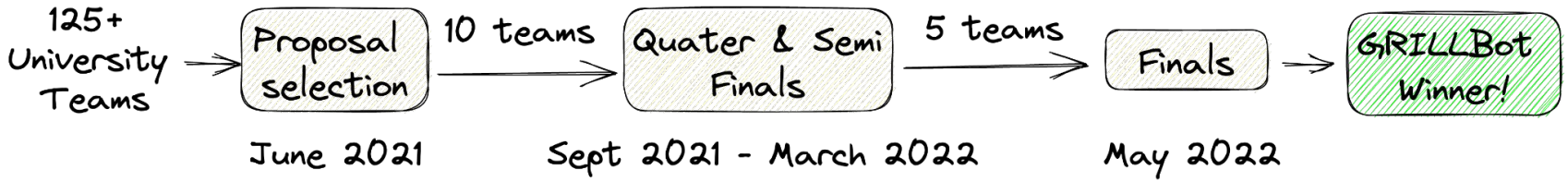
University of Glasgow

GRILLBot: An assistant for solving real-world tasks

Carlos Gemmell
Twitter: @carlos_gemmell
Date: Nov 29th 2022

1st Place in Alexa Prize TaskBot Competition



Iain Mackie - Search, Data acquisition

Paul Owoicho - Dialogue and system initiative

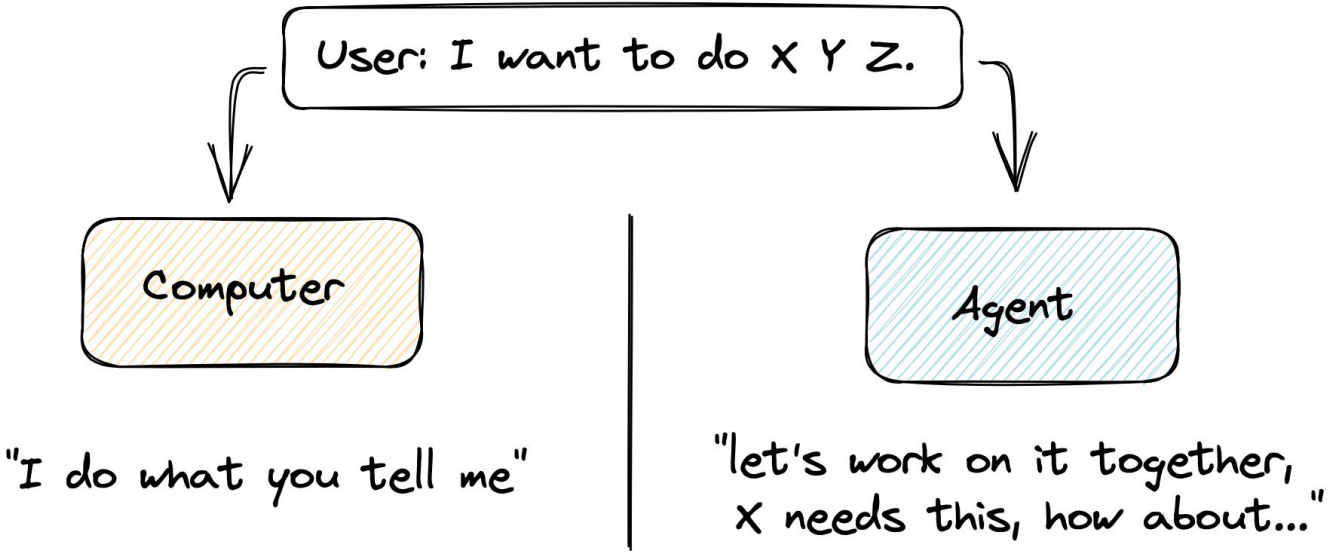
Federico Rossetto - Task Representation

Sophie Fisher - Multi-modal UI

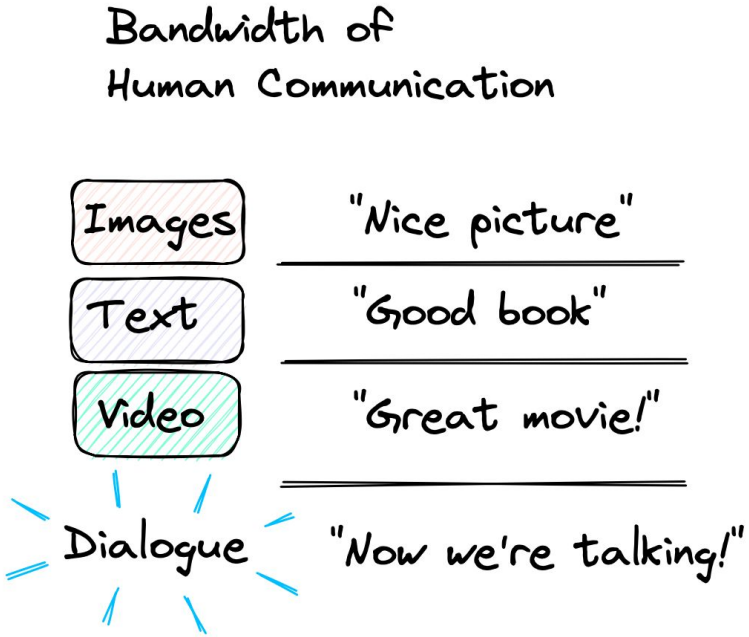
Carlos Gemmell - [Lead] Neural models

Jeff Dalton - [Faculty Advisor]

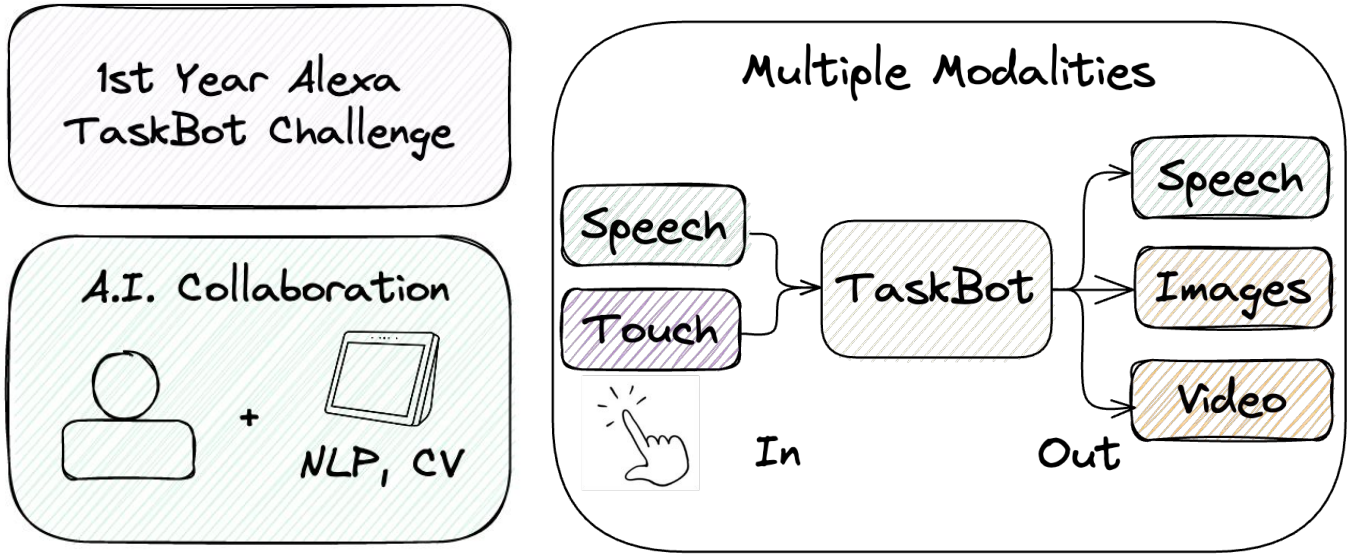
Towards Deeper Human-Machine Collaboration



Vision: Communication Bandwidth



Alexa TaskBot Challenge



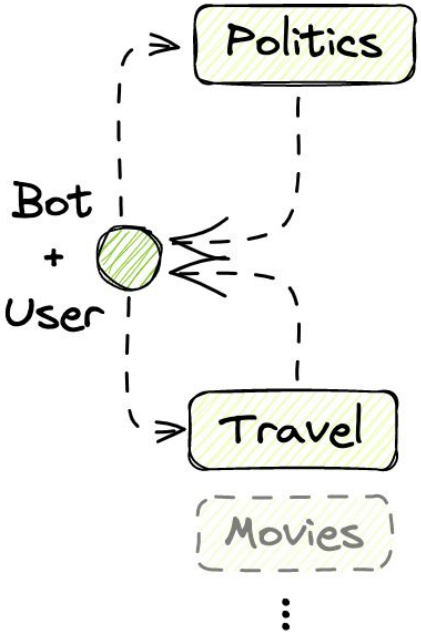
Alexa TaskBot Challenge: Setting

Talk Outline

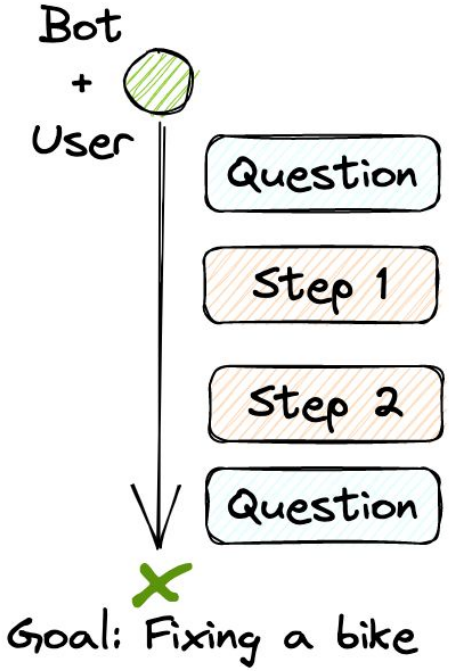
- Conversation Flow System Overview
- TaskGraphs
- TaskGraph text and image augmentations
- Question Answering
- Neural Decision Parser
- Conclusion & Published Works
- 🔥 Recorded and Live Demo 🔥

