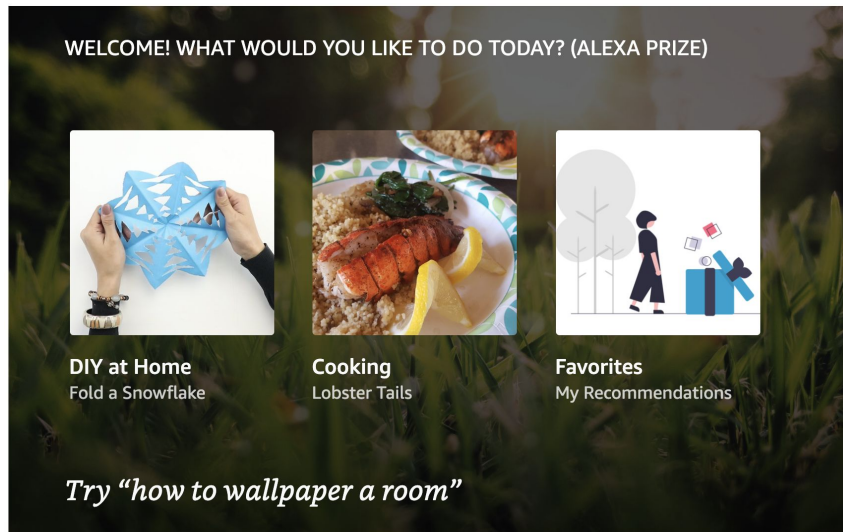


Bootstrapping a User-Centered Task-Oriented Dialogue System



Huan Sun
Associate Professor
The Ohio State University

<https://sunlab-osu.github.io/tacobot/>



Background

- ❖ First Alexa Prize TaskBot Challenge (06/2021 - 06/2022)
- ❖ 10 participant teams selected across the world (out of 125 initiated applications)
- ❖ 5 finalists
 - Top-3 performers
 - GRILLBot: University of Glasgow
 - Twiz: Nova School of Science and Technology
 - TacoBot: The Ohio State University
 - The other two finalists
 - QuakerBot: UPenn
 - GauchosBot: UC Santa Barbara

Team TacoBot



Left to right, top row: Zhen Wang and Samuel Stevens; second row: Tianshu Zhang, **Shijie Chen** (student co-lead), and Yu Su (faculty co-advisor); third row: Lingbo Mo, Xiang Yue, Xiang Deng; and bottom row: Ashley Lewis, **Ziru “Ron” Chen** (student co-lead, **undergraduate**), and Huan Sun (faculty advisor).

Goal

🤔 Alexa, help me fix a scratch on my car

Sure, I am happy to help. I learned how to do this on WikiHow.com Scratches are easy to remove. Ready to get started?

😊 Sure!

Ok, Hose down the area to rinse off dirt from the surface. Take your hose and aim a concentrated jet directly onto the scratches.

🤔 Won't it make the scratches worse?

Don't worry, spraying the fine scratches won't make them worse or damage your paint



😊 Ok, done. What next?

Scrub the scratched area using soapy water and a sponge.

😊 I did that, and rinsed off the soap. Easy!

That's great! Now Choose a rubbing compound and polish designed for cars.

🤔 Oh no! I don't have a rubbing compound. Is there something else I could use?

That's OK. I found this on a Website eurotechcarcare.com "Old-fashioned shoe polish is a great homemade recipe for fixing scratches in your car's exterior." Would like to use it or try something else?



Why do we care?

- ❖ A multimodal collaborative & accessible assistant to provide useful guidance in daily life



Why do we care?

- ❖ A multimodal collaborative & accessible assistant to provide useful guidance in daily life
- ❖ A rich problem space to study AI, e.g.,
 - Natural language understanding (with noisy user input, e.g., rambling, ASR error)
 - Question answering
 - Dialogue management
 - User engagement

