

Summary message

Confirmation message

Confirmation code

Receipt Image

Expense Type List

Expense Summary

Confirmation Message
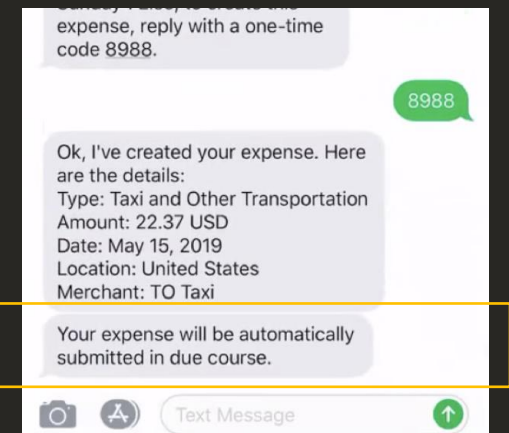
Disconnects with the chatbot conceptual model

(1) **Initiate**

👎 How to begin?

👎 Why so long? Is it working?

(2) **Enter details**

👎 Misread some receipt details

(3) **Confirm details**

👎 Can't make changes after the confirmation code is sent

(4) **Submit**

👎 Is an expense being created or being submitted?
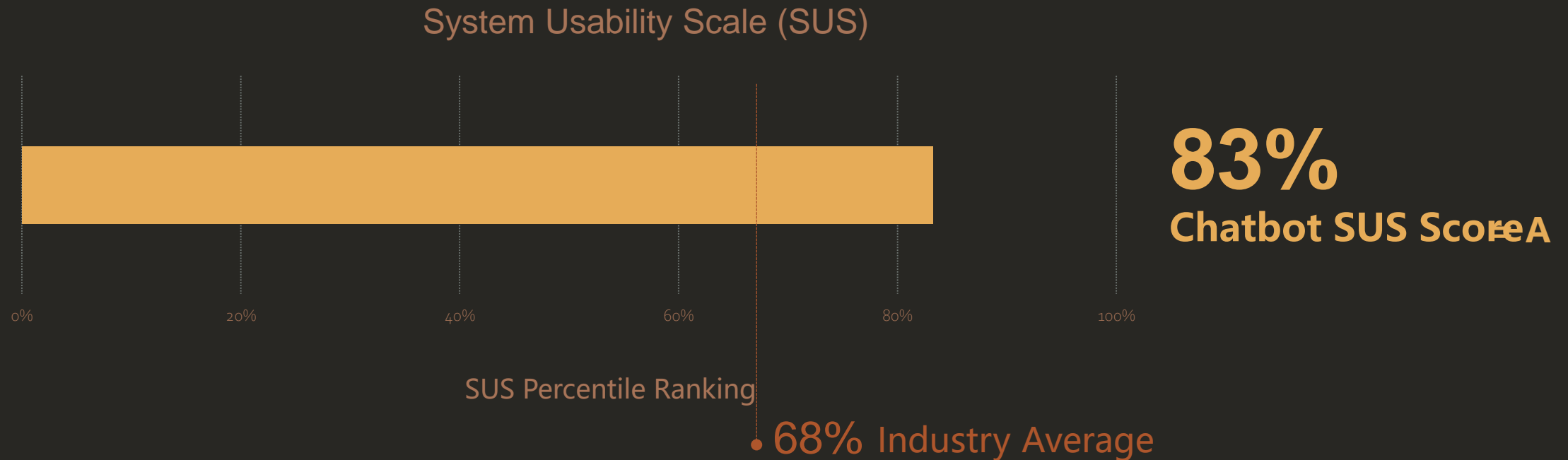
👎 What does "due course" mean?