Real-World Task Assistance

SocialBot



TaskBot

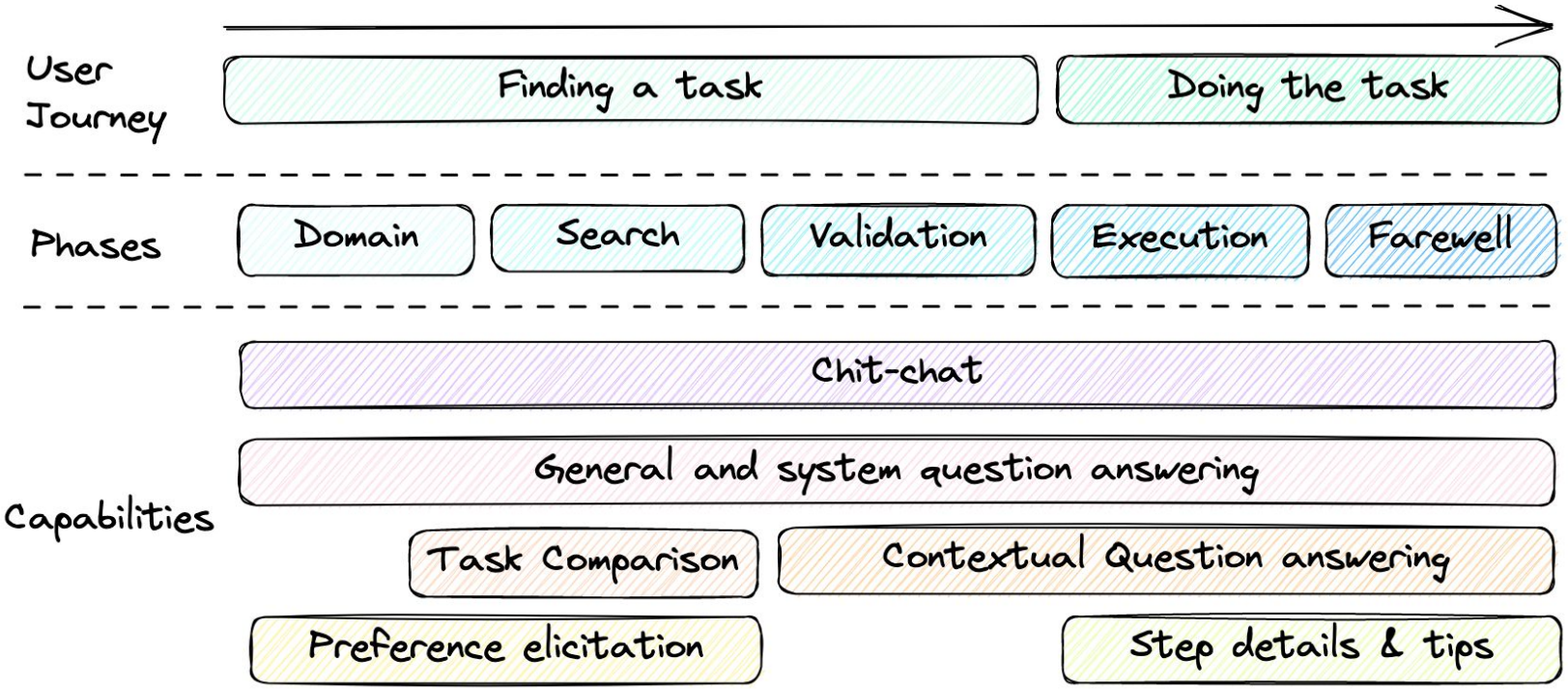


2022 AlexaPrize

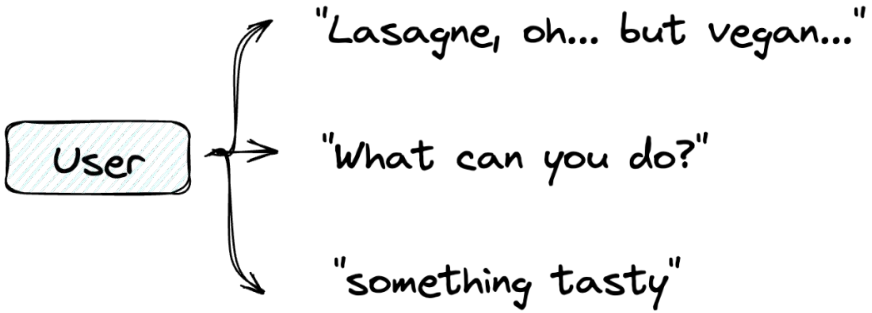
Cooking

Crafts & DIY

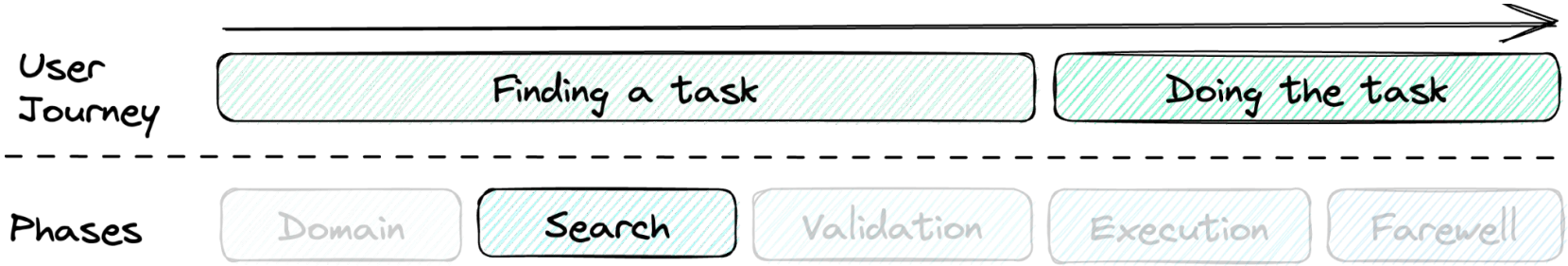
Conversation flow: A phased approach

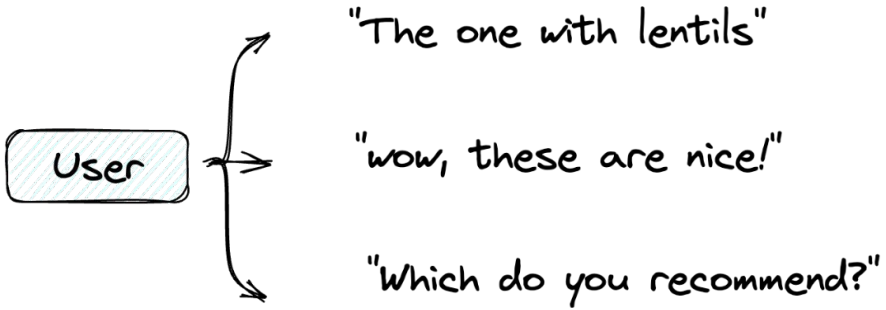


Conversation flow: Domain phase

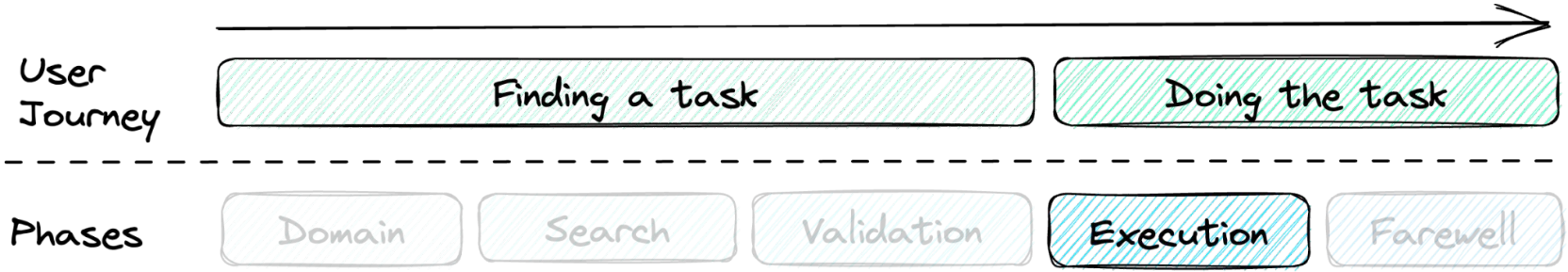


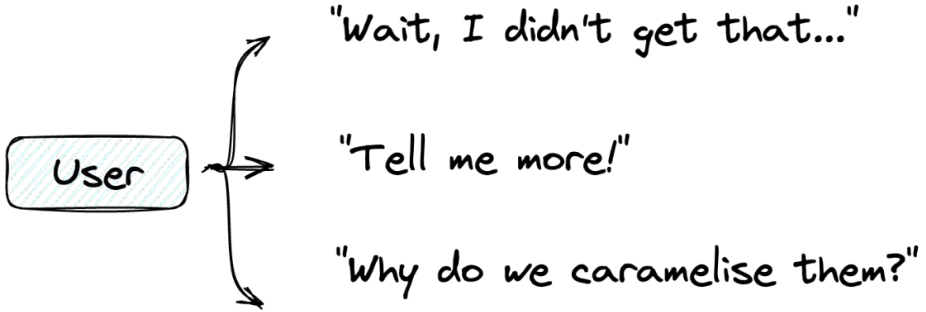
Conversation flow: Search phase





Conversation flow: Execution phase



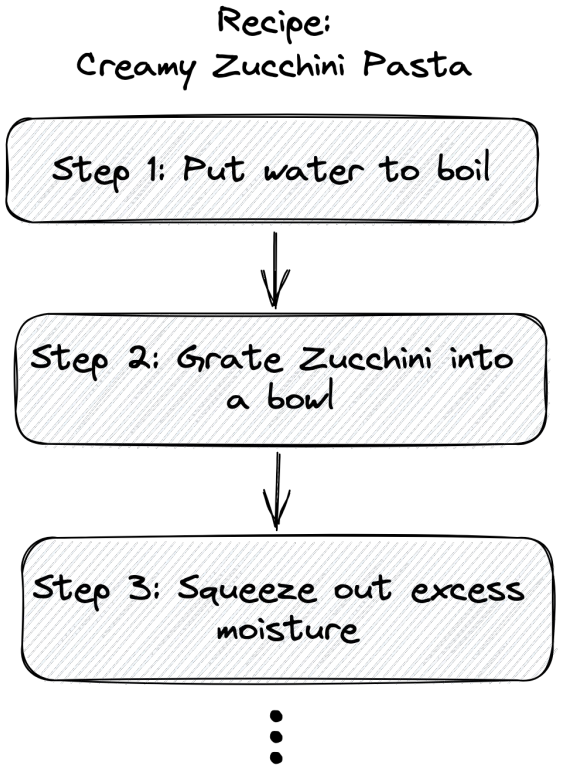


Talk Outline

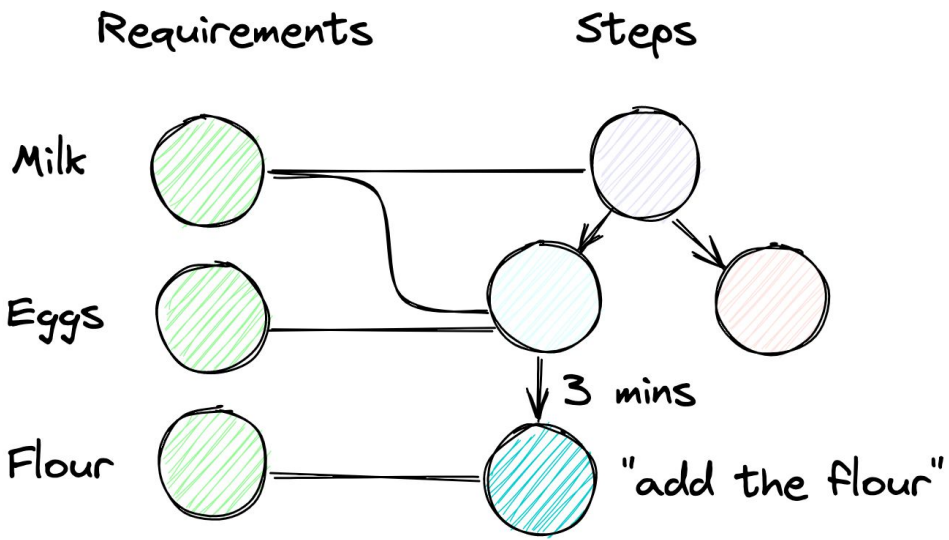
- ~~Conversation Flow System Overview~~
- TaskGraphs
- TaskGraph text and image augmentations
- Question Answering
- Neural Decision Parser
- Conclusion & Published Works
- 🔥 Recorded and Live Demo 🔥