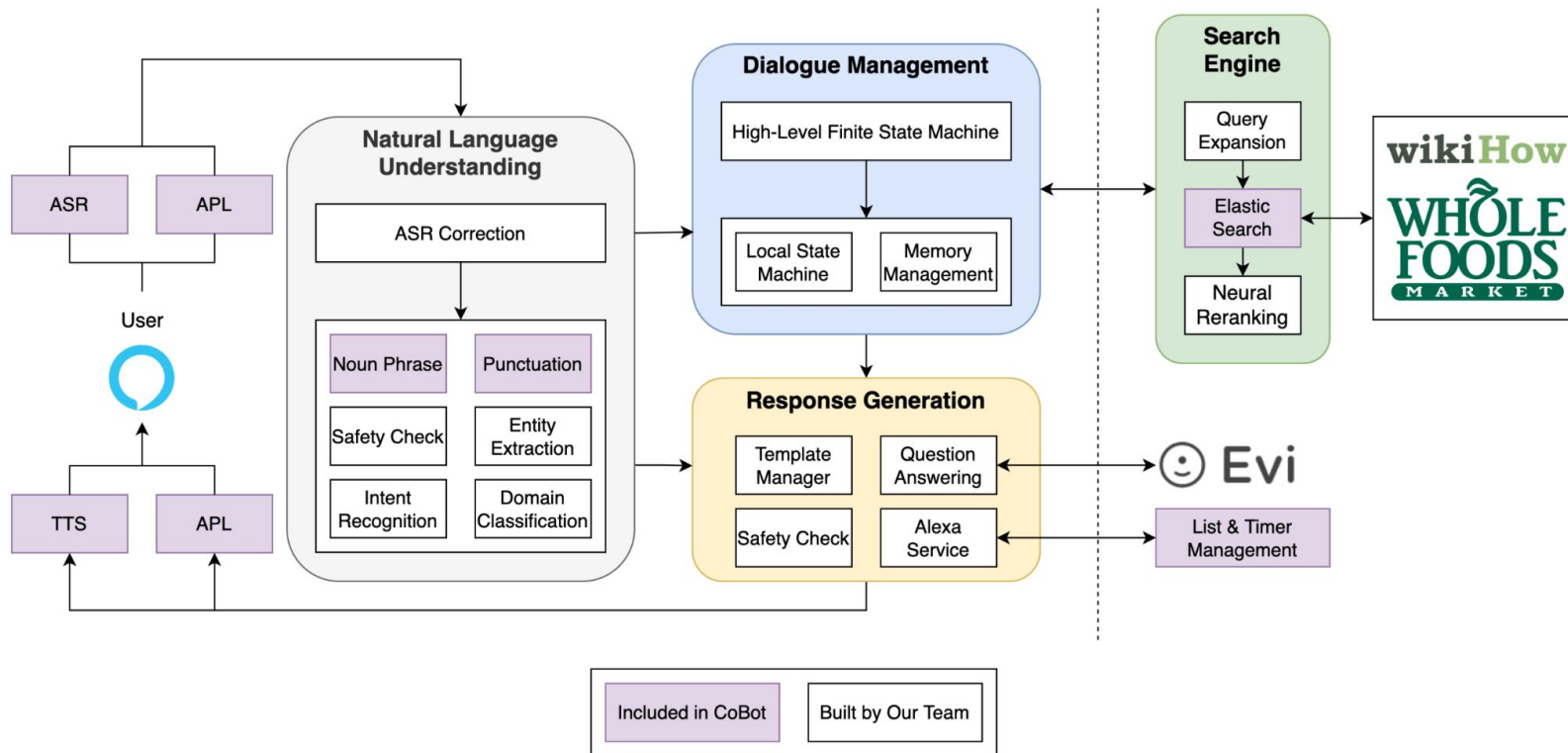


Why do we care?

- ❖ A multimodal collaborative & accessible assistant to provide useful guidance in daily life
- ❖ A rich problem space to study AI, e.g.,
 - Natural language understanding (with noisy user input, e.g., rambling, ASR error)
 - Question answering
 - Dialogue management
 - User engagement
- ❖ Possible to be embedded in physical robots in the future to execute the tasks for humans



TacoBot System Design (2021-2022)



Natural Language Understanding (NLU)

- ❖ Automatic Speech Recognition (ASR) Error Correction
- ❖ Intent Recognition
- ❖ Task Name Extraction
- ❖ Task Domain Classification

Natural Language Understanding (NLU)

❖ Automatic Speech Recognition (ASR) Error Correction

- **ASR error: A big source of user frustration**
- Current error corrector: rule-based, very limited
- Ongoing: contextualized error correction

User: “blueberry crepes” / “three bean salad”

Bot: blueberry creeps / three been salad

User: “sand art”

Bot: send art

User: “go to step four”

Bot: go to step for / go to stop for



Natural Language Understanding (NLU)