👎 When will my manager see the expense?
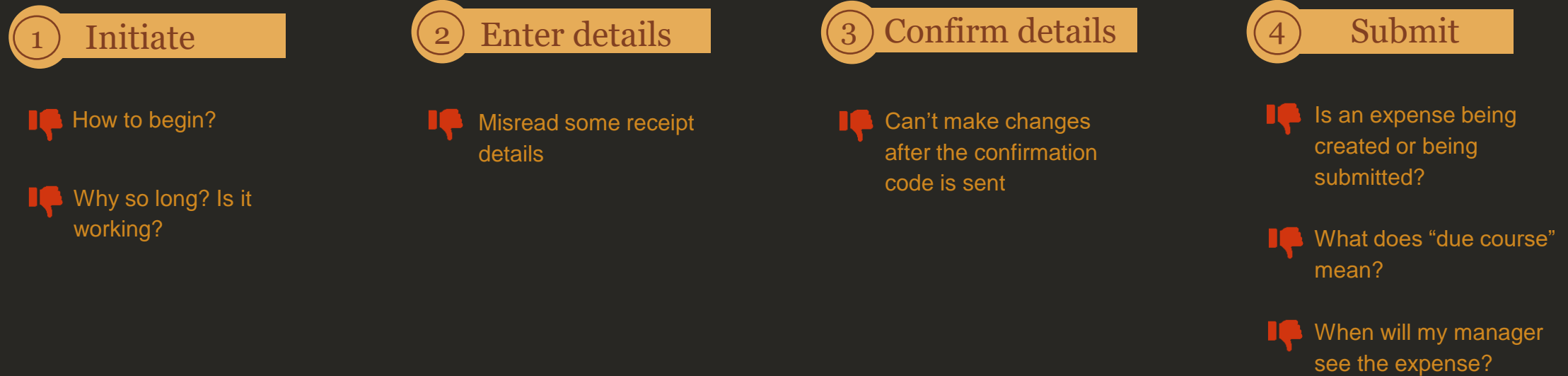
The idea of automating the expense filing process had high appeal

User perception of the Chatbot is very positive

Net Promoter Score (NPS): **55%**

10%	25%	65%

■ Detractors   ■ Neutral   ■ Promoters
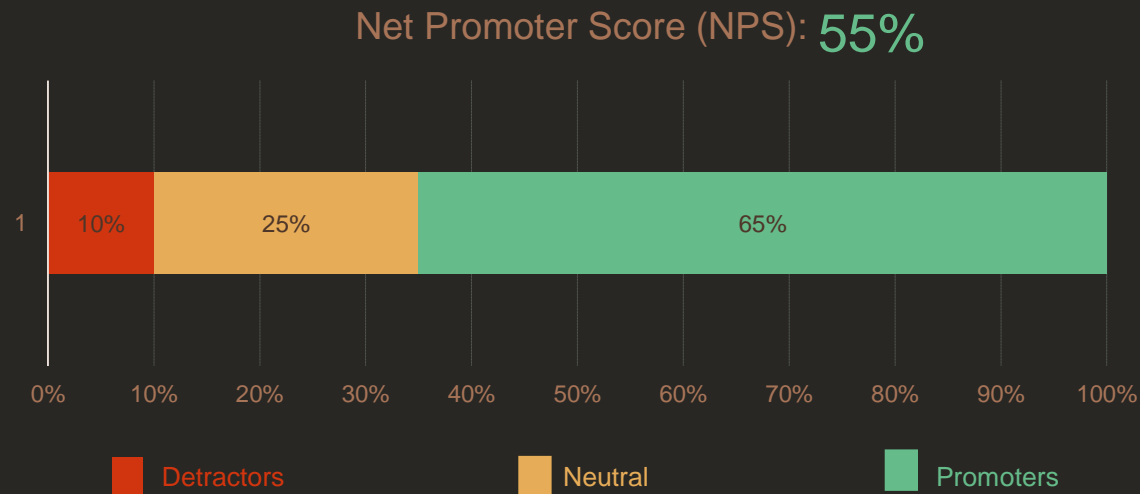
- NPS = % Promoters - % Detractors
- Scores above 0% are good

The chatbot SUS score above industry average

System Usability Scale (SUS)

