

Michelle Brachman

UX Research Portfolio

Connecting users with complex systems for 13+ years

About Me

I am a UX Researcher with a Ph.D. in Computer Science/Human-Computer Interaction and five years of experience specializing in Human-Centered AI.

I bridge the gap between complex systems and the reality of human needs and behavior using qualitative and quantitative methods.

My background in CS allows me to work with complex technical constraints, collaborate on working prototypes, and dig into large scale quantitative data.



When I'm not analyzing mental models or understanding user workflows, I'm usually hitting the pavement. There's nothing like a 10-miler to find the solution to a tricky problem.

Toolbox

Methods

Interview Design

Survey Design

Usability Study

A / B Study

Skills

Statistical Methods

Wireframing

Prototyping

Cross-functional
communication

Platforms

Qualtrics

Figma

Mural

Mechanical Turk

Technologies

SQL

Python / R

Javascript

HTML / CSS

Calibrating Trust in Agentic AI: A Foundational Study for BeeAI

Challenge: BeeAI needed a design framework for transparency.

Question: How do you design for appropriate trust in a complex agentic system?

Outcome: Impacted the interface for an open-source platform.

The screenshot shows a user interface for a conversational AI system. At the top, a message from "Michelle Brachman" says "provide best flight times from EWR to Costa rica in january". Below it, a response from "Peri" provides a general overview: "Based on the Google tool output, it seems that there are many flights available from EWR to Costa Rica in January, with prices ranging from \$92 to \$915. To find the best flight times, I'll need to narrow down the options based on your preferences, such as budget, departure and arrival times, and layovers." A follow-up question asks if the user wants to search for specific flight times or routes. The interface includes a sidebar titled "Sources" listing various websites and tools used for the search, such as Kayak, American Airlines, Skyscanner, and Orbitz. A search bar at the bottom contains the query "flights from EWR to Costa Rica in January".

Overview

- Goal:** Understand user information needs to support appropriate trust
- Role:** Lead UX Researcher/ HCI Research Scientist
- Timeline:** ~ 3 months
- Team:** PM, Design lead, designer, UX researcher, research scientists

Research Questions

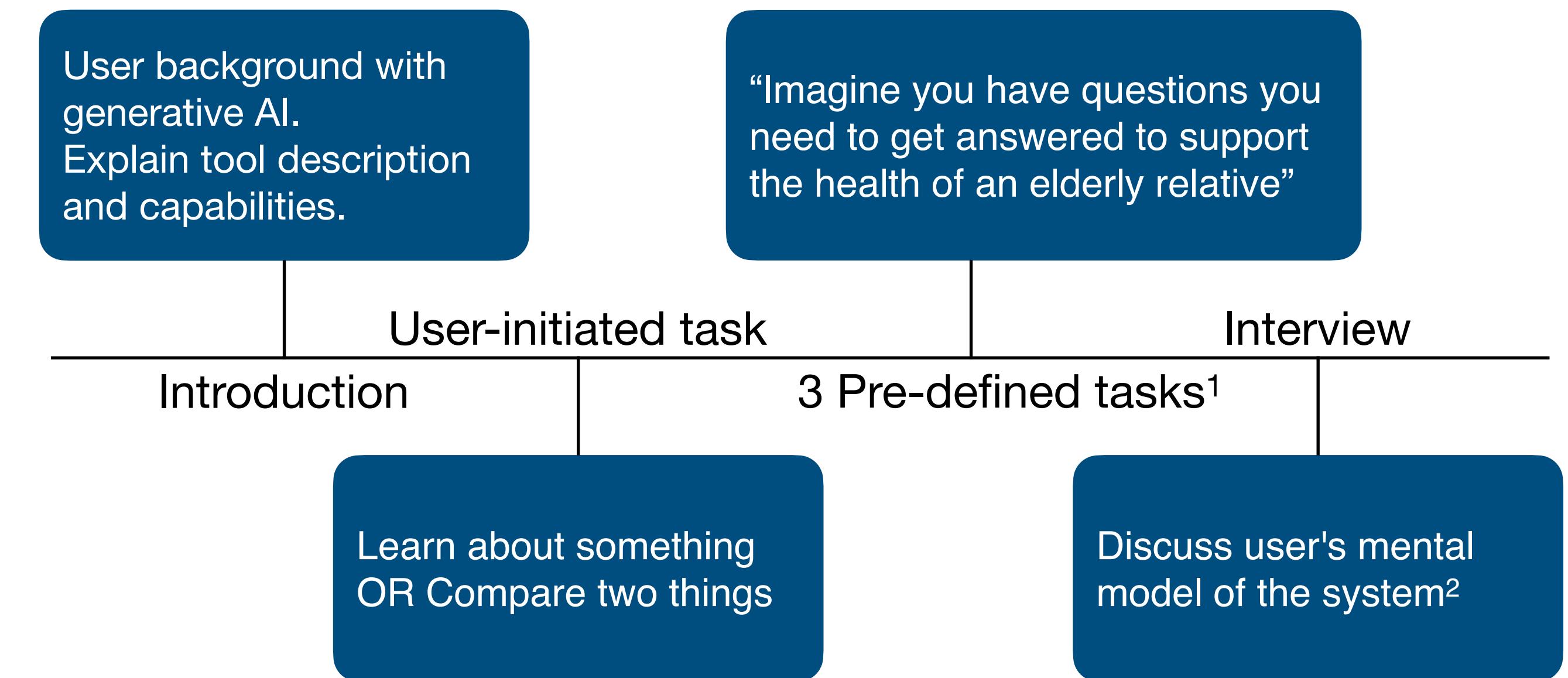
How did participants describe how the system works and the risks of using the system?

