

# Combining Discourse Coherence with LLMs for More Robust and Equitable Task-Oriented Dialogue

Katherine Atwell, Mert Inan,  
Anthony Sicilia, Malihe Alikhani

# Effective collaboration requires understanding user goals



Goal: get assistance with making a chicken sandwich

I want to make a chicken sandwich.

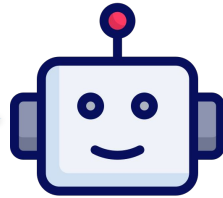


Umm...can you help me?



Perceived Goal: make small talk

Good idea! Chicken sandwiches are high in protein



# Effective collaboration requires understanding user goals



Goal: get assistance with making a chicken sandwich

I want to make a chicken sandwich.

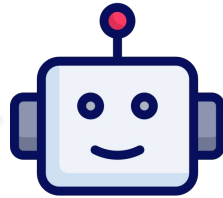


Yes!



Perceived Goal: get assistance with making a chicken sandwich

Would you like to hear this highly-rated chicken sandwich recipe?



Effective collaboration requires understanding user goals

→ But LLMs frequently get distracted by irrelevant context (Shi et al., 2023)

**Is there a way we can better use context to understand user goals?**

# Task Setup: Alexa Prize Taskbot Challenge

- **Goal:** Walk users through articles and recipes from online data sources to help them achieve their task
  - ◆ Provide suggestions, answer questions, navigate the task, etc.

The logo for wikiHow, featuring the word "wikiHow" in white text on a green rectangular background.The logo for Whole Foods Market, featuring the words "WHOLE FOODS MARKET" in white text on a dark green rectangular background.

External data sources for  
recipes and task instructions



Here are some recipes for pie...

Let's do the the first one

Great! Start by mixing in butter and...

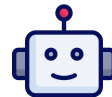
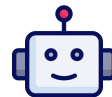
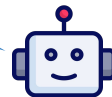


Okay, how much butter

1lb

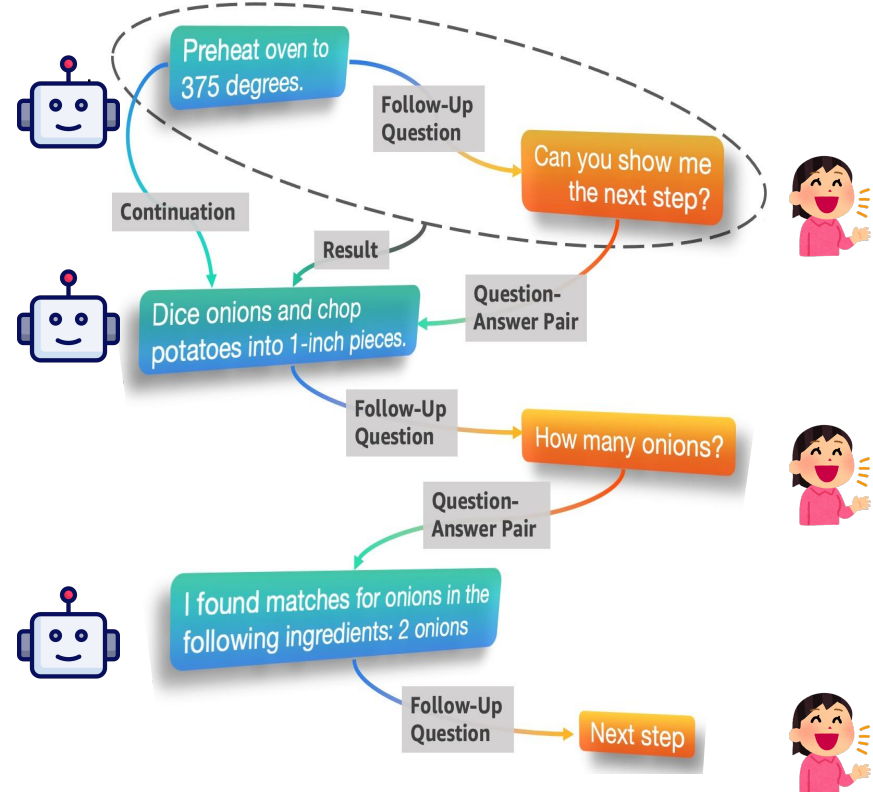


Okay, what's the next step



# Can Discourse Coherence Theories Help?

- Analyze dialogue beyond the sentence or utterance level
- Represent a dialogue as a series of interconnected units
- Provide theoretical framework for tracking context and goals within a dialogue



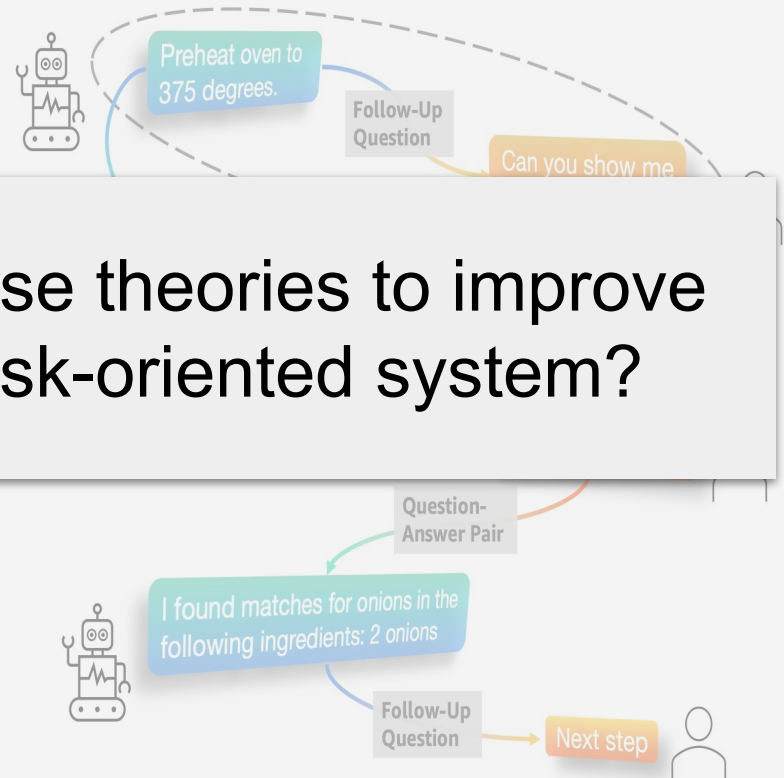
# Can Discourse Coherence Theories Help?