Standard Task Representation

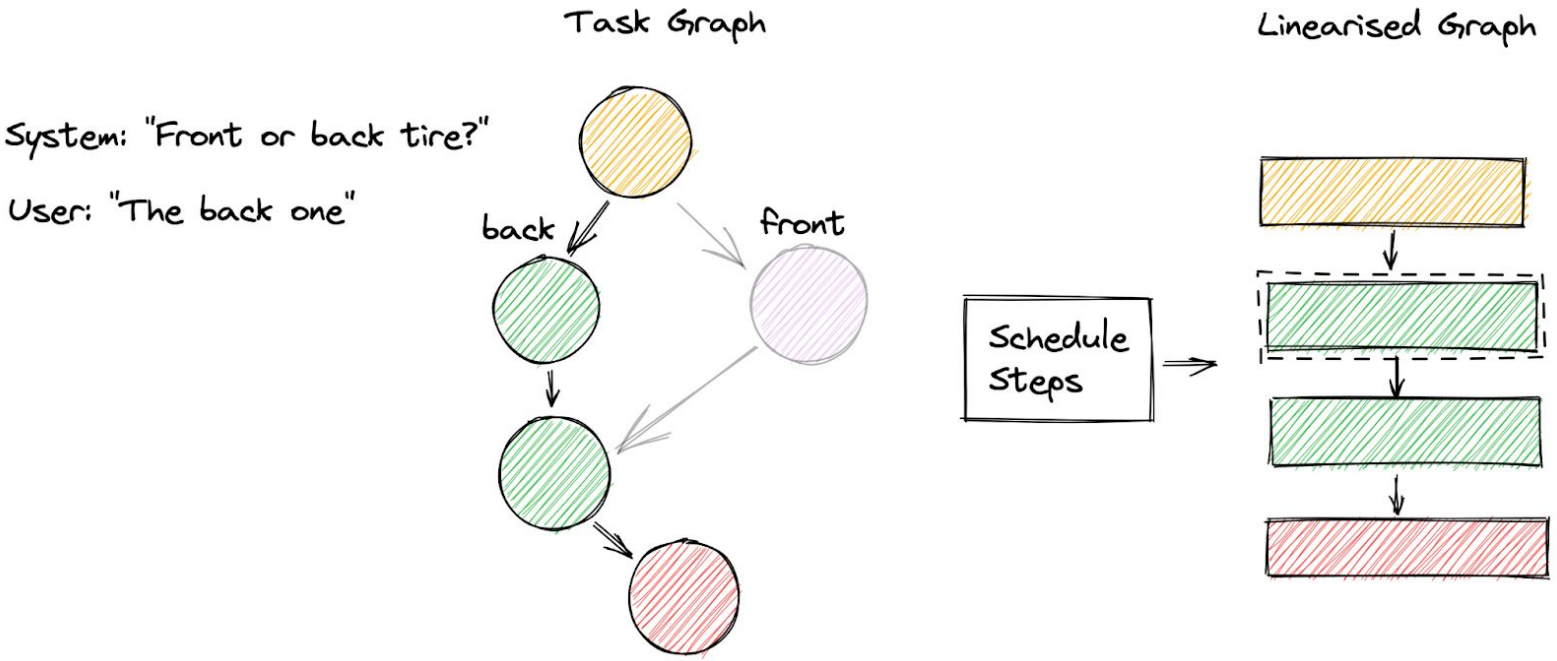
- No system initiative
- Lacking personalisation
- Dry speech interactions



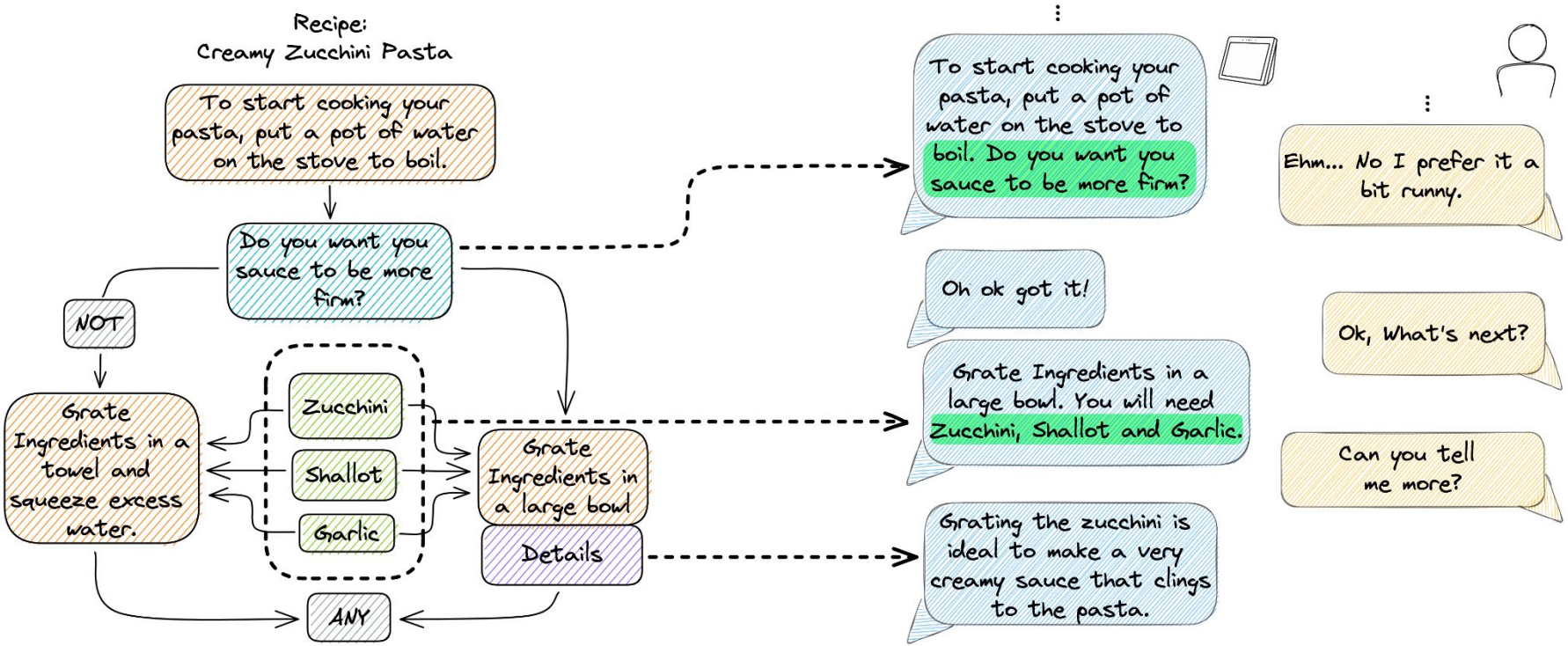
Task Graphs: Ingredient Linking



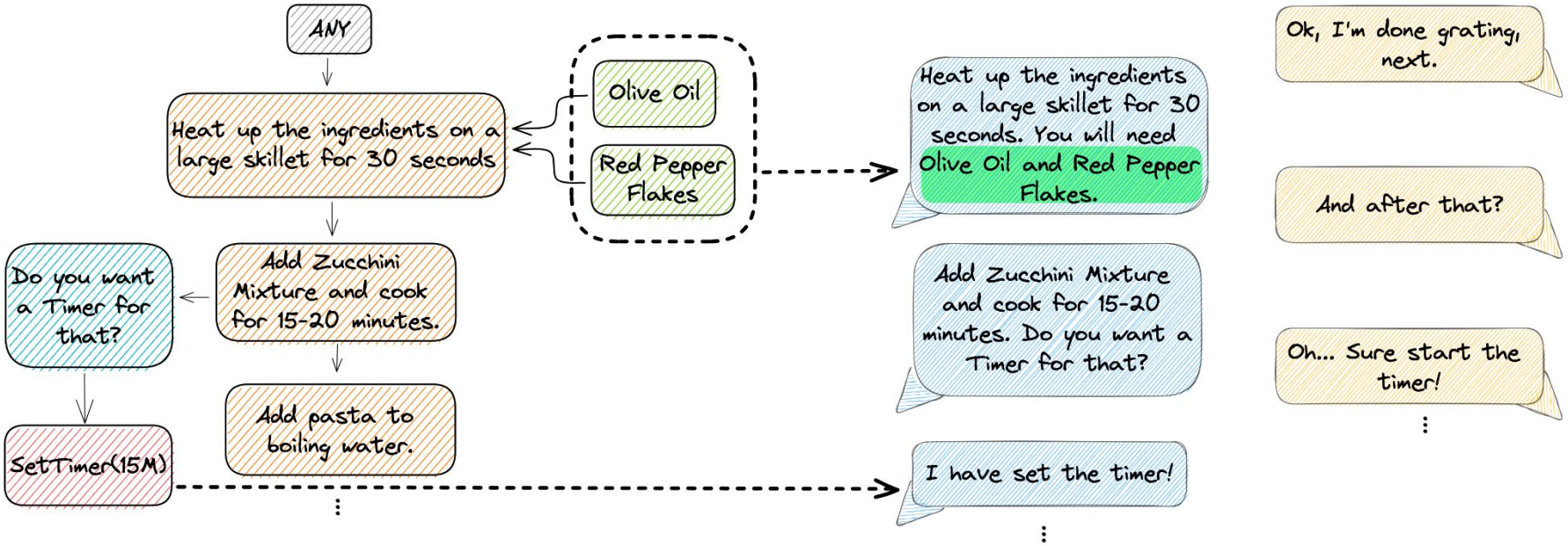
Task Graphs: Conditional Execution



Task Graphs: a live example



Task Graphs: function initiative



Talk Outline

- ~~Conversation Flow System Overview~~
- ~~TaskGraphs~~
- TaskGraph text and image augmentations
- Question Answering
- Neural Decision Parser
- Conclusion & Published Works
- 🔥 Recorded and Live Demo 🔥

Available Data

Amazon APIs
- Only Online

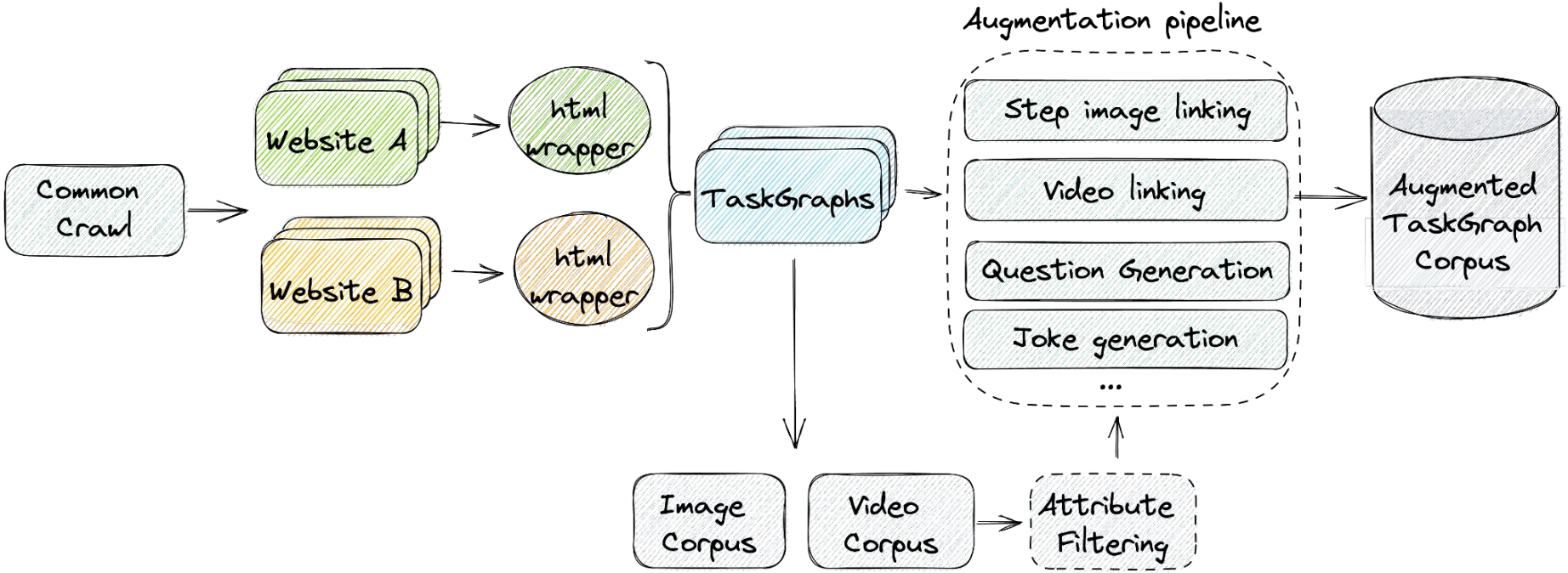
Common Crawl
- 8 selected sites
- 200k total pages