What information did participants use to judge their confidence in the accuracy of outputs?

What information did participants want to have about how the system works?

Methods & Protocol

- ◆ Semi-structured think-aloud/interview study with 24 participants recruited internally.
- ◆ Four information seeking tasks
- ◆ Questions after each task:
 - Rate confidence in accuracy
 - Information needed about system behavior
 - Is any information unnecessary or unclear



1. Kim, Sunnie SY, Q. Vera Liao, Mihaela Vorvoreanu, Stephanie Ballard, and Jennifer Wortman Vaughan. "" I'm Not Sure, But...": Examining the Impact of Large Language Models' Uncertainty Expression on User Reliance and Trust." In Proceedings of the 2024 ACM Conference on Fairness, Accountability, and Transparency, pp. 822-835. 2024.

2. Kulesza, Todd, Simone Stumpf, Margaret Burnett, Sherry Yang, Irwin Kwan, and Weng-Keen Wong. "Too much, too little, or just right? Ways explanations impact end users' mental models." In 2013 IEEE Symposium on visual languages and human centric computing, pp. 3-10. IEEE, 2013.

Key Insights

Mental Model

Participants primarily assumed that the system **worked like a search system.**

Participants were concerned about the **risks of data sources.**

So from the user experience, when I see it, it looks more like a search tool rather than generating answer on its own.

Key Insights

Mental Model

Participants primarily assumed that the system **worked like a search system**.

Participants were concerned about the **risks of data sources**.

So from the user experience, when I see it, it looks more like a search tool rather than generating answer on its own.

Information used to evaluate

Sources were the primary way users judged their confidence in the accuracy of system outputs. They also evaluated the system's **actions** and used **their own prior knowledge**.

When I see that it's going out on Google or Wikipedia, I feel confident that the information it's gathering is accurate.

Key Insights

Mental Model

Participants primarily assumed that the system **worked like a search system**.

Participants were concerned about the **risks of data sources**.

So from the user experience, when I see it, it looks more like a search tool rather than generating answer on its own.

Information used to evaluate

Sources were the primary way users judged their confidence in the accuracy of system outputs. They also evaluated the system's **actions** and used **their own prior knowledge**.