→ Provide theoretical framework for tracking context and goals within a

→ **How can we use discourse theories to improve dialogue quality in a task-oriented system?**

→

interconnected units

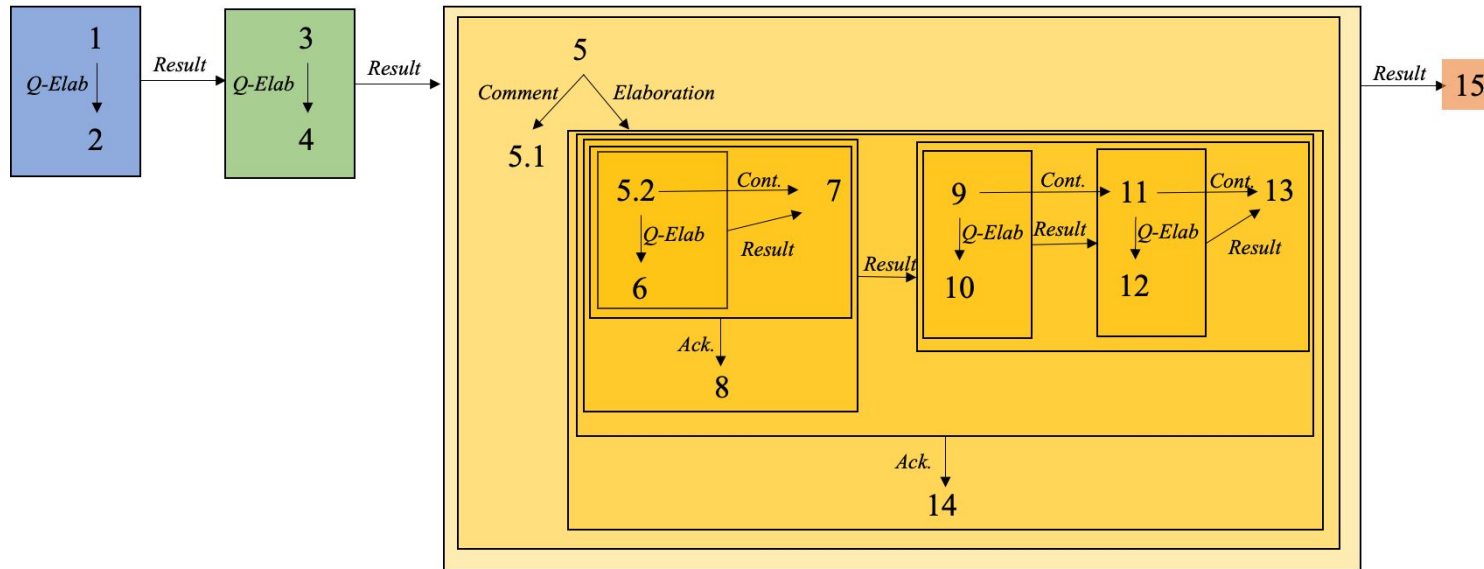


# To answer, we...

1. Conduct a preliminary discourse analysis of previous TaskBot dialogues
2. Model user flow, and use insights from this and our discourse analyses to track dialogue context
3. Use contextual information to improve downstream components
  - a. Intent Recognition
  - b. Repair & Clarification Questions
  - c. Natural Language Generation

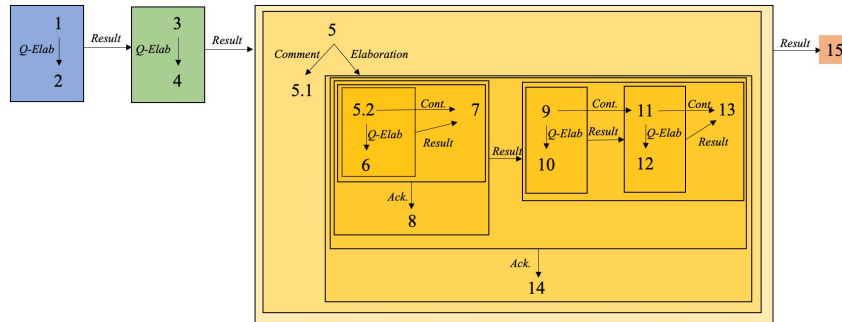
# Preliminary Discourse Analysis on TaskBot Dialogues

- We have annotated previous TaskBot dialogues using SDRT framework and STAC discourse relations
  - We created discourse graphs connecting utterances in a dialogue