Recipe 1M

WikiHow Open Dump

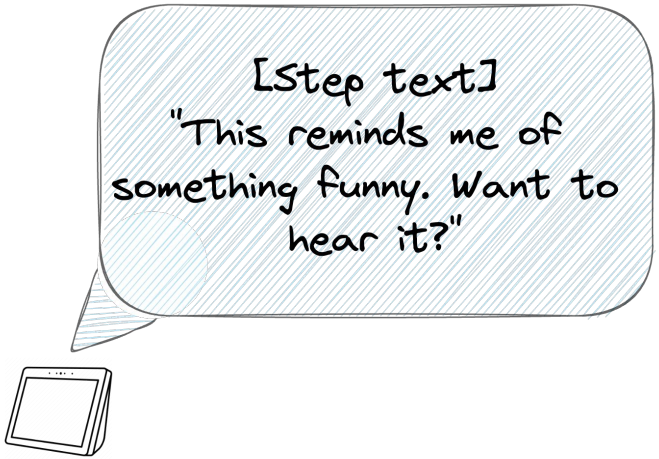
Suitable for offline processing

System Overview: Offline / Async Pipeline

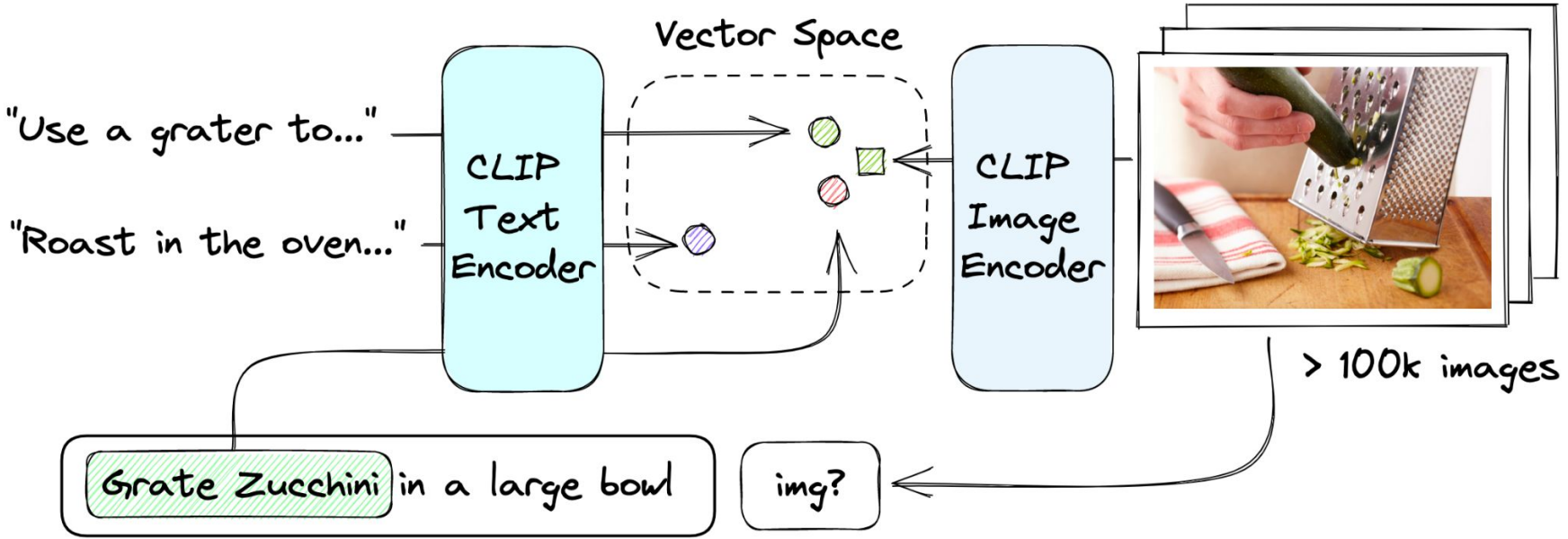


Augmentations to TaskGraphs

- Only listening to steps is very dry
  - Online recipes are meant to be **read**, but **not spoken**
- We pre-process TaskGraphs before displaying to the user
  - **Image and Video Linking**
  - **Step details**: Make steps concise by truncating
  - **Requirement linking**
  - **Joke Generation**: jokes based on the step content
  - **Question generation**: System lead task initiative

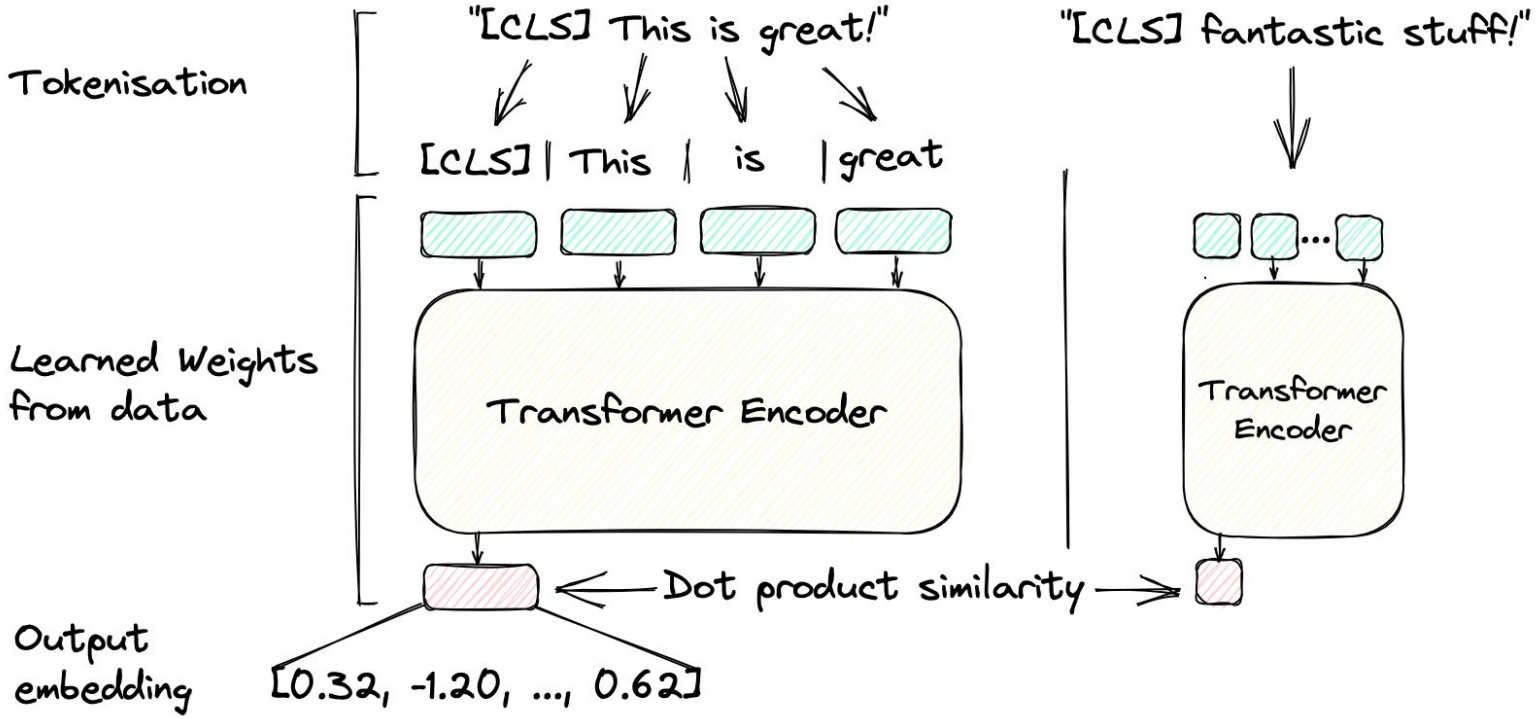


Task Graphs: Image Linking

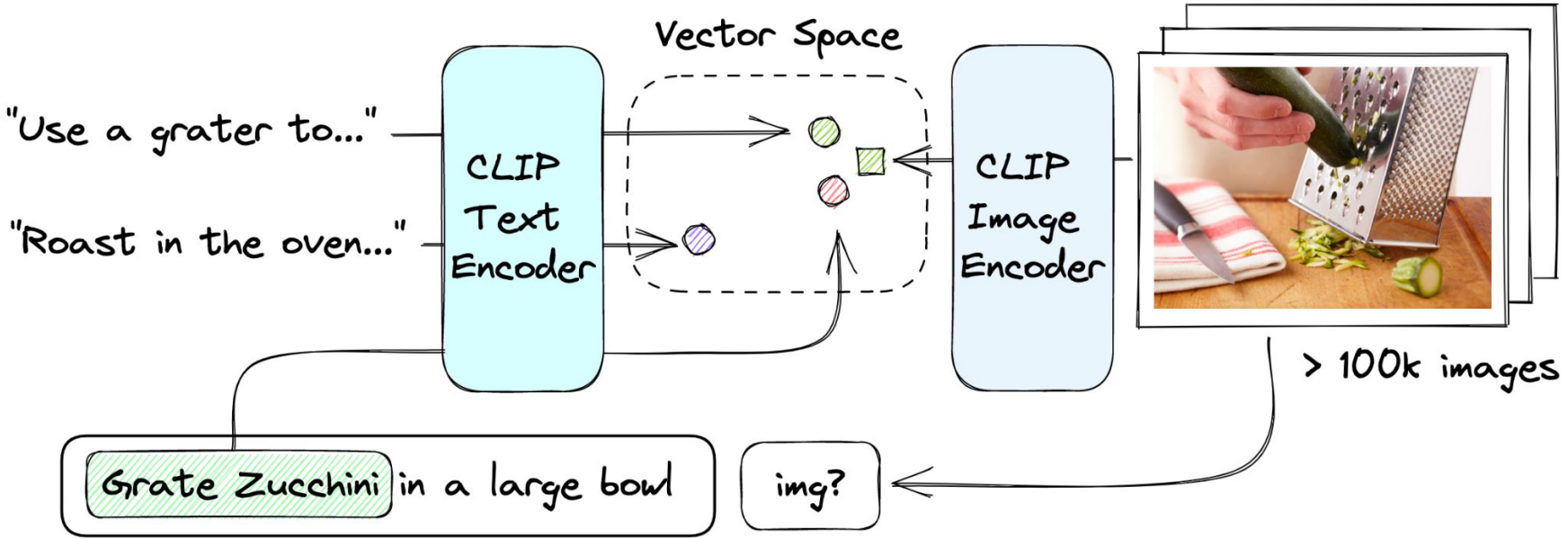


Learning Transferable Visual Models From Natural Language Supervision (2021, Radford et al.)