When I see that it's going out on Google or Wikipedia, I feel confident that the information it's gathering is accurate.

Information needs

Participants wanted to better understand **how the system worked**.

They asked questions like:

- *Why is it prioritizing certain sources/ how is it prioritizing?*
- *How are the answers created from the sources? Which sources are used?*
- *What is the system's confidence in its response?*

Impact

Context: Leadership asked for a clean/minimalist UI, like other well-known AI assistants.

However, the data showed that for an agentic system, **users wanted to know more about how the system worked.**

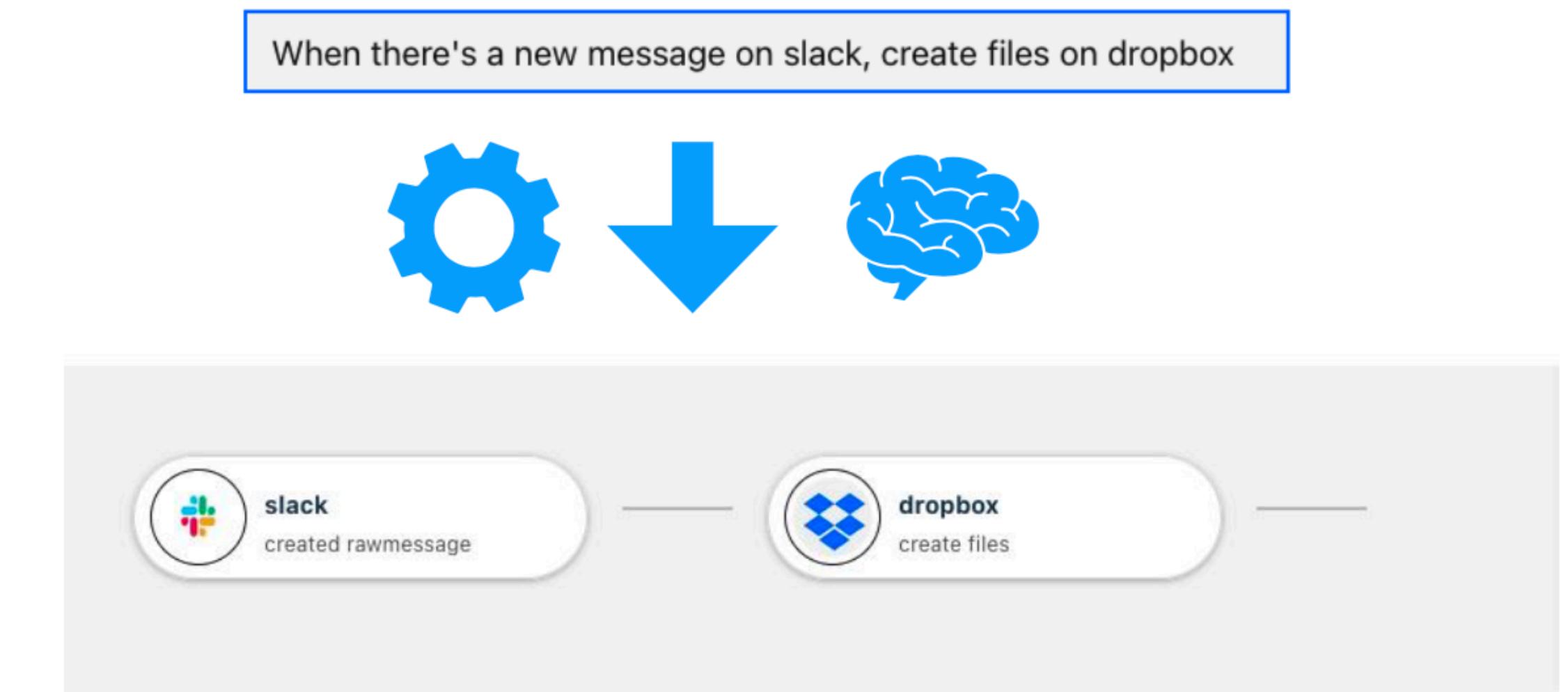
Through presentations of my findings and discussions with the design team, we found a common ground that exposed the agent's reasoning steps and sources, while minimizing the distraction of this information.