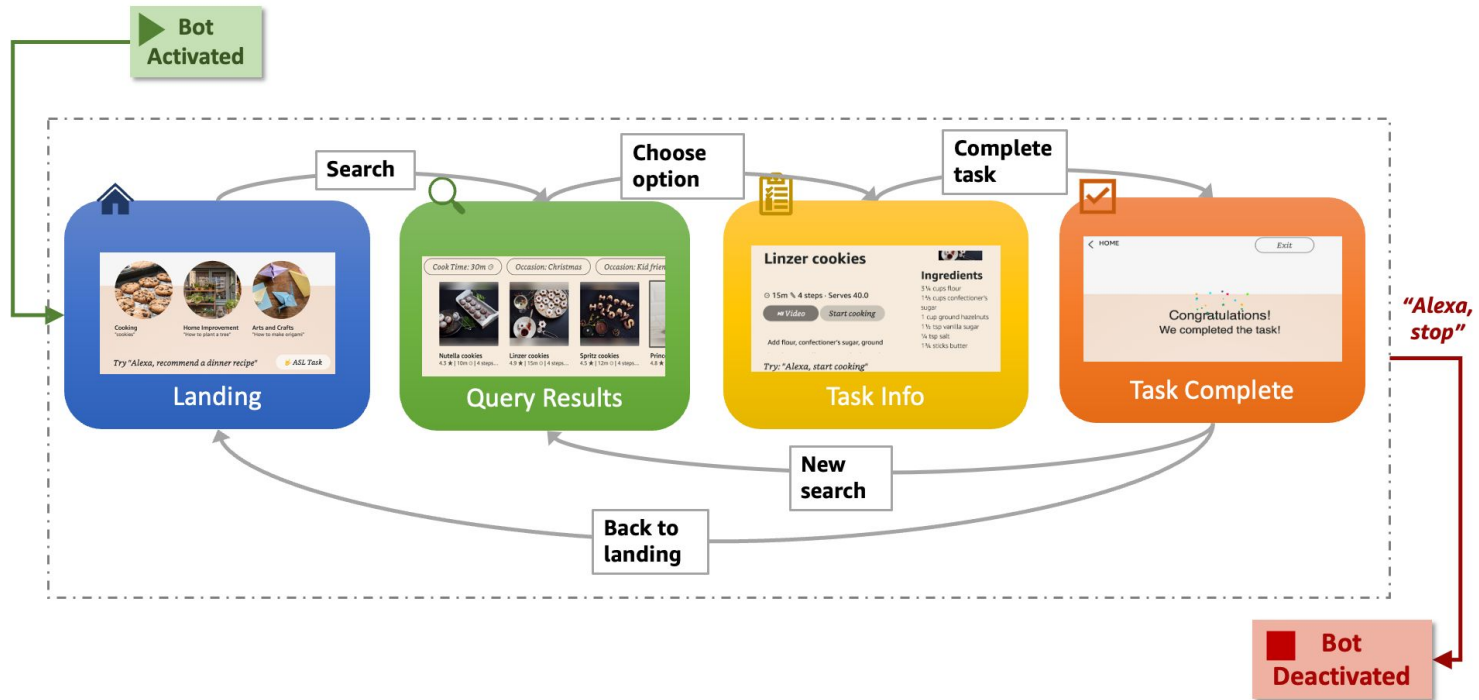
# Insights from These Preliminary Analyses

- ❖ Users can only make subordinating discourse moves, and cannot progress the dialogue forward themselves.
  - So valid user moves depend on the dialogue context
- ❖ We identify that each successful dialogue can be split up into **four distinct stages**
- ❖ At each stage, the set of discourse relations that can result in dialogue progression is constrained



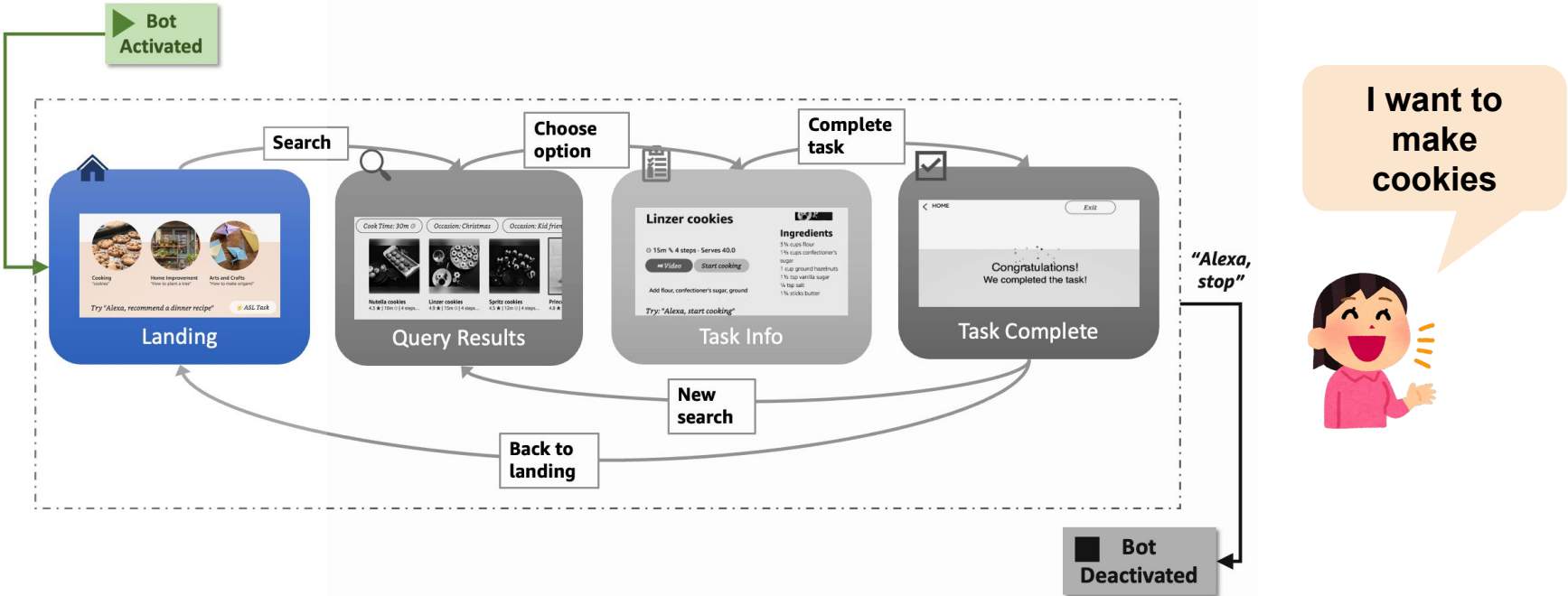
# Modeling User Flow Using Insights from Discourse