83%
Chatbot SUS Score A

SUS Percentile Ranking

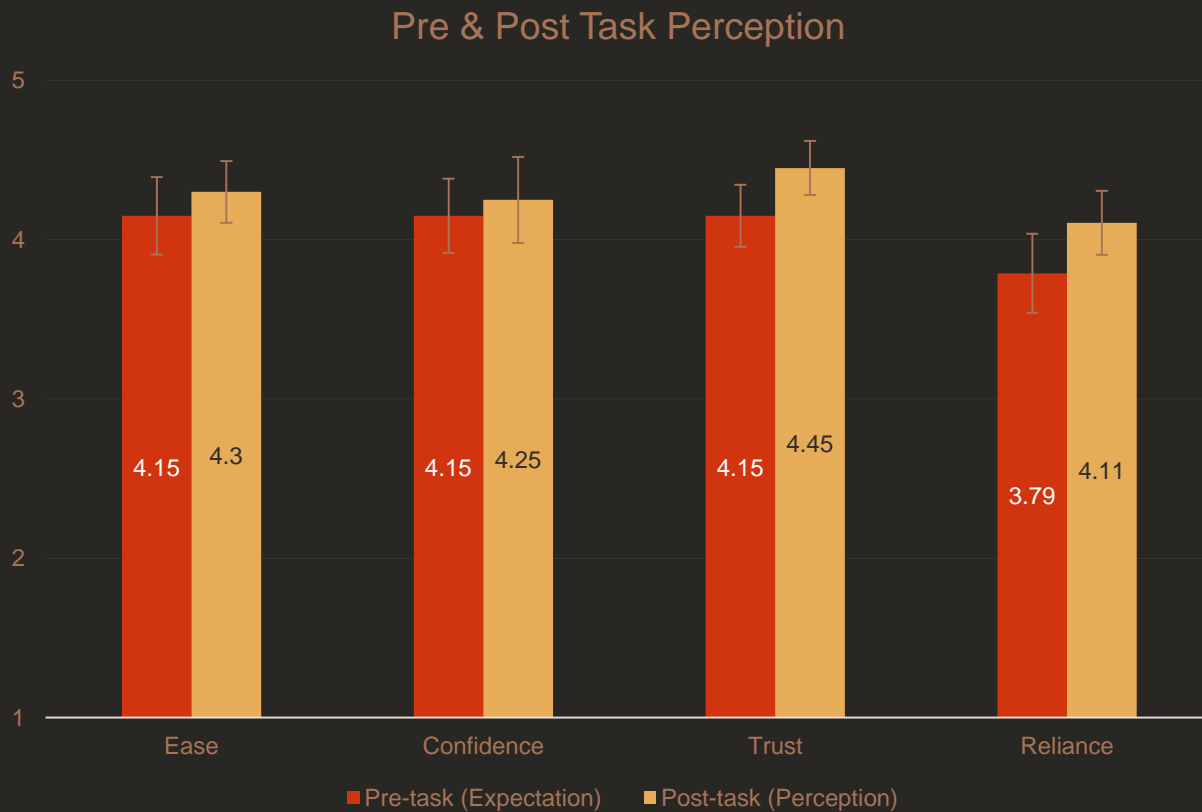68% Industry Average

0%   20%   40%   60%   80%   100%

The outcome of such positive perception was **high trust**

Metrics to measure trust: Trust in Automation scale

- Psychometrically developed model

- 6 underlying dimensions, 19 5-point scales

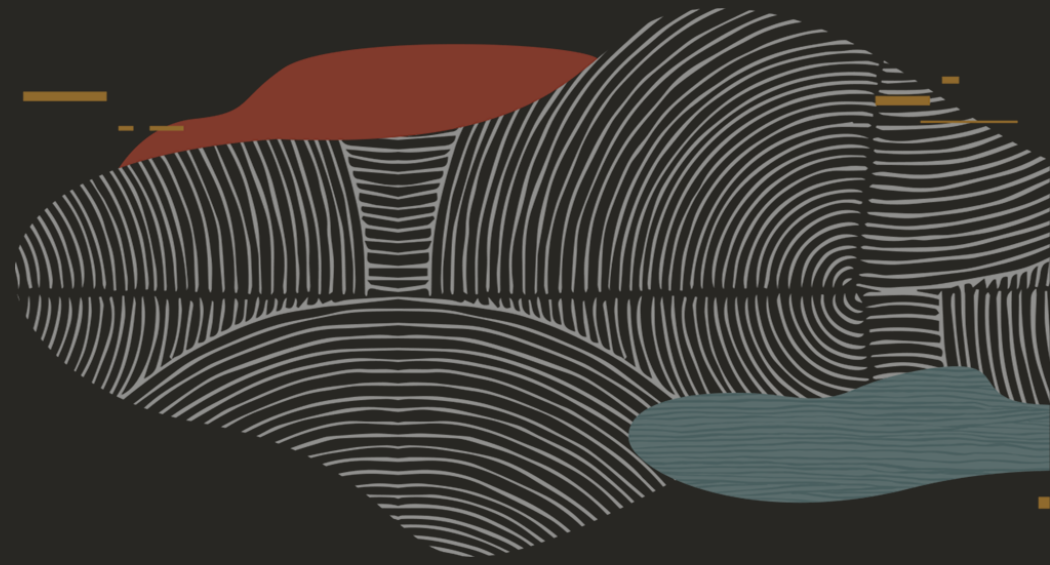- Participants rated 2 TIA scales before and after chatbot

	Strongly disagree → Strongly agree				
	1	**2**	**3**	**4**	**5**
I trust the Expenses Chatbot					
I can rely on the Expenses Chatbot					

Körber, M. (2019) https://osf.io/y3jn5/

https://github.com/moritzkoerber/TiA_Trust_in_Automation_Questionnaire

Users trusted the Chatbot despite making many errors

Pre & Post Task Perception

Participants' trust ratings increased after interacting with the chatbot.