Explanations to Support use of a Natural Language System for Automation

Challenge: IBM AppConnect wanted to provide natural language automation but users didn't understand how to use it.

Question: How can we design explanations that support effective use of a natural language system?

Outcome: Influenced the design of the AppConnect natural language generation experimental feature.



Overview

Goal: Evaluate explanation types

Role: Lead UX Researcher/ HCI Research Scientist

Timeline: ~ 1 year

Team: PM, designers, AI engineer, full-stack engineers, AI researchers

Research Questions

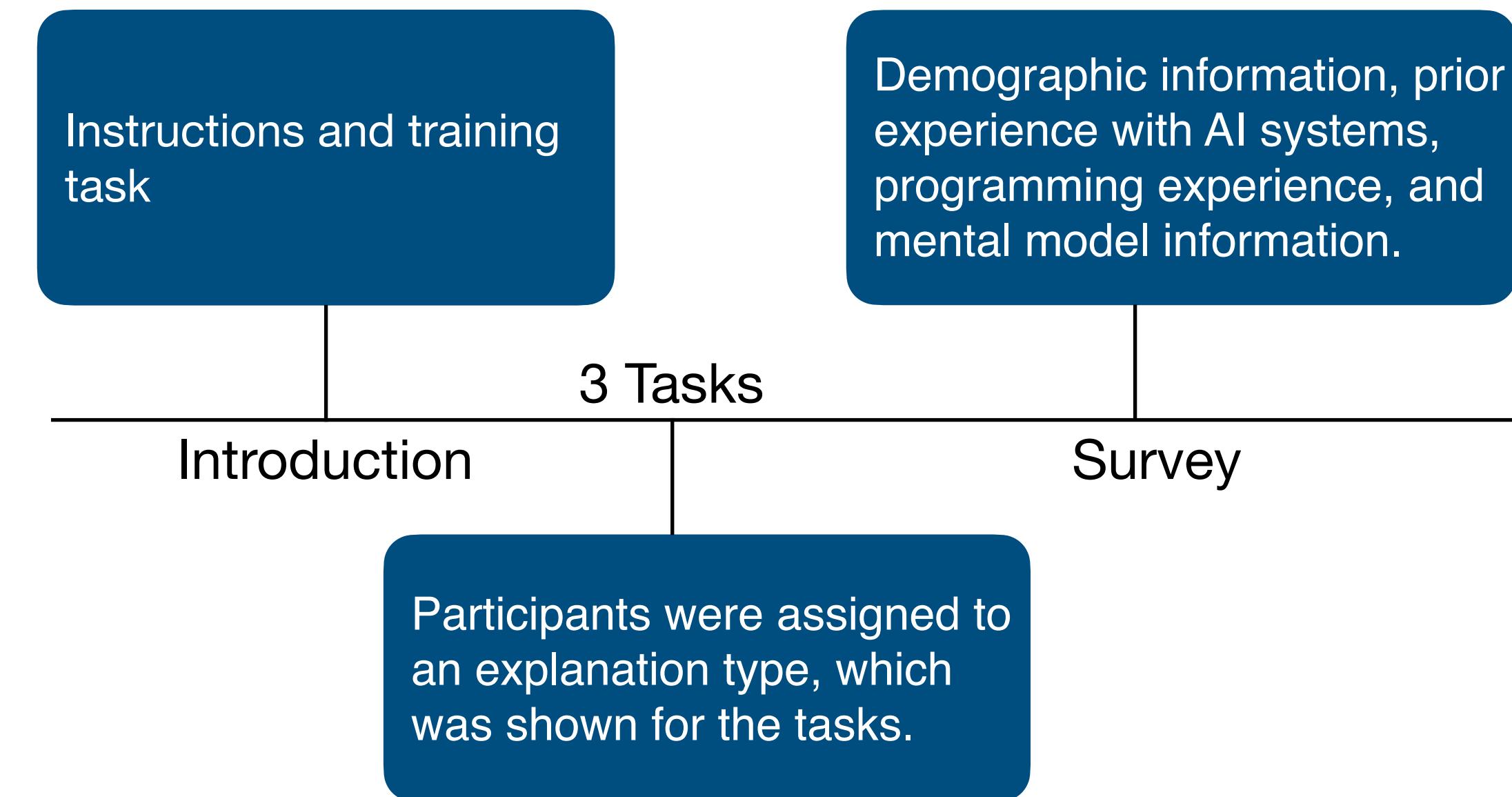
How did explanations impact overall correctness and efficiency?