Quick Intro: Text Embeddings

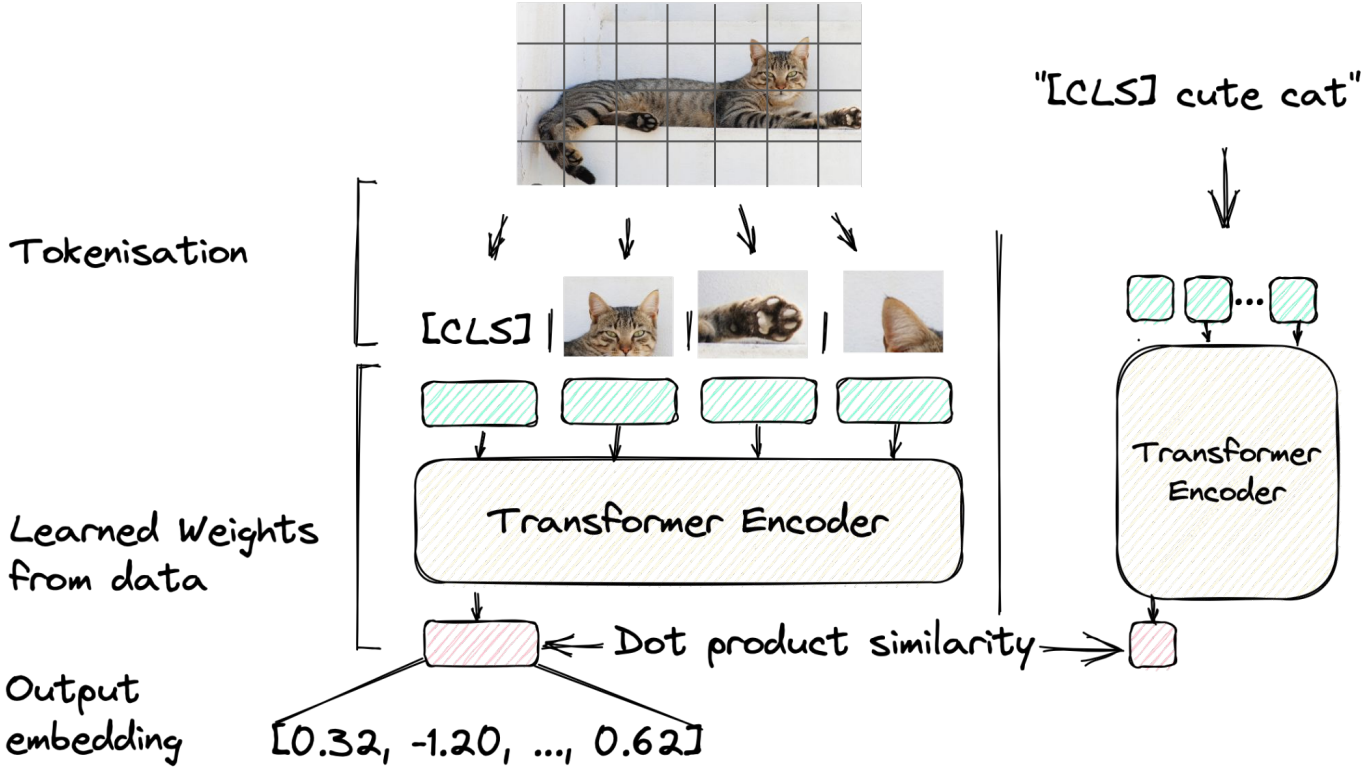


Task Graphs: Image Linking

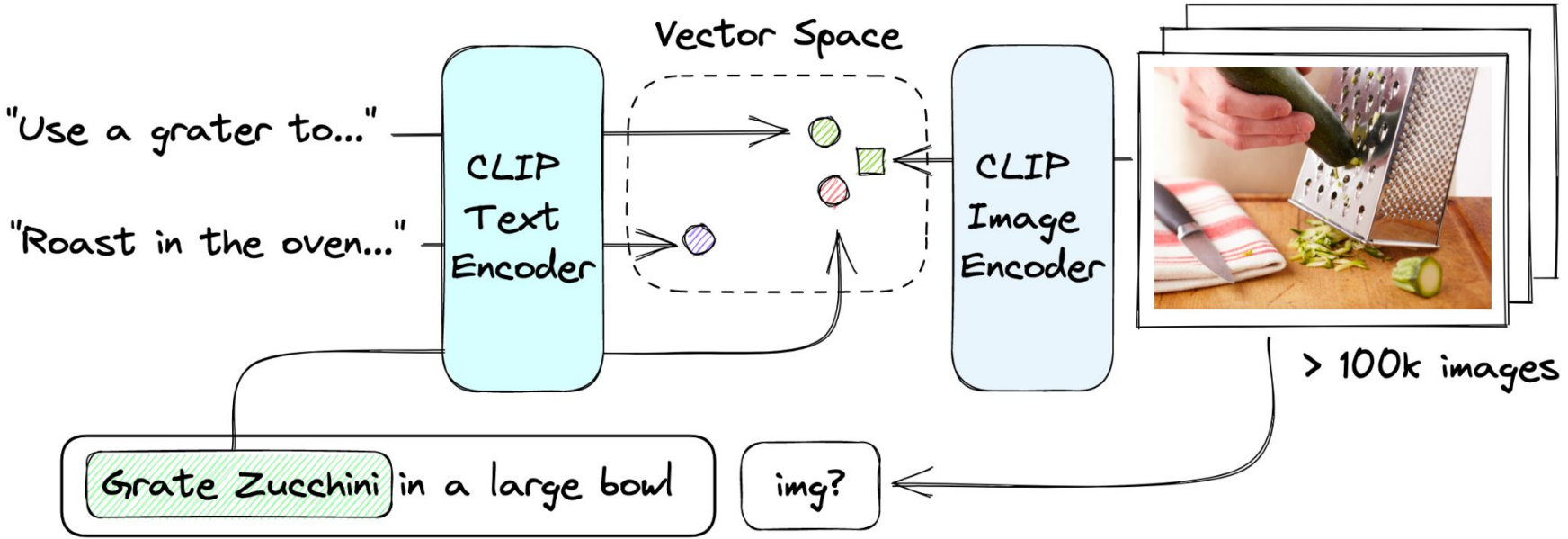


Learning Transferable Visual Models From Natural Language Supervision (2021, Radford et al.)

Quick Recap: Image-Text Embeddings

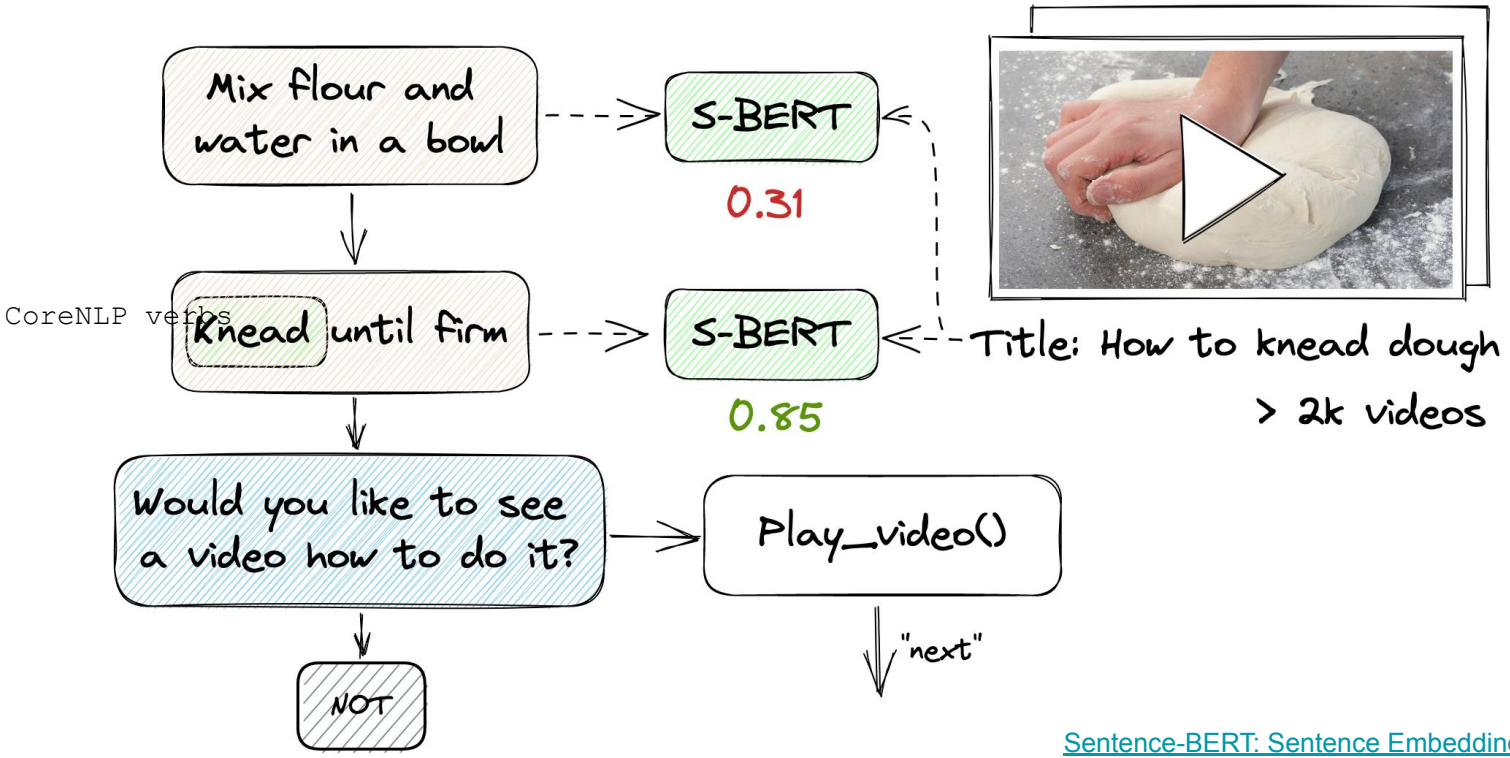


Task Graphs: Image Linking



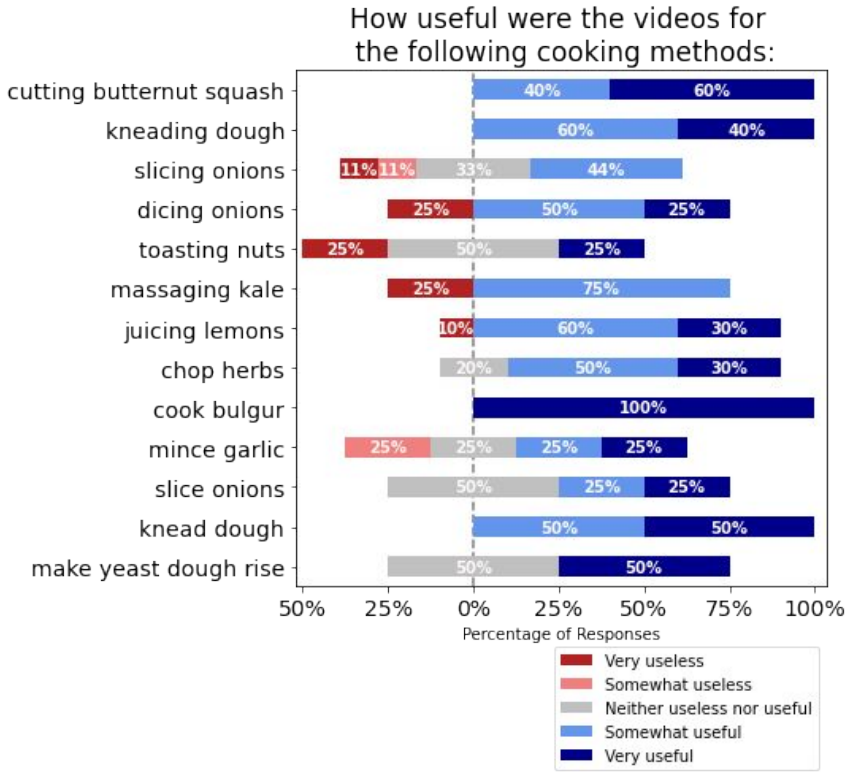
Learning Transferable Visual Models From Natural Language Supervision (2021, Radford et al.)