- ❖ Intent Recognition
- ❖ Task Name Extraction
- ❖ Task Domain Classification

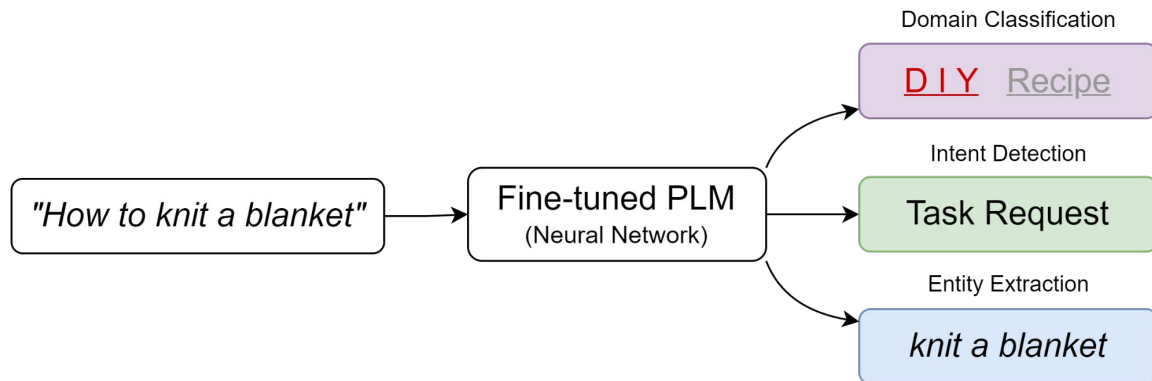
Baseline: Alexa Skill Kit (ASK) Interaction Model

- Rule based black box with poor generalization
 - User: “I’m done”
 - ASK: Self Intro, First Name = “Done”
 - User: “How to decorate for valentine’s day”
 - ASK: Set Timer

Natural Language Understanding (NLU)

- ❖ Intent Recognition
- ❖ Task Name Extraction
- ❖ Task Domain Classification

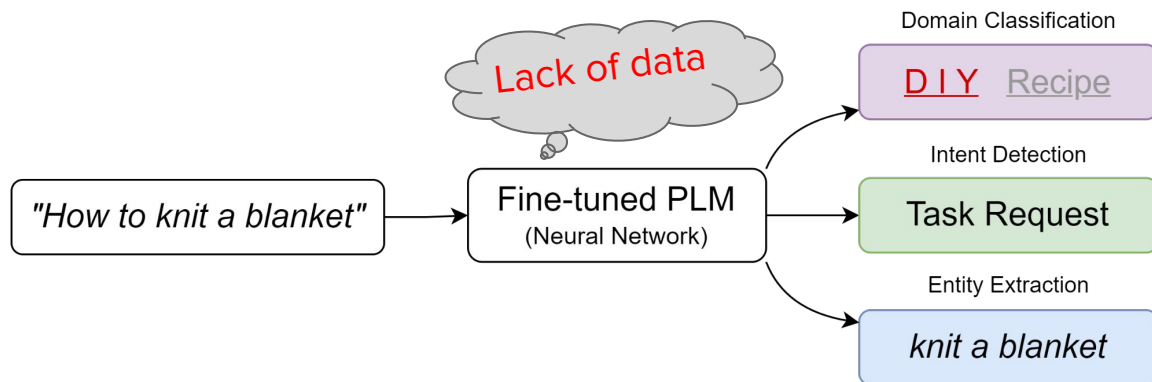
Desideratum:



Natural Language Understanding (NLU)