How did participants use explanations?

How did explanations impact users' mental models of the system?

Methods & Protocol

- ◆ Comparative study with 7 conditions (baseline and 3 variations per main explanation style).
- ◆ Unmoderated task-based user study on Amazon Mechanical Turk with 252 participants.
- ◆ Tasks asked user to write a natural language sentence to generate an automation flow.
- ◆ Measures: correctness, efficiency, use of explanations, mental model



Explanation Styles

System map

Example: When there's a new message on Slack, create file sharing on Dropbox

when there's a new incident on servicenow, create an issue on jira and send mail on domino

Submit

Predicted flow based on your submitted input:

when there's a new incident on servicenow, create an issue on jira and send mail on domino

```
graph LR; A(servicenow) --> B(jira); B --> C(domino)
```

System top

Top 5 predictions based on your input sentence
Only the top 1 predictions are used to generate the flow

Connector 1 Predictions		
servicenow incident UPDATED	triggers	84.97% match
servicenow incident CREATED	triggers	83.11% match
servicenow problem UPDATED	triggers	69.49% match
servicenow problem CREATED	triggers	67.63% match
servicenow ticket UPDATED	triggers	67.39% match

Connector 2 Predictions		
jira Issue CREATE	actions	100.00% match
jira Project CREATE	actions	83.57% match
jira Issue RETRIEVEALL	actions	82.79% match
jira Issue DELETEALL	actions	80.46% match
jira Issue UPDATEALL	actions	78.44% match

Connector 3 Predictions		
domino MailMessage CREATE	actions	69.48% match
domino Document CREATE	actions	66.82% match
domino MailMessage RETRIEVEALL	actions	64.37% match
domino MailMessage DELETEALL	actions	63.95% match
domino MailMessage UPDATEALL	actions	62.86% match

Challenge: Because we were running a live prototype, we had to work within the technical constraints of the project, which limited what we were able to provide.

Explanation Styles

System map

Example: When there's a new message on Slack, create file sharing on Dropbox

when there's a new incident on servicenow, create an issue on jira and send mail on domino

Submit

Predicted flow based on your submitted input:

when there's a new incident on servicenow, create an issue on jira and send mail on domino

```

graph LR
    A(servicenow  
updated incident) --> B(jira  
create issue)
    B --> C(domino  
create mailmessage)
  
```

System top

Top 5 predictions based on your input sentence
Only the top 1 predictions are used to generate the flow

Connector 1 Predictions			Connector 2 Predictions			Connector 3 Predictions		
servicenow incident UPDATED	triggers	84.97% match	jira Issue CREATE	actions	100.00% match	domino MailMessage CREATE	actions	69.48% match
servicenow incident CREATED	triggers	83.11% match	jira Project CREATE	actions	83.57% match	domino Document CREATE	actions	66.82% match
servicenow problem UPDATED	triggers	69.49% match	jira Issue RETRIEVEALL	actions	82.79% match	domino MailMessage RETRIEVEALL	actions	64.37% match
servicenow problem CREATED	triggers	67.63% match	jira Issue DELETEALL	actions	80.46% match	domino MailMessage DELETEALL	actions	63.95% match
servicenow ticket UPDATED	triggers	67.39% match	jira Issue UPDATEALL	actions	78.44% match	domino MailMessage UPDATEALL	actions	62.86% match