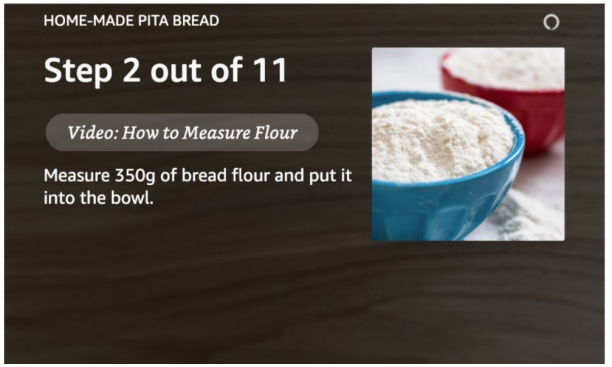
Task Graphs: “How-to” Video Linking

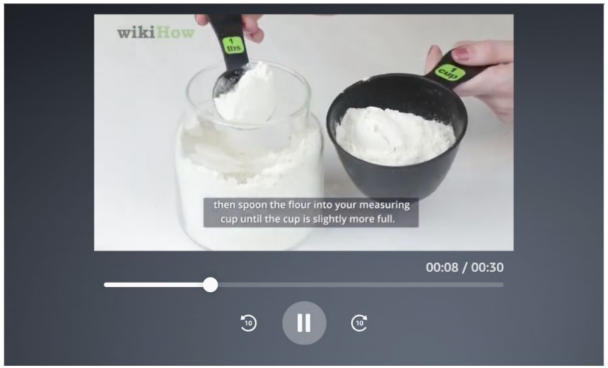


Sentence-BERT: Sentence Embeddings using Siamese BERT-Networks (2019, Reimers and Gurevych)

Task Graphs: Effects of Video on Users







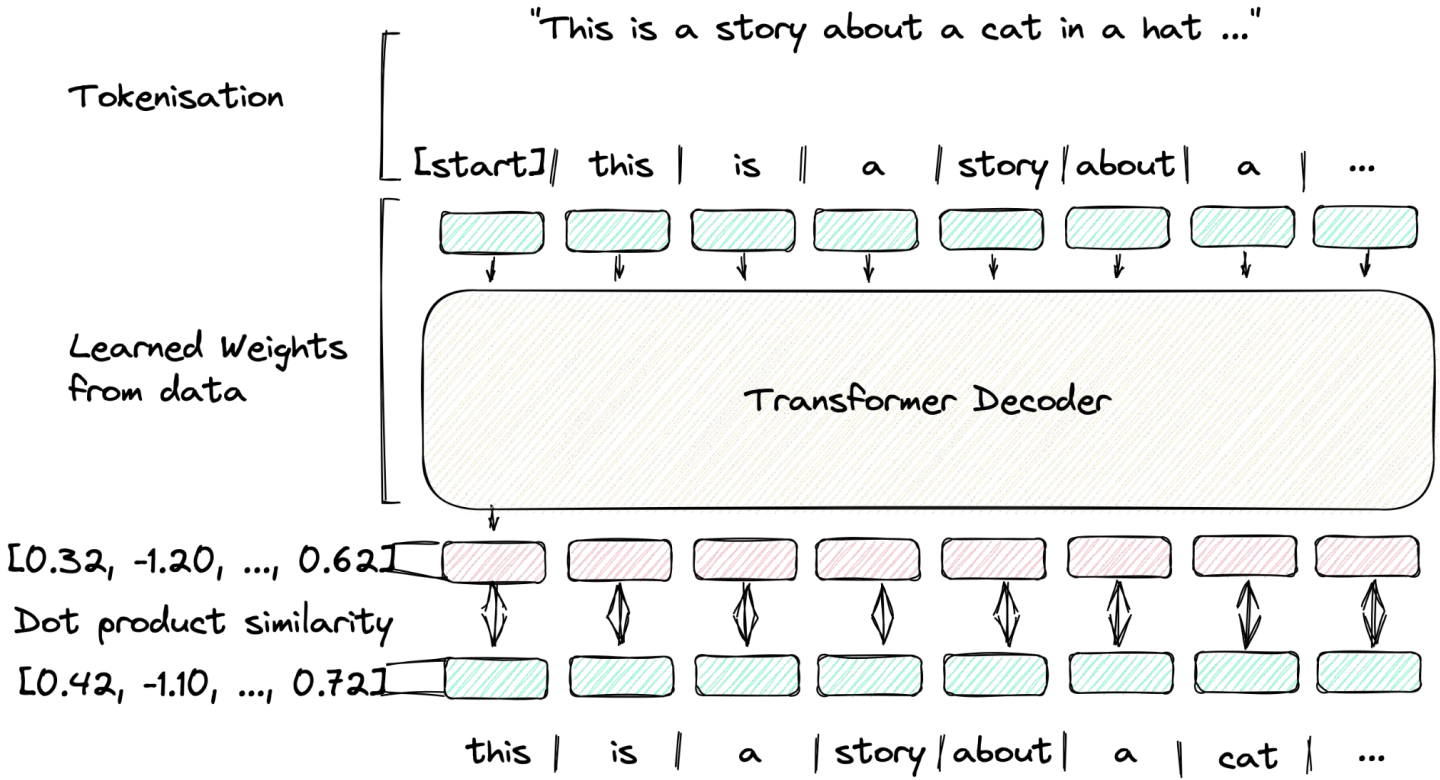
(Fischer et al. SIGDial 2022)

Video Linking User Study - Observations

Poor semantic content of titles leads to distracting / “entertaining” the user

	Automatic	Perfect
"Remove the stems and roughly chop the herbs"		
"Using a tablespoon measure, scoop out heaping spoonfuls of the mixture into your hand."		

Quick Intro: Text Generation



GPT3/J
175B params

As much data as privacy allows

Offline Question Generation

Task: Fried Chicken
Previous steps:
- "In a small bowl, combine the remaining 1/4 cup buttermilk and the milk."
- "Pour the milk mixture into the flour mixture and, with a pastry cutter or fork, gradually mix until there are little lumps throughout."

Future steps:
- "If necessary, add a little more flour or milk to the bowl in order to make it slightly lumpy."
- "Heat 1 1/2 inches of oil in a deep skillet or Dutch oven over medium-high heat until a deep-fry thermometer inserted in the oil reaches 365 degrees F. Lower the heat slightly, if necessary, to keep the oil from getting hotter."

The following is a question from a virtual assistant that first gives a hint of future steps and ends the question asking how the past steps are going.
Format: future hint + previous step question

"How is the mixture looking? If it's too thick, add a little more milk. If it's too thin, add a little more flour."

"How's the milk mixture coming along? If you're satisfied with it, it's time to heat up the oil for frying."

Offline Joke Generation

Wrap the punchline of the joke in the following tags <premise></premise> <punchline></punchline>

John: Omg omg, I came up with a great joke about salmon fish cakes!

Mary: hahahah no way! go for it!

John: Alright, here goes: <premise>What do you call a fish with no eyes?</premise><punchline>A fsh!</punchline>

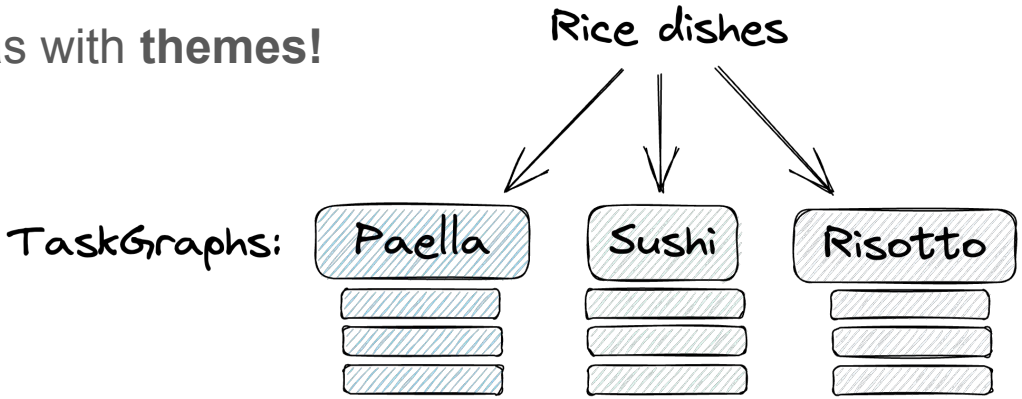
P.S. Look out for this later in the demo :)

Augmentations for TaskGraph groups

- Queries like “make pasta”, “cook dinner”, “something healthy”
  Are vague and return poor search results

- How to create groupings with little manual effort?