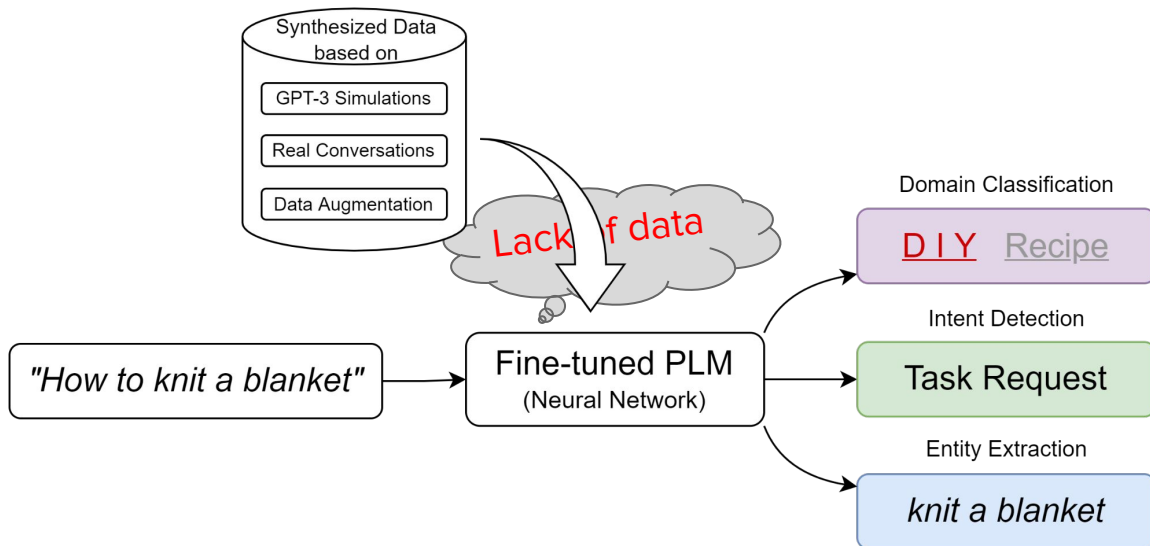
- ❖ Intent Recognition
- ❖ Task Name Extraction
- ❖ Task Domain Classification

Desideratum:



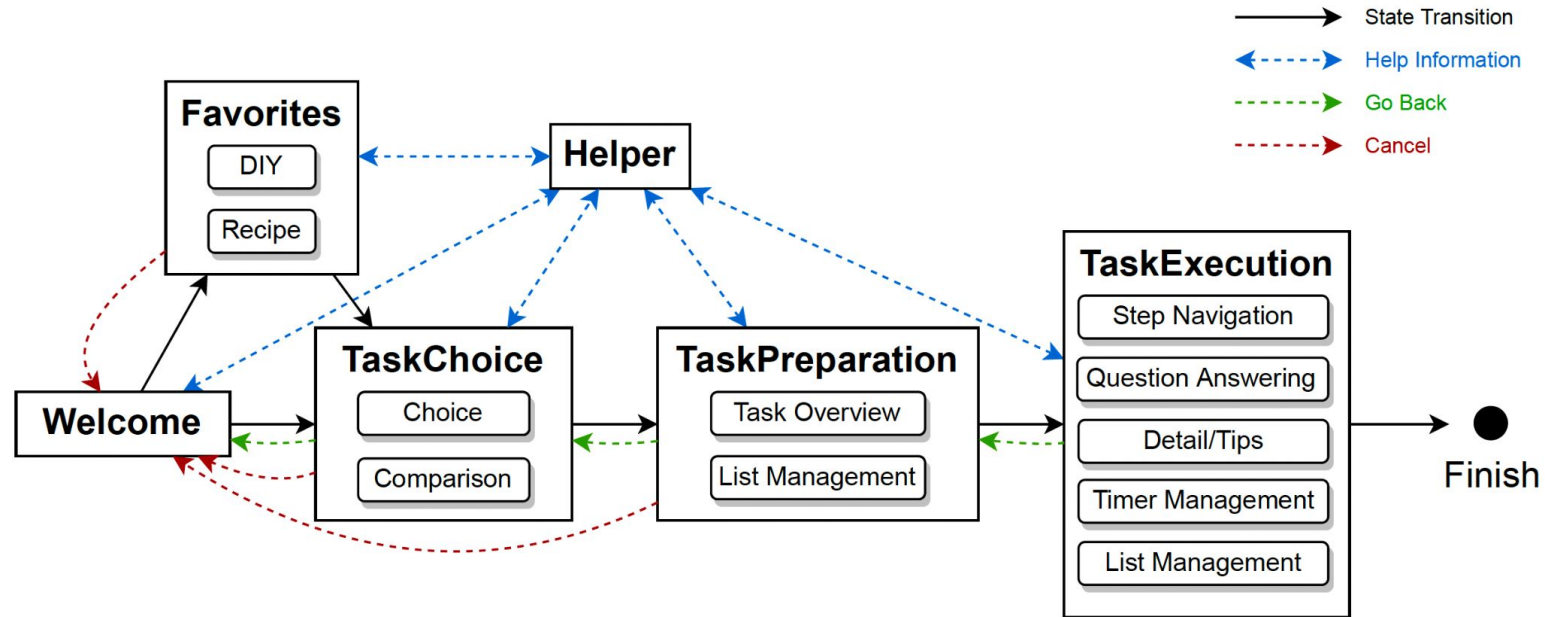
Robust Neural NLU

- ❖ Data synthesis
- ❖ Data augmentation
- ❖ Real conversations



Flexible Dialogue Management

A finite state machine



(State transition graph)