- Users progress through **four phases** to complete their task



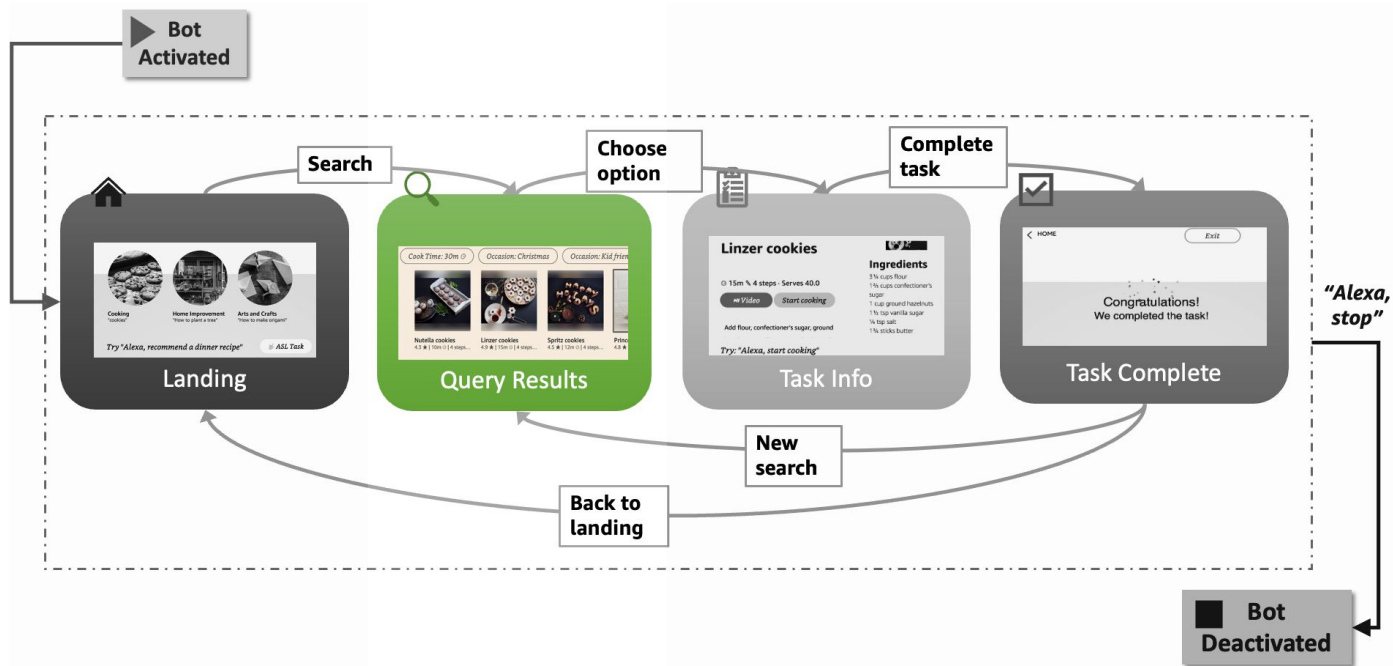
# Modeling User Flow Using Insights from Discourse

1. **Landing:** “home page” - users start here after activating the bot and can choose or request specific tasks



# Modeling User Flow Using Insights from Discourse

2. **Query Results:** recipes and articles that best match the user's search or selection are presented