- Create trivia to give users ideas with **themes!**



Offline Theme Generation

Theme Schema

Field	Type	Description	Required
`theme`	String	The name of the theme	Yes
`example_queries`	Array of Strings	A list of example queries to be used in the query bar	Yes
`description`	String	A description of the theme	Yes
`trivia`	Array of Strings	A list of fun facts about the theme	Yes
`popular_dishes`	Array of Strings	A list of popular dishes associated with the theme	Yes
`clarifying_questions`	Array of Strings	A list of general questions about the qualities of the dish name i.e. spice level, intolerances...	Yes

```
theme = {
    'theme':'rice'
```

Augmentation Annotation with LLMs: theme creation

Theme Schema

Field	Type	Description	Required
`theme`	String	The name of the theme	Yes
`example_queries`	Array of Strings	A list of example queries to be used in the query bar	Yes
`description`	String	A description of the theme	Yes
`trivia`	Array of Strings	A list of fun facts about the theme	Yes
`popular_dishes`	Array of Strings	A list of popular dishes associated with the theme	Yes
`clarifying_questions`	Array of Strings	A list of general questions about the qualities of the dish name i.e. spice level, intolerances...	Yes

theme = {
'theme':'rice'
'example_queries':['rice dishes','rice cuisine', 'rice']
'description':'Rice dishes are enjoyed by people all over the world and come in many different forms. In Asian cultures, rice is often eaten with chopsticks and is used as a base for many different dishes. In Japanese cuisine, sushi is a popular dish that consists of cooked rice that is vinegar flavored and rolled with seafood or vegetables. In Chinese cuisine, rice is often stir fried with vegetables or meat. In India, rice is a staple food and is often served with curries or lentils. Rice dishes vary greatly in flavor and ingredients depending on the culture in which they are made.'
'trivia':['The most popular type of rice is white rice.', 'Rice is a grain that is grown in warm climates.', 'Rice is the staple food for about half the world's population.']
'popular_dishes':['Sushi', 'Fried Rice', 'Risotto', 'Paella']
'clarifying_questions':['What type of rice would you like?', 'Are there any flavors would you prefer in your rice dish?', 'What meat or vegetables would you like in your rice dish?']
}

TaskGraph Augmentations Outcomes

- Over 200 themes automatically created containing
  - Sample recipes
  - Possible queries
  - Trivia and Fun facts

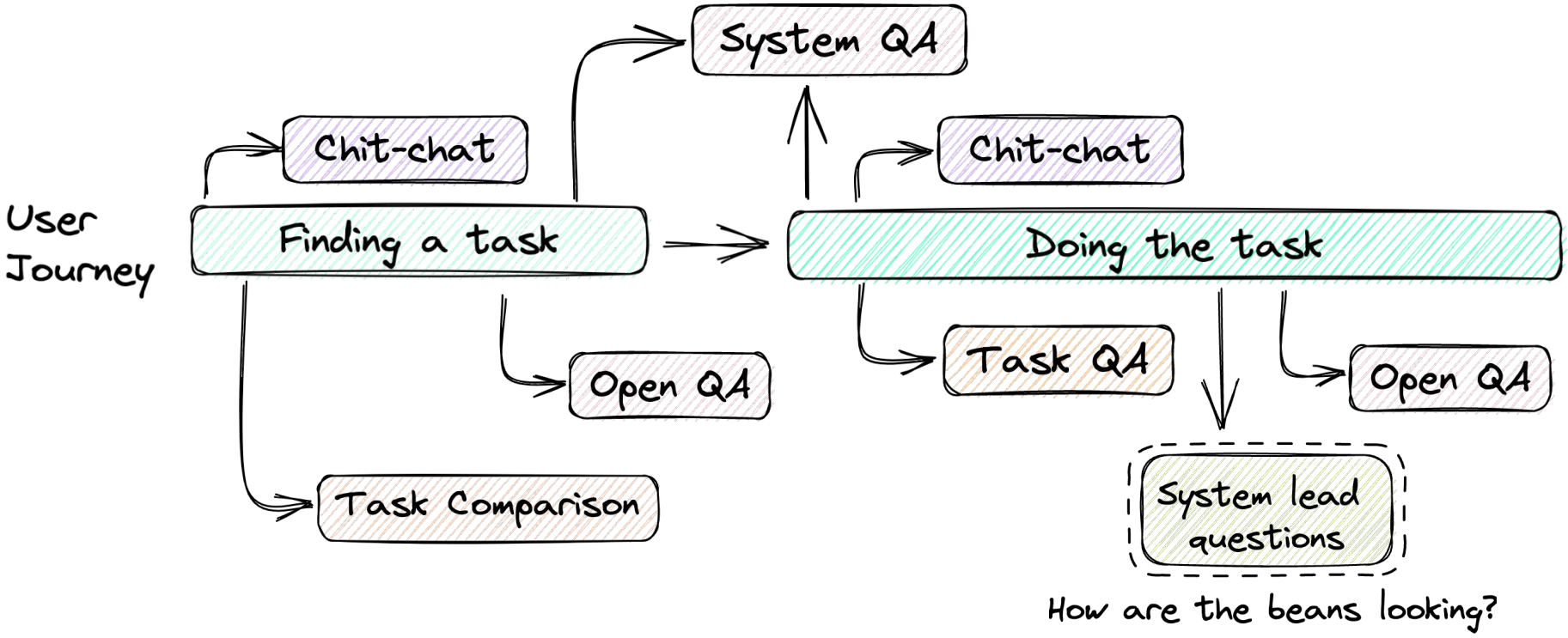
- Every task has per-step questions, details, fun facts and jokes
  - Pre-compute with offline corpus: GPT-3, GPT-J
  - Async augmentations with GPT-J

- Every task has per-step images and selected videos if relevant
  - Videos especially help communicate techniques compared to voice only

Talk Outline

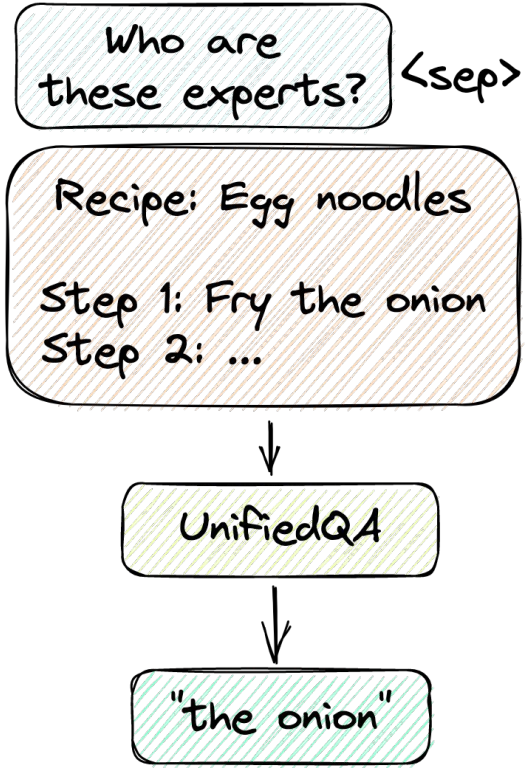
- ~~Conversation Flow System Overview~~
- ~~TaskGraphs~~
- ~~TaskGraph text and image augmentations~~
- Question Answering
- Neural Decision Parser
- Conclusion & Published Works
- 🔥 Recorded and Live Demo 🔥

GRILLBot Question Answering

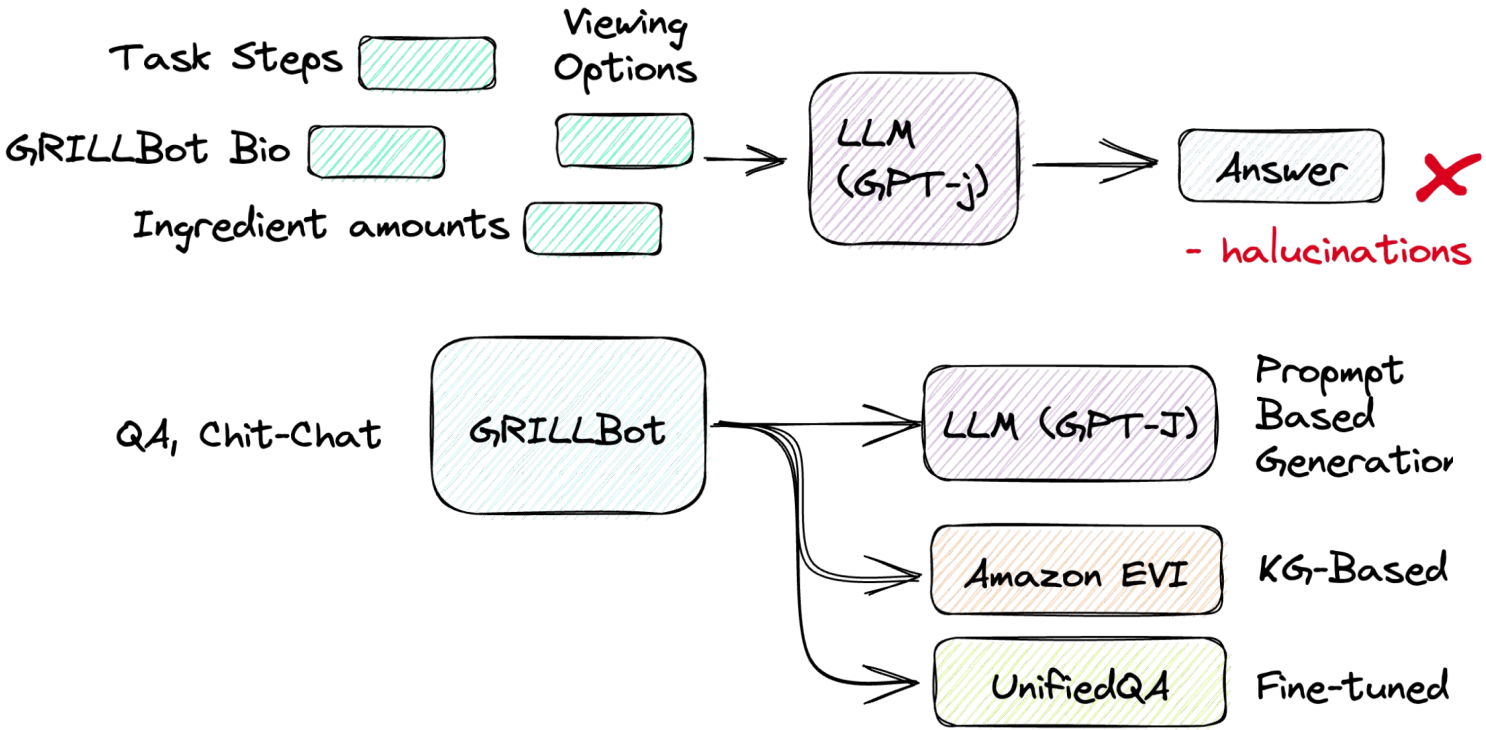


Building a TaskBot QA System

- Scarce domain specific contextual QA data
- High cost to model iteration
- Using Pre-trained QA systems
  - Trained on academic reading comprehension datasets
  - Unstructured text data as context
- Limitations
  - Domain shift results in poor answers
  - Model hallucinations

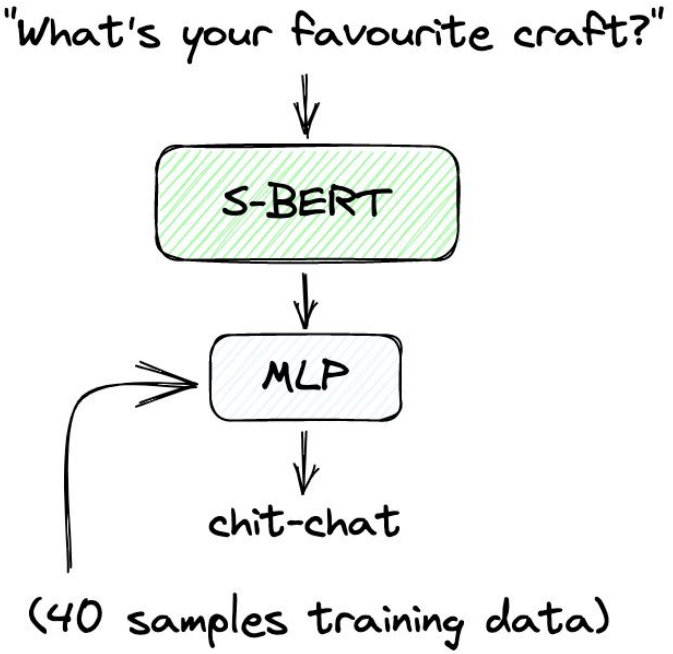


Large Language Models (LLM): tools for rapid iteration



GRILLBot QA: Type classification

Ingredient question	"How much oil do I need?"
Current task question	"How hot was the oven again?"
Step question	"How many steps are left?"
General cooking or DIY question	"How does wood glue work?"
Other domain question	"What's the capital of Paris?"
Chit-chat	"What's your favourite recipe?"
System capabilities question	"how do I set a timer?"
Current viewing options question	"Can you compare these recipes?"
Ingredient substitution	"I'm dairy-free, is oat milk alright?"