	Ease	Confidence	Trust	Reliance
Pre-task (Expectation)	4.15	4.15	4.15	3.79
Post-task (Perception)	4.3	4.25	4.45	4.11

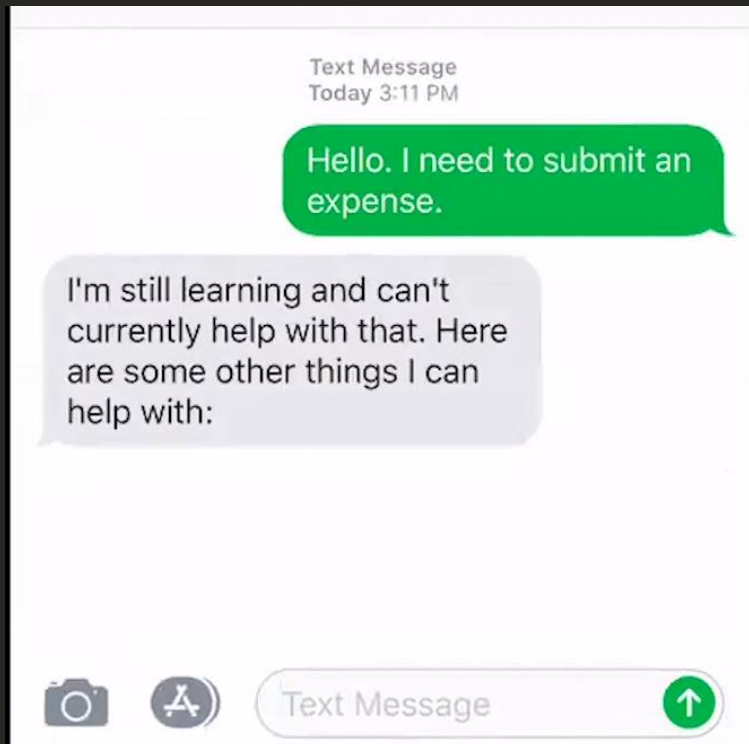■ Pre-task (Expectation)   ■ Post-task (Perception)