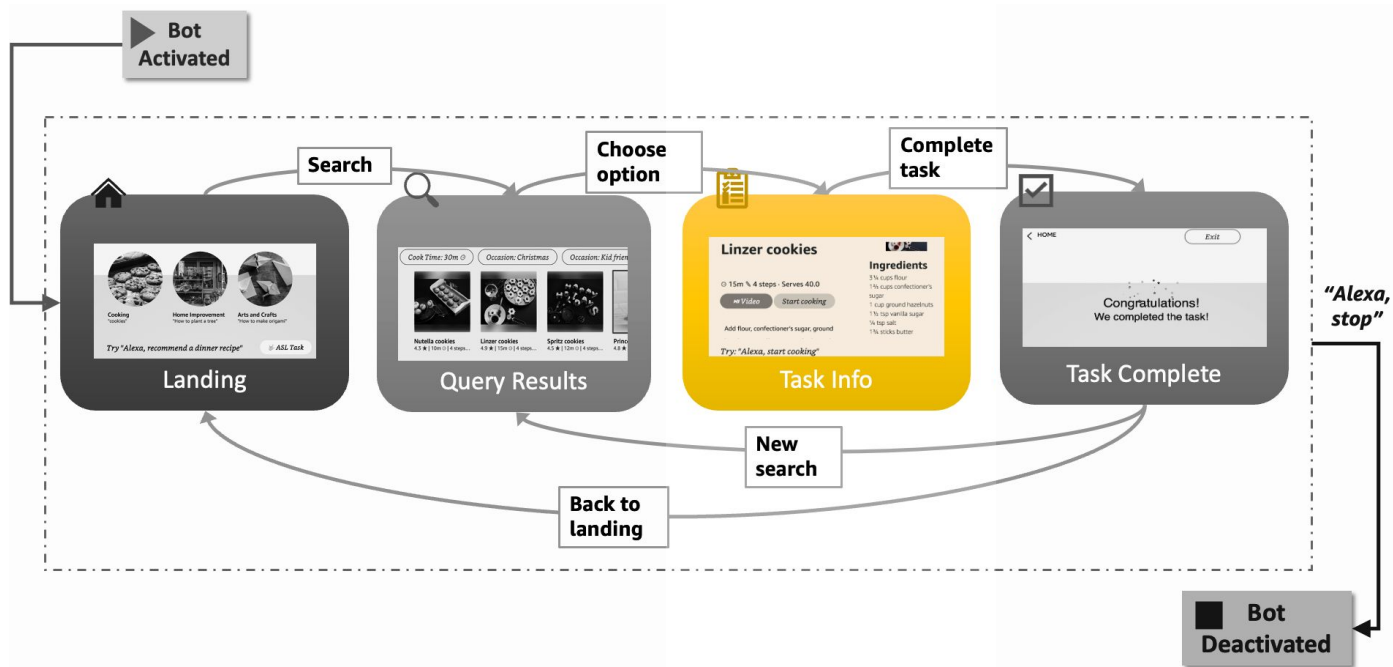


I'll do the second one



# Modeling User Flow Using Insights from Discourse

3. **Task Info:** information and instructions are provided for the selected task



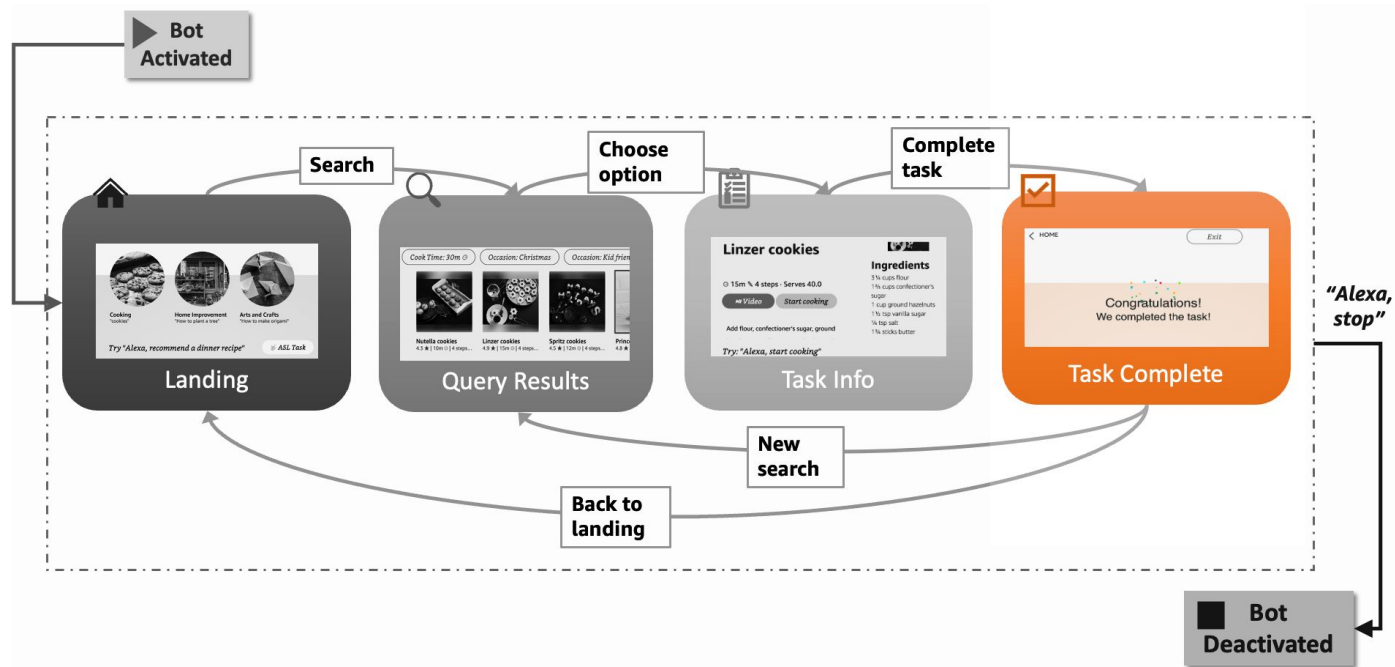
Okay, I'm ready for the next step



"Alexa, stop"

# Modeling User Flow Using Insights from Discourse

4. **Task Complete:** when the user finishes the task, they are congratulated and given the option to start another task.