Search Engine

Example Query: “Fix scratches on the car”

Original Elastic Search Result

- 1.How to Fix Scratches on Hardwood Floors
- 2.How to Fix Scratches on an Aquarium
- 3.How to Fix the Alignment on a Car

+ Query Expansion

- 1.How to Fix Cat Scratches in Leather
- 2.How to Fix Stainless Steel Scratches with Scratch-B-Gone
- 3.How to Touch up Scratches on Your Car

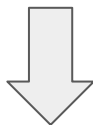
+ Neural Reranking

- 1.How to Remove Scratches from a Car
- 2.How to Repair a Deep Scratch on Car
- 3.How to Protect Cars from Scratches

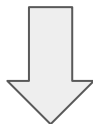
Search Engine

Bootstrapping a Neural Reranker

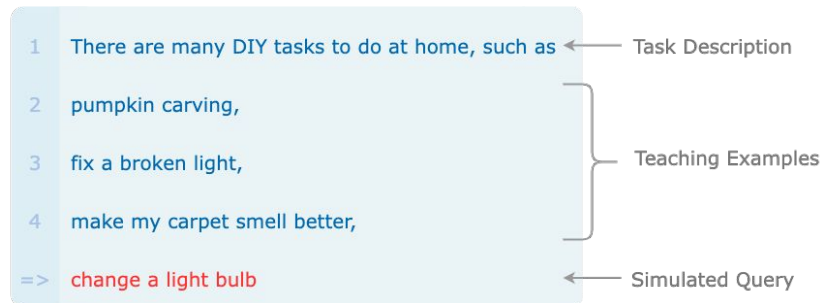
Simulate User Queries with GPT-3



Collect Weak Supervision with Google
Search as Oracle



Fine-tune a BERT Model



Query Simulation by GPT-3



Example Training Data Pair

How to make search & recommendation better?

❖ **Conversational search** (a distinctive feature implemented in the last phase)

User: “Let’s cook king crab legs”

Bot: “Sure! king crab legs sounds yummy. Do you have any diet constraints, such as gluten-free, low carb?”

User: “No” / “gluten-free” / ...



Pinpoint user preferences

How to make search & recommendation better?

❖ Conversational search

User: “Let’s cook king crab legs”

Bot: “Sure! king crab legs sounds yummy. Do you have any diet constraints, such as gluten-free, low carb?”

User: “No” / “gluten-free” / ...



Pinpoint user preferences

Challenges we addressed:

(1) Ask a lot, find few

User: “Let’s cook king crab legs”

Bot: “...Do you have any diet constraints, such as gluten-free, low carb?”

User: “...”

Bot: “...Which cuisine do you like?”

User: “...”

Bot: “I cannot find any recipe...”

(2) Expect more complex user queries

User: “Find recipes for king crab legs that are gluten free and low carb”

Bot: “Do you have any diet constraints, such as gluten-free, low carb?”

Question Answering

User:



Bot: “add a 14-ounce can of sweetened condensed milk, unsweetened natural cocoa powder... use a whisk to blend the ingredients until they’re completely mixed, and set the bowl aside.”

What questions can the user ask next?

Question Answering

User:

Bot: “add a 14-ounce can of sweetened condensed milk, unsweetened natural cocoa powder... use a whisk to blend the ingredients until they’re completely mixed, and set the bowl aside.”

What questions can the user ask next?



Question Type	Example
MRC	Use what tool to blend?
Factual	How much is 14 ounce in gram?
FAQ	How should I know the ingredients are completely mixed?
Substitute	I don't have condensed milk, can I use something else?
Ingredient	How much cocoa powder do I need?

Question Answering

User:

Bot: “add a 14-ounce can of sweetened condensed milk, unsweetened natural cocoa powder... use a whisk to blend the ingredients until they’re completely mixed, and set the bowl aside.”

What questions can the user ask next?