Social remove

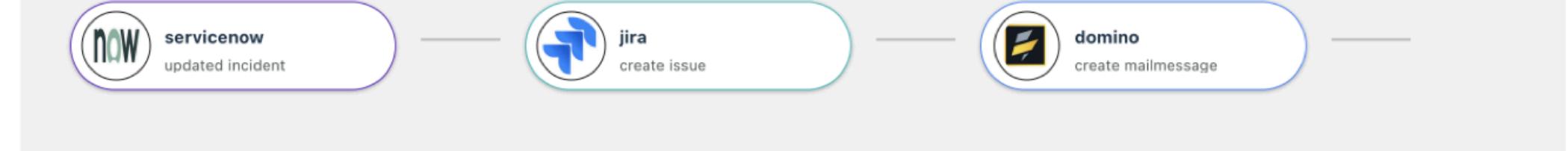
when there's a new incident on servicenow, create an issue on jira and send mail on domino

Submit

Consider modifying or removing the highlighted and underlined words
Based on other successful users with similar sentences

when there's a new incident on servicenow, create an issue on jira **and** send mail on domino

Predicted flow based on your submitted input:



Social add

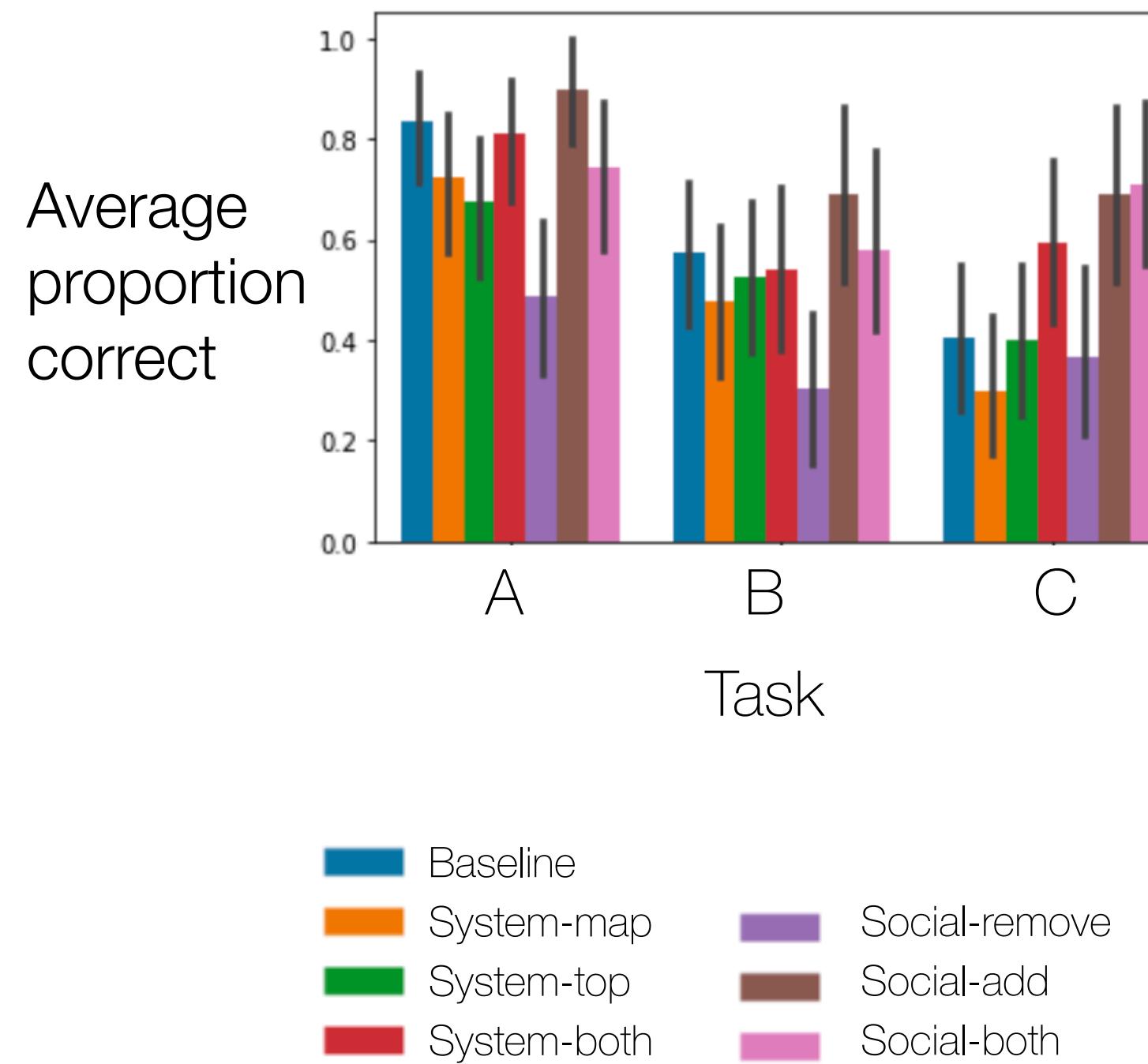
Suggested words & phrases you might want to add to your sentence
Based on other successful users with similar sentences