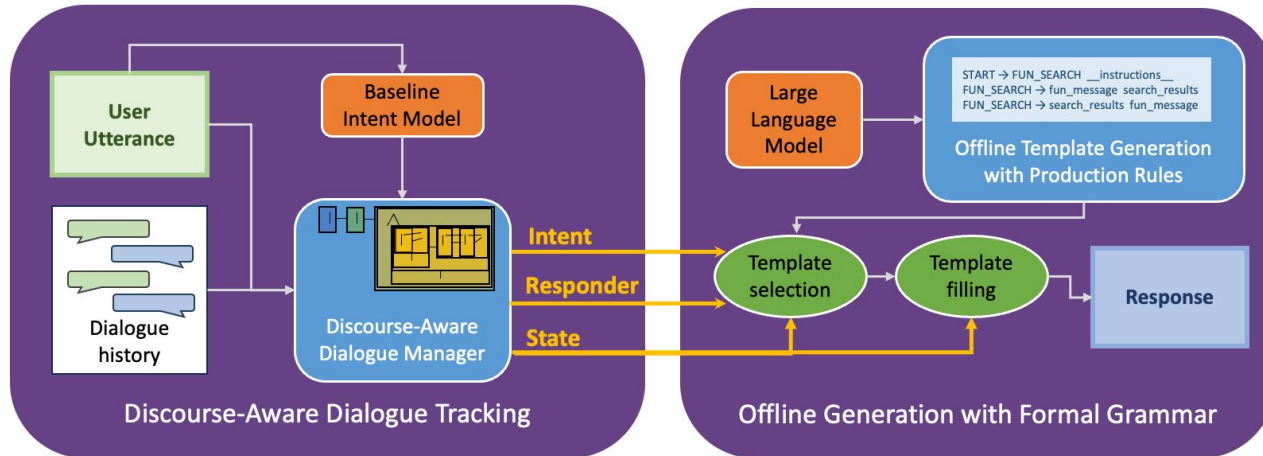


Okay, I'd like to stop

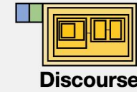


# Incorporating Discourse Into Our System

1. Conduct a discourse analysis of our system dialogues
2. Track dialogue context using insights from our analyses and from existing discourse theories
3. Use contextual information to improve downstream components



## Better Intent Recognition



“Next”  
“Continue”

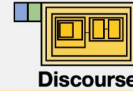


“Can I go to the next one?”  
“I did it!”

## Improved Clarification + Repair Strategies



“Show me recipes for chicken catty”

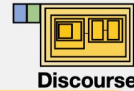


I’m sorry, I didn’t understand



Did you mean chicken patty?

## Context-aware Response Generation

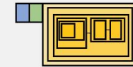


Here’s what I found in WikiHow:  
- How to Play Tennis,  
- How to Play a Tennis Tiebreaker, and  
- How to Play Ping Pong (Table Tennis).



Get ready to unleash your creativity!  
Check out these cool options I found on WikiHow for “play tennis”. The top results are:  
- How to Play Tennis,  
- How to Play a Tennis Tiebreaker, and  
- How to Play Ping Pong (Table Tennis).

## Better Intent Recognition



Discourse



“Next”  
“Continue”



“Can I go to the next one?”  
“I did it!”

## Improved Clarification + Repair Strategies



“Show me recipes for chicken catty”



Discourse



I’m sorry, I didn’t understand



Did you mean chicken patty?

## Context-aware Response Generation



LLM



Discourse



Here’s what I found in WikiHow:  
- How to Play Tennis,  
- How to Play a Tennis Tiebreaker, and  
- How to Play Ping Pong (Table Tennis).



Get ready to unleash your creativity!  
Check out these cool options I found on WikiHow for “play tennis”. The top results are:  
- How to Play Tennis,  
- How to Play a Tennis Tiebreaker, and  
- How to Play Ping Pong (Table Tennis).

# Improving Intent Recognition with Discourse







For each stage, there are a limited number of valid discourse moves the user can make to progress through the task (also dependent on the previous dialogue context)

- So, with an understanding of the relevant dialogue context, we can constrain the possible intent labels and slots
- We can also treat all invalid discourse moves in the same manner

# Method: Updating Intents

We can use these insights to update the intents classified by our baseline intent model using:

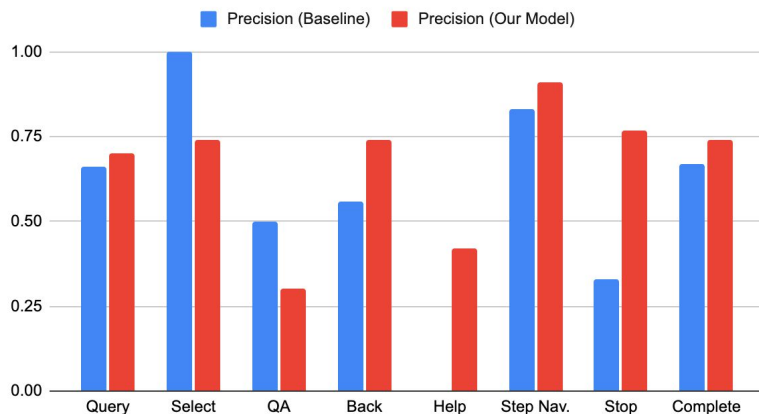
- Dialogue state information
- Knowledge of previous dialogue acts

User Utterance	Stage	Baseline Intent Label	Context-Aware Intent Label
<i>I'm thinking of blueberry muffins</i>	Opening	General 	Query 
<i>Option two</i>	Query Results	General 	Select 
<i>The last step</i>	Task Info	General 	Previous 

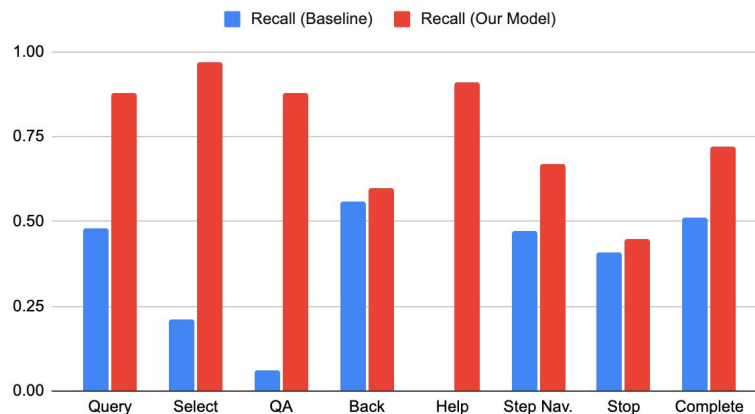
# Results: Discourse Improves Intent Recognition