Question Type	Example
MRC	Use what tool to blend?
Factual	How much is 14 ounce in gram?
FAQ	How should I know the ingredients are completely mixed?
Substitute	I don't have condensed milk, can I use something else?
Ingredient	How much cocoa powder do I need?

Our solution: Question type classifier (94% acc on 500 test questions)
+ QA module for each type

Question Answering

User:



Bot: “add a 14-ounce can of sweetened condensed milk, unsweetened natural cocoa powder... use a whisk to blend the ingredients until they’re completely mixed, and set the bowl aside.”

What questions can the user ask next?

Question Type	Example
MRC	Use what tool to blend?
Factual	How much is 14 ounce in gram?
FAQ	How should I know the ingredients are completely mixed?
Substitute	I don’t have condensed milk, can I use something else?
Ingredient	How much cocoa powder do I need?

Our solution: Factual -> EVI (provided by Amazon)

FAQ -> FAQ dataset (collected by us)

Substitute -> ingredient substitute KB (constructed by us)

Question Answering

Challenges

- ❖ What are the comprehensive knowledge sources for such Qs? Knowledge graphs / Web corpus / tables?

How to effectively retrieve?

Or delegate to GPT-3?

- ❖ Lack of realistic questions for model development
- ❖ Reasoning (commonsense, multi-modal & heterogeneous data, etc.) for more complex questions

Question Type	Example
MRC	Use what tool to blend?
Factual	How much is 14 ounce in gram?
FAQ	How should I know the ingredients are completely mixed?
Substitute	I don't have condensed milk, can I use something else?
Ingredient	How much cocoa powder do I need?

Question Answering

Challenges

- ❖ What are the comprehensive knowledge sources for such Qs? Knowledge graphs / Web corpus / tables?

How to effectively retrieve?

Or delegate to GPT-3?

- ❖ Lack of realistic questions for model development
- ❖ Reasoning (commonsense, multi-modal & heterogeneous data, etc.) for more complex questions

Question Type	Example
MRC	Use what tool to blend?
Factual	How much is 14 ounce in gram?
FAQ	How should I know the ingredients are completely mixed?
Substitute	I don't have condensed milk, can I use something else?
Ingredient	How much cocoa powder do I need?

Side note:

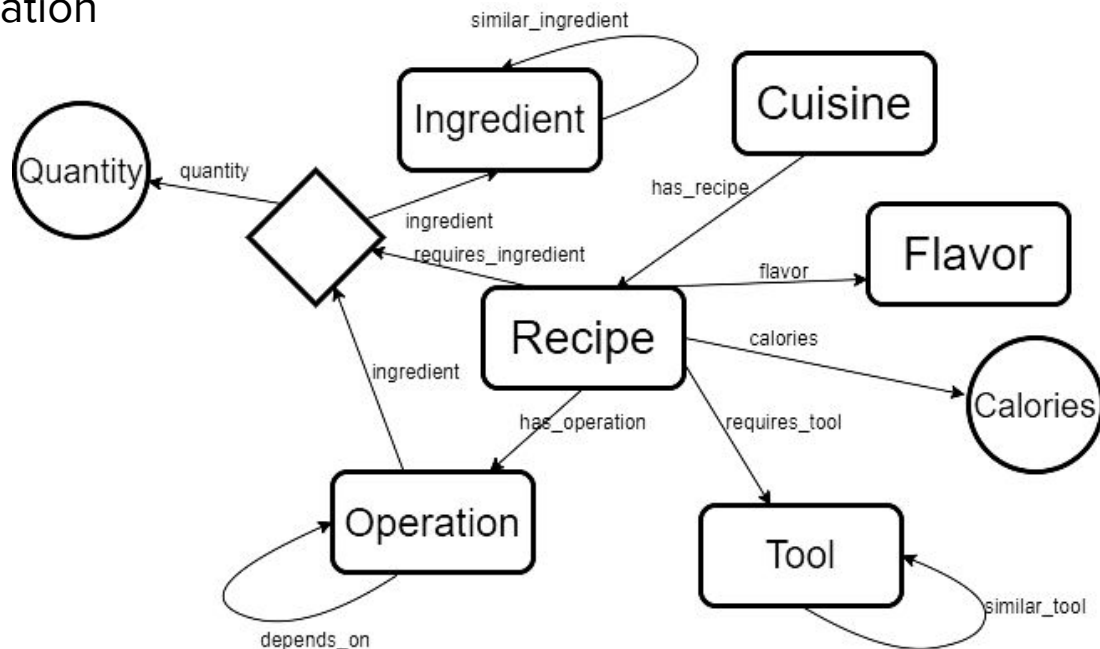
1. Poor QA capability -> few users ask questions -> hard to collect real questions -> hard to improve QA (vicious cycle)
2. QA is important for TaskBot user experience

Concluding Remarks

How can graphs help TaskBots?