Top chatbot bloopers
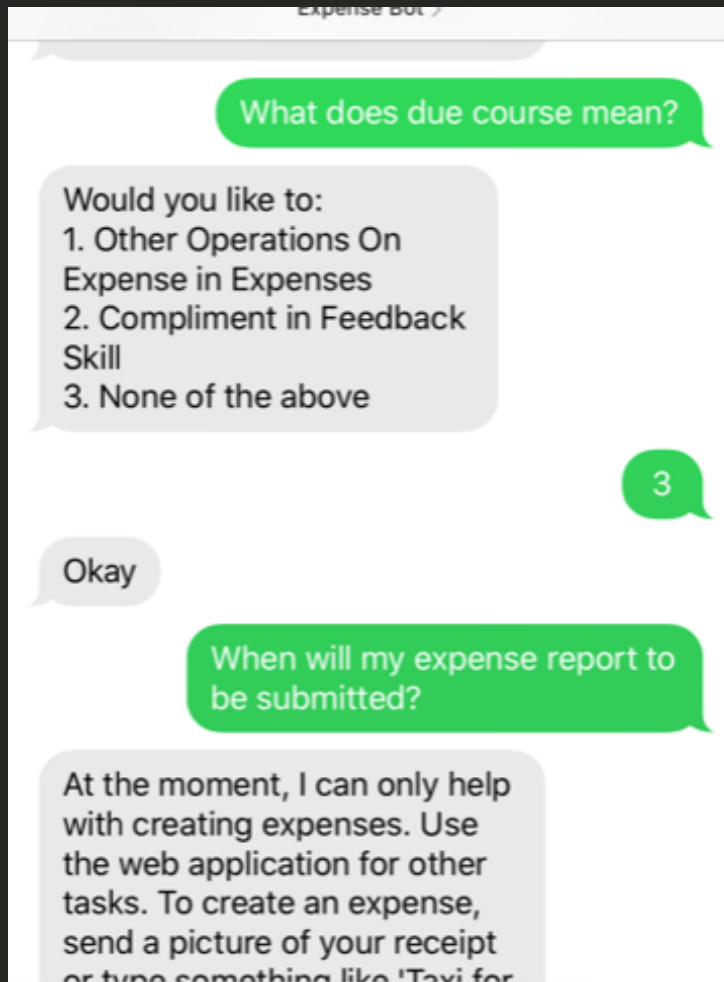
Yikes!....boy was that an expensive meal!
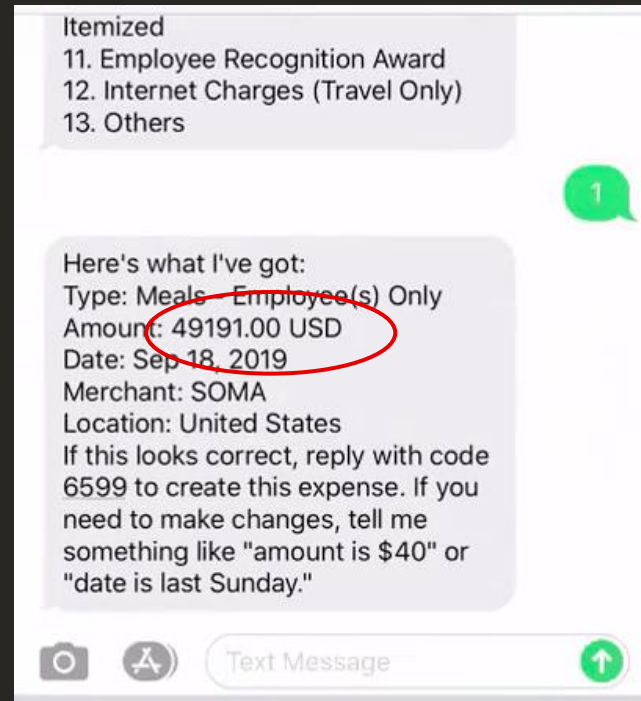
NLU Issues

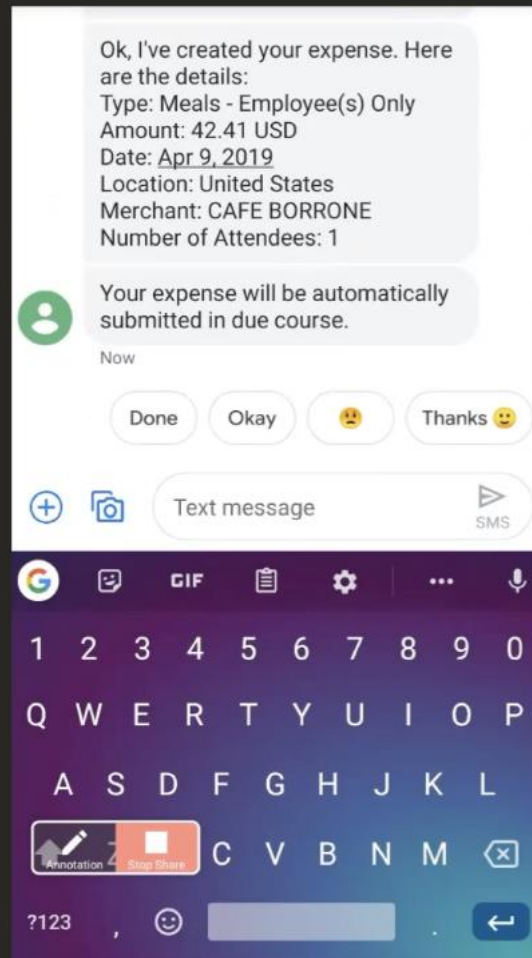Is the expense Chatbot telling you it can't help you submit an expense?

Unclear when the expense will be submitted

User wants a date or time.

User tries to rephrase the question. That doesn't work either.

Doesn't (yet) forgive errors easily

Participant forgets to include tip in expense amount and submits code.

Bot reads his correction text as a fresh expense submission attempt.

Do users trust the chatbot more than it deserves?

Yes

4.5/5
Trust rating

Subjective trust was high...

...despite errors and the limited use cases it supports.

We need to make sure the technology is robust enough to be worthy of the high trust.

Thank you