- Our context-aware intent recognition system improved classification accuracy for our sampled user interactions

Intent Classification Precision



Intent Classification Recall



## Results: Discourse Helps Mitigate Bias

	Accuracy (Baseline)	Accuracy (Our Approach)
<b>want to</b>	100%	100%
<b>wanna</b>	7%	43%

**AAVE keyword “wanna” vs. “want to”**

## Results: Discourse Helps Mitigate Bias

	Accuracy (Baseline)	Accuracy (Our Approach)
<b>want to</b>	100%	100%
<b>wanna</b>	7%	43%

**93% gap**

AAVE keyword “wanna” vs. “want to”

# Results: Discourse Helps Mitigate Bias

	Accuracy (Baseline)	Accuracy (Our Approach)
want to	100%	100%
wanna	7%	43%

93% gap

57% gap

AAVE keyword “wanna” vs. “want to”

# Results: Mitigating Speech Recognition Biases

	Accuracy (Baseline)	Accuracy (Our Approach)
want to	100%	100%
wanna	7%	43%

93% gap

Performance gap reduced by 36%

57% gap

AAVE keyword “wanna” vs. “want to”

## Better Intent Recognition



“Next”  
“Continue”

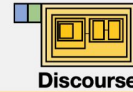


“Can I go to the next one?”  
“I did it!”

## Improved Clarification + Repair Strategies



“Show me recipes for chicken catty”



I’m sorry, I didn’t understand



Did you mean chicken patty?

## Context-aware Response Generation



Here’s what I found in WikiHow:  
- How to Play Tennis,  
- How to Play a Tennis Tiebreaker, and  
- How to Play Ping Pong (Table Tennis).



Get ready to unleash your creativity!  
Check out these cool options I found on WikiHow for “play tennis”. The top results are:  
- How to Play Tennis,  
- How to Play a Tennis Tiebreaker, and  
- How to Play Ping Pong (Table Tennis).

# Improving Clarification and Grounding with Discourse

- ❖ Simplifying our intent classification allows us to more easily identify invalid moves (**Help** intent)
- ❖ We respond to these moves in three different ways
  - Question tracking when the user gives an invalid answer to a question from the system
  - Clarification question when the system is unsure what the user wanted but has a guess
  - Help messages otherwise

# Examples: Clarification and Grounding Strategies

## Clarification Questions

These Whole Foods recipes are sure to impress your taste buds! Check out what I found for your search on chicken burger: The top results are Chicken Burgers, Moroccan Chicken Burgers, and Cheese and Apple Chicken Burgers. The first one has the highest rating, and for a quick recipe, check out the second one. Select an option, summarize it, or search for something else - it's all up to you!

Chicken birder

Hmm I didn't quite catch that. Did you mean Chicken Burgers?

Yes

Great! Before we get started...

## Question Tracking

In a large bowl add the Ground Chicken, Bread, Egg, Onion Powder, Ground White Pepper, Garlic Powder, Salt, Milk, Fresh Parsley, and Aged Goat Cheese. Mix together until well combined.

I want to make cookies

I couldn't find a quick answer for your question. Do you wanna start a new task?

Make cookies

Sorry, I don't quite understand your response. I asked if you would like to start a new task for cookies. Please respond with yes or no.

Yes

## Help Messages

In a large bowl add the Ground Chicken, Bread, Egg, Onion Powder, Ground White Pepper, Garlic Powder, Salt, Milk, Fresh Parsley, and Aged Goat Cheese. Mix together until well combined.