Summary: How graphs can help TaskBots?

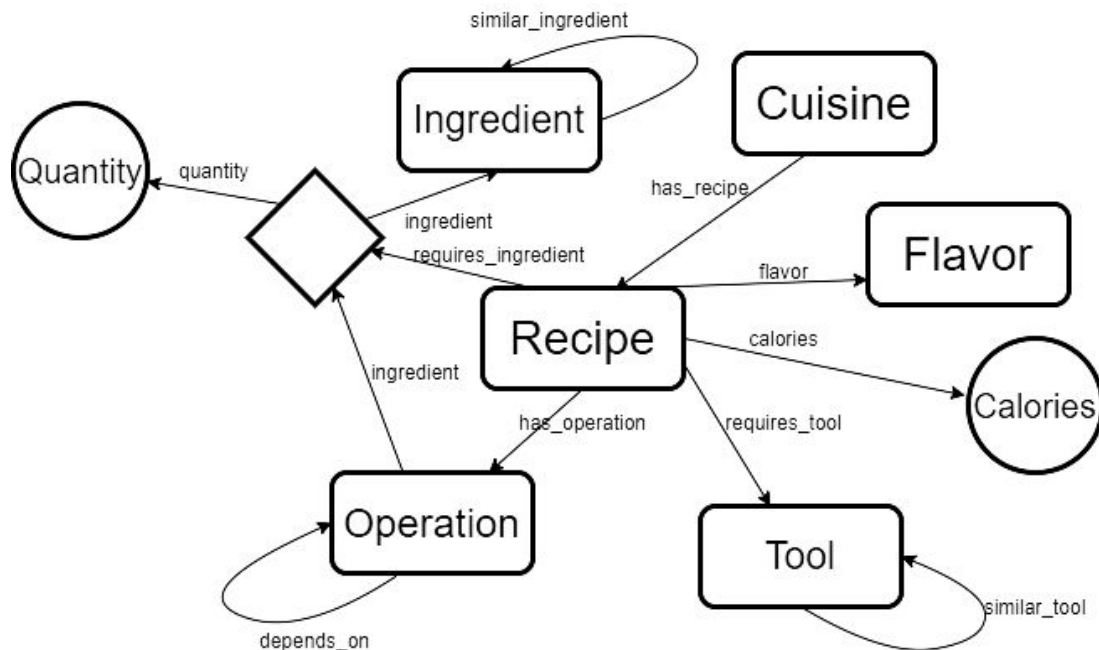
- ❖ Knowledge graph construction to make recipes/task instructions structured
 - QA
 - Search & recommendation
 - Create novel task execution plans



Ontology example

How can graphs make search & recommendation better?