created		message		then	
Usage Examples	Number of Users	Usage Examples	Number of Users	Usage Examples	Number of Users
incident is created	31	a mail message	23	then create a	12
is created on	18	mail message in	17	jira then create	11
created on servicenow	17	message in domino	17	jira and then	9
created in servicenow	13	mail message on	11	on jira then	9
		message on domino	11	and then create	6

Challenge: Because we were running a live prototype, we had to work within the technical constraints of the project, which limited what we were able to provide.

Key Insights

Task Correctness



Participants who had social-add explanations had significantly higher task correctness than system-map ($p < .05$), system-top ($p < .05$), and social-remove ($p < .001$).

Impacts differed across tasks (C was the most difficult task), in which having the system information helped participants.

Other

During pilot testing, we discovered that providing an example input was highly effective in helping users understand what to type.

Impacts

Context: Design and technical simplicity were high priorities for the PMs and product designers.

The AppConnect team added explanations to the experimental feature, based on our findings and limited by design criteria and technical feasibility.

They also added an example input based on our recommendation.

Example: When there's a new message on Slack, create file sharing on Dropbox

when there's a new incident on servicenow, create an issue on jira and send mail on domino

Submit

Predicted flow based on your submitted input:

when there's a new incident on servicenow, create an issue on jira and send mail on domino



Top 5 predictions based on your input sentence
Only the top 1 predictions are used to generate the flow

Connector 1 Predictions		
servicenow incident UPDATED	triggers	84.97% match
servicenow incident CREATED	triggers	83.11% match
servicenow problem UPDATED	triggers	69.49% match
servicenow problem CREATED	triggers	67.63% match
servicenow ticket UPDATED	triggers	67.39% match

Connector 2 Predictions		
jira Issue CREATE	actions	100.00% match
jira Project CREATE	actions	83.57% match
jira Issue RETRIEVEALL	actions	82.79% match
jira Issue DELETEALL	actions	80.46% match
jira Issue UPDATEALL	actions	78.44% match

Connector 3 Predictions		
domino MailMessage CREATE	actions	69.48% match
domino Document CREATE	actions	66.82% match
domino MailMessage RETRIEVEALL	actions	64.37% match
domino MailMessage DELETEALL	actions	63.95% match
domino MailMessage UPDATEALL	actions	62.86% match

Michelle Brachman, PhD

UX Researcher with experience in human-centered AI, B2B, developer experience, and zero-to-one / innovation.

Thanks so much for your time!

[Email](#) [LinkedIn](#) [Webpage](#) [Phone](#)