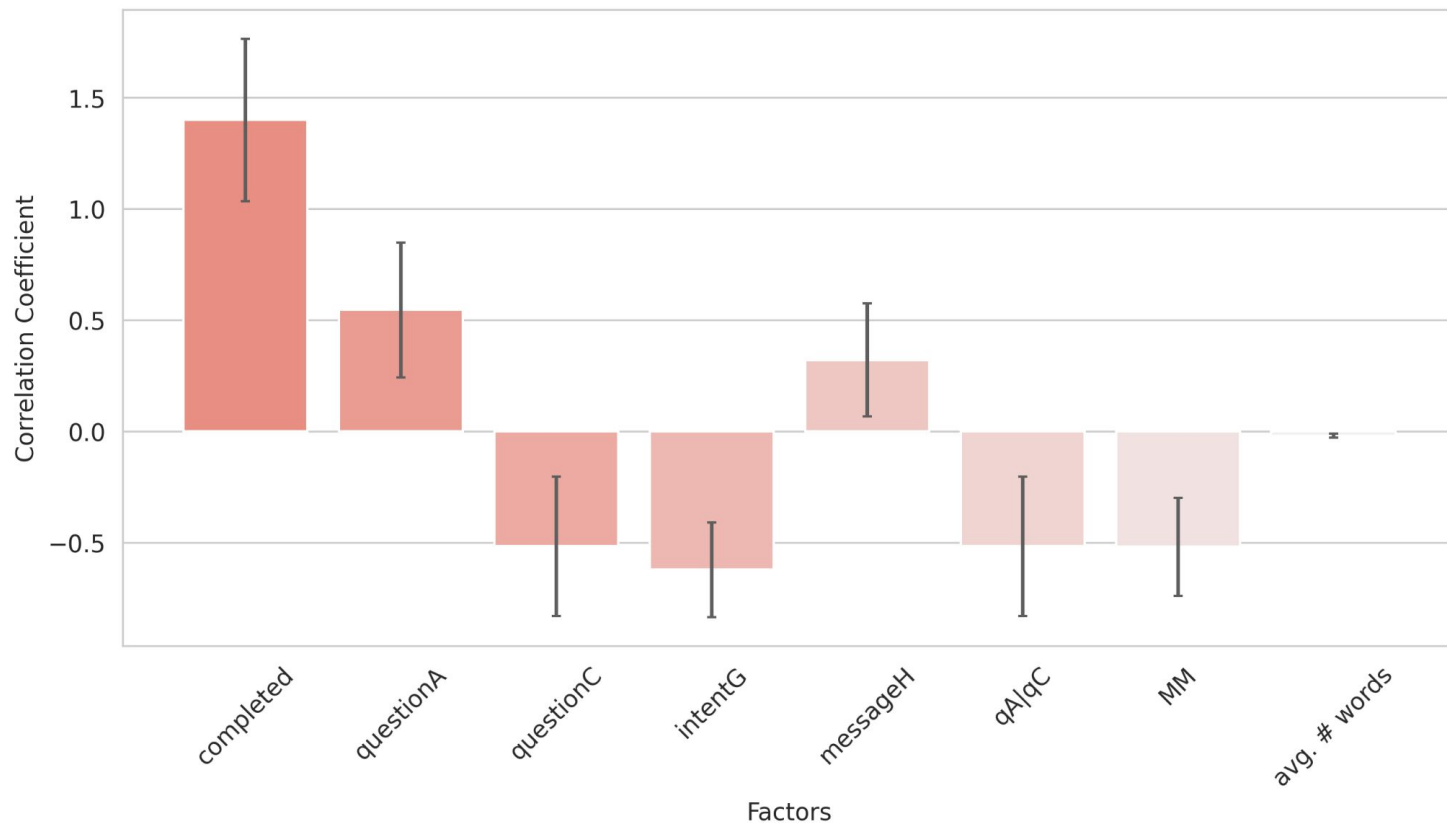
Blueberry muffins

Hmm, I'm having trouble understanding you. I was talking about the steps of our task. Try saying next

Next

Once combined add around half a cup of Breadcrumbs a little bit at a time to absorb the excess moisture.

# Results: Discourse Strategies Improve User Ratings



## Better Intent Recognition



“Next”  
“Continue”



“Can I go to the next one?”  
“I did it!”

## Improved Clarification + Repair Strategies



“Show me recipes for chicken catty”

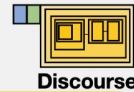


I’m sorry, I didn’t understand



Did you mean chicken patty?

## Context-aware Response Generation



Here’s what I found in WikiHow:  
- How to Play Tennis,  
- How to Play a Tennis Tiebreaker, and  
- How to Play Ping Pong (Table Tennis).



Get ready to unleash your creativity!  
Check out these cool options I found on WikiHow for “play tennis”. The top results are:  
- How to Play Tennis,  
- How to Play a Tennis Tiebreaker, and  
- How to Play Ping Pong (Table Tennis).

# Context-Aware Generation with LLMs

Our rule-based algorithms for response generation often led to **repetitive text**

→ Can we incorporate LLMs to reduce repetition and create more engaging, varied responses?

Problems with prompting LLMs on-the-fly:

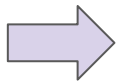
- Can generate incorrect, irrelevant, or unsafe responses
- Making API calls using dialogue information can compromise user privacy

**Solution: offline neurosymbolic generation using formal grammar**

# Method: Neurosymbolic Generation via Grammars

How can we combine generative AI and symbolic algorithms for more human-like responses?

Prompt ChatGPT  
using computational  
grammar to form  
*rules* for generation



Are you ready for some DIY fun? Check  
out what I found on WikiHow for  
`__input_search__`: `__list_options__`.  
Don't forget to `select` an option,  
`summarize` it, or `search something new`

ACC  
98%

Rules respect expected **inputs** and **outputs** of symbolic algorithm

# Results: Context-Aware Generation Performs Well in Human Eval

We conducted a **human evaluation** to test our approach

- 86% of generated utterances met all the required constraints
- 100% were as engaging or more engaging than the baseline utterances designed by humans
- Themes and partial utterances were repeated only 31% of the time

# Conclusion

- Using simple approaches based on discourse theories can improve the quality of dialogues in a task-oriented setting
- LLMs can be prompted to generate responses offline using a formal grammar
  - ◆ These responses can be used as templates to fill in with dialogue state information during the conversation itself
- Context-aware approaches can result in more equitable dialogue systems

# Deployment and System Success

We were able to deploy our system to thousands of users to measure its overall success

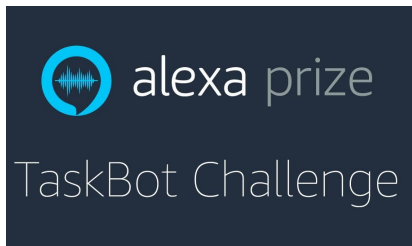
Final result: our system won third place in the competition!

★ Highest-placing team that had not previously competed



**KHOURY INCLUSIVE AI TEAM PLACES THIRD  
IN AMAZON'S ALEXA PRIZE TASKBOT  
CHALLENGE**




**MON 01.29.24** / BENJAMIN HOSKING



# Thank you!






**Katherine Atwell**

 katherine-atwell  
 @katwellll  
 <https://katherine-atwell.github.io/>






**Mert Inan**

 merterm  
 @merterm  
 <https://merterm.github.io/>



**Anthony Sicilia**

 anthonsicilia  
 @asicil96  
 <https://anthonsicilia.tech/>



**Malihe Alikhani**

 malihealikhani  
 @malihealikhani  
 <https://www.malihealikhani.com/>