- ❖ Knowledge graph of recipes/tasks for **constrained search**

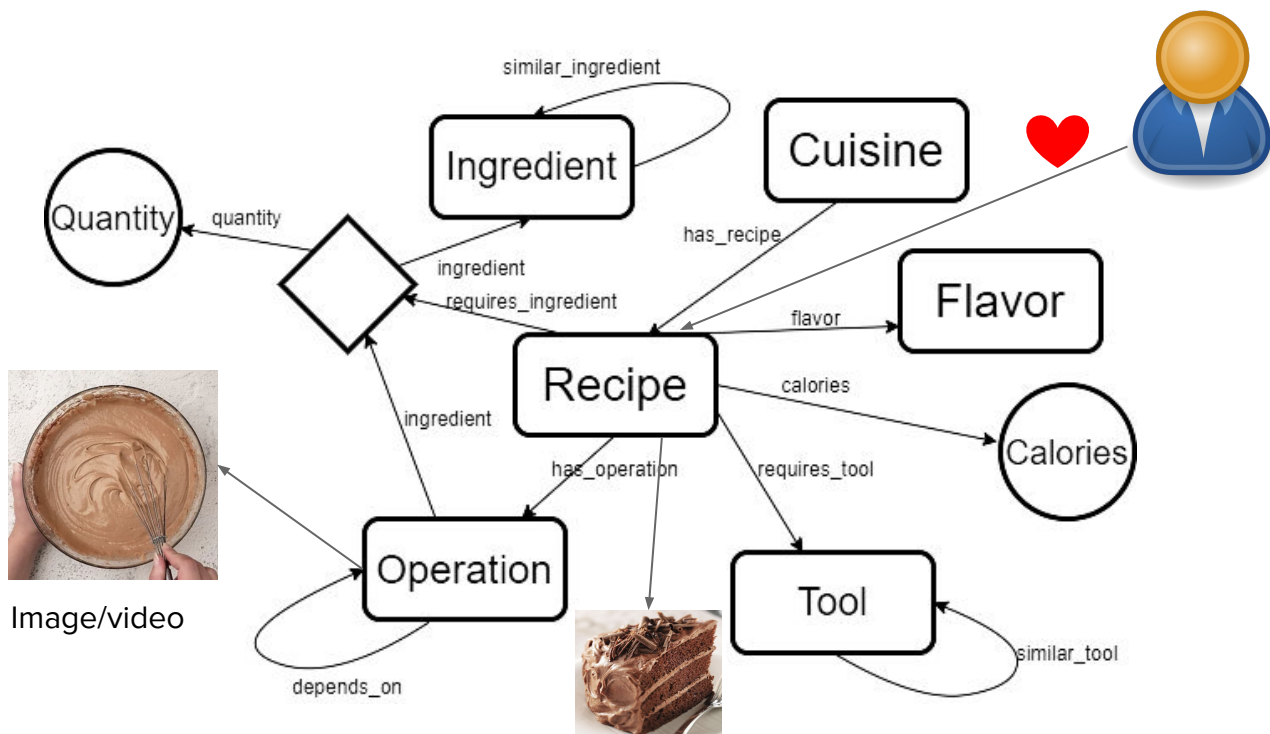


User: "I have tomatoes, pork, and beef and only one hour to cook. What do you recommend?"

Bot: "How about stir-fried beef with tomatoes?" (recommendation based on user constraints)

How can graphs make search & recommendation better?

- ❖ Construct a **heterogeneous multi-modal** graph for recommendation



Summary: How can graphs help TaskBots?

❖ Reasoning for exception handling

- Understand the exception
 - Missing ingredient/tool?
 - Certain steps already done?
 - User request to do things in parallel?
- Construct a dependency graph for each task
- Adjust the step order or revise relevant steps

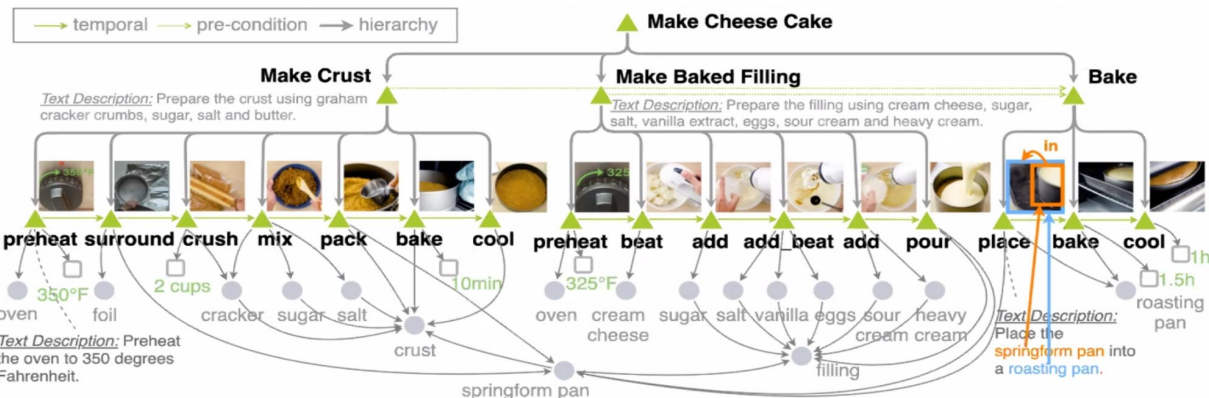


Figure source: Invited talk by Prof. Heng Ji at NAACL'22 SUKI workshop



Try “Alexa, Assist me” on your Echo show or Alexa app!
(One of the top-3 bots will help you)

Check out <https://sunlab-osu.github.io/tacobot/> for more details!