Rapid QA Iteration with LLMs

Current Viewing Options Question

Current Candidates:
- egg noodles by J Kenji Lopez
- noodle stir fry
- vegan noodles

The following is a conversation with an AI assistant. The assistant is helpful, creative, clever, and very friendly. The AI is helping the user choose one of the candidates and explaining why.

Human: Which do you recommend?
AI: I recommend the egg noodles by J Kenji Lopez. They are authentic and delicious, and he is a well-respected authority on noodles.

Current Task Question

Step 1:
....

Step 2:
...

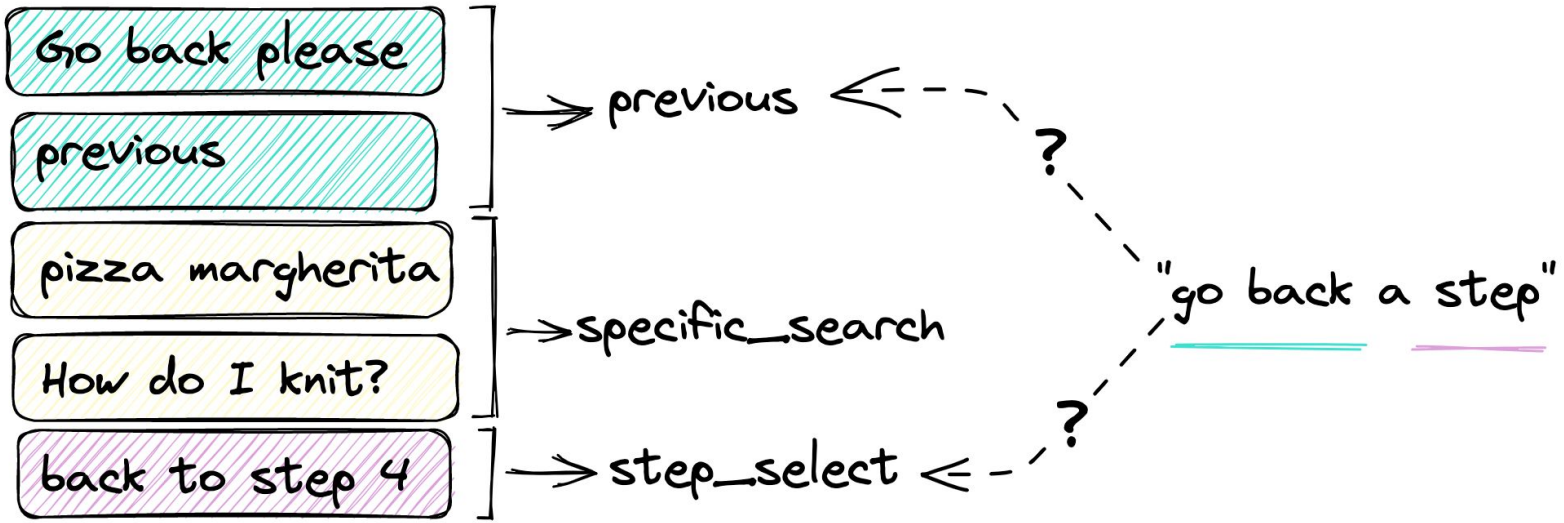
I am a highly intelligent question answering bot. If you ask me a question that is rooted in truth about {request.query.taskmap.title}, I will give you a friendly answer. If you ask me a question that is nonsense, trickery, or has no clear answer, I will respond with "Good question about {topic}, but the task we\'re following doesn\'t say".

Q:{request.query.text}
A:

Talk Outline

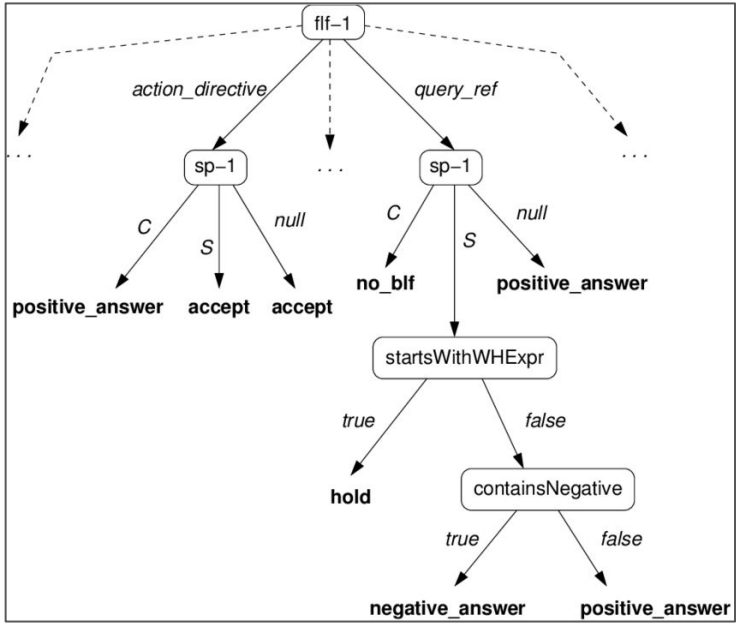
- ~~Conversation Flow System Overview~~
- ~~TaskGraphs~~
- ~~TaskGraph text and image augmentations~~
- ~~Question Answering~~
- Neural Decision Parser
- Conclusion & Published Works
- 🔥 Recorded and Live Demo 🔥

Traditional Intent classification



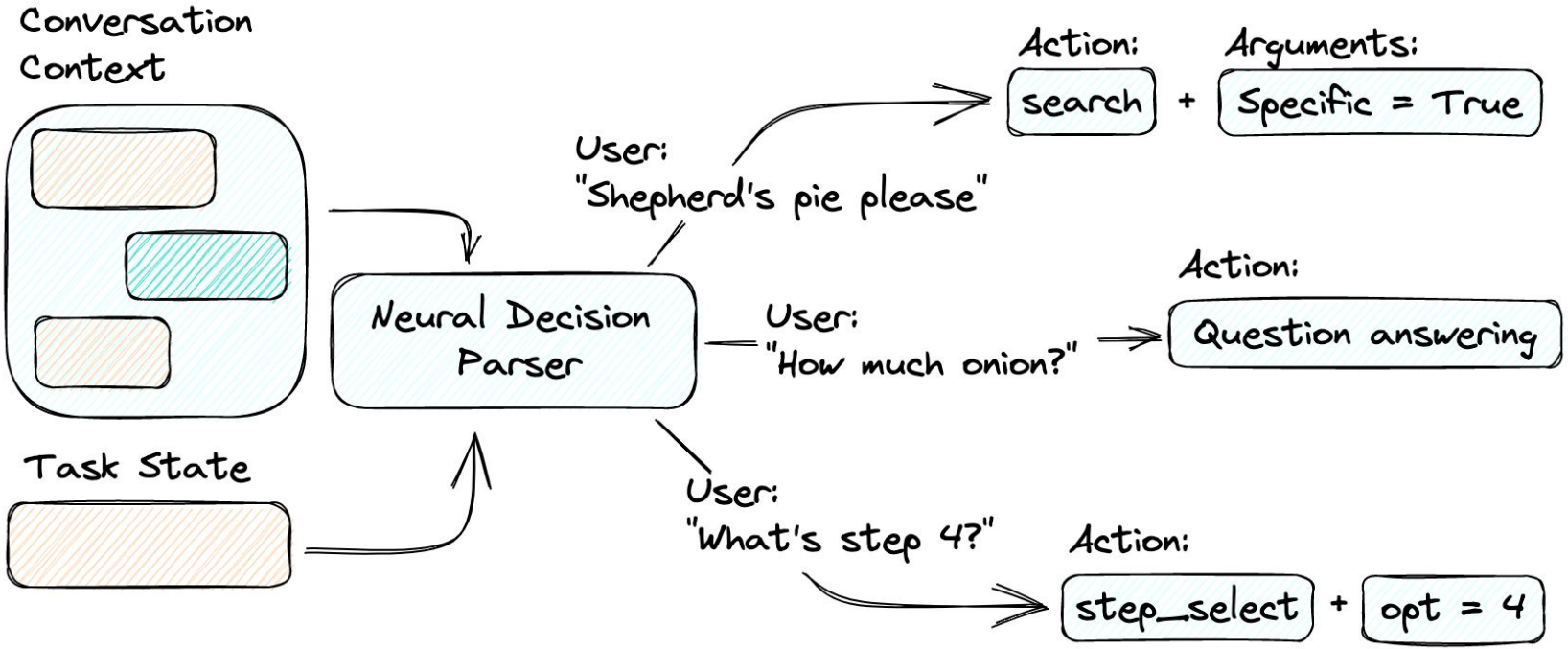
Traditional Intent classification

- Simple when starting
- Becomes very complex with many intents
- Only operates on the utterance at hand (Amazon intents)
- No compositionality between branches
- Brittle when understanding complex utterances



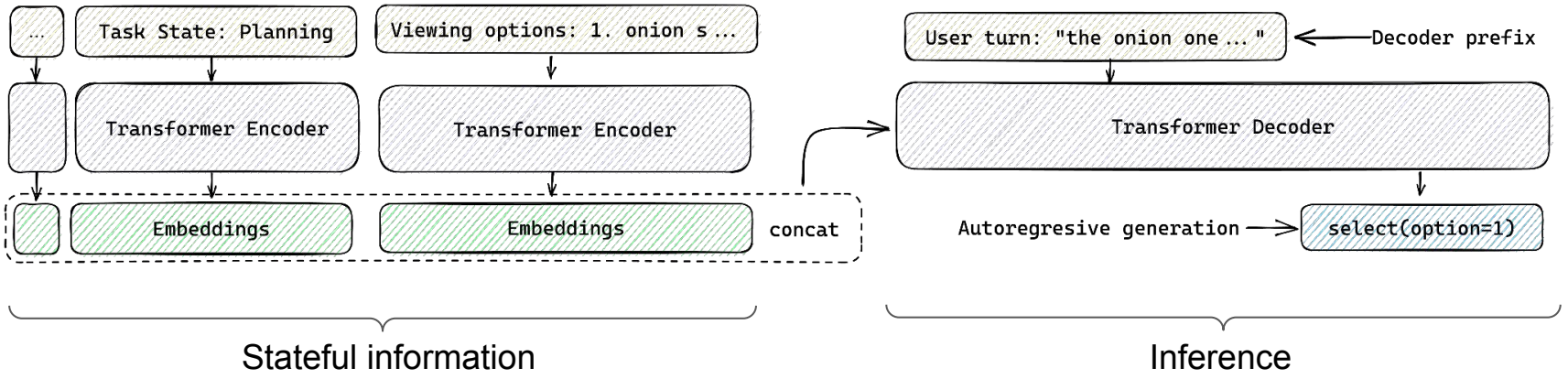
Example of a decision tree for dialogue act classification (Keizer et al. 2015)

Moving away from static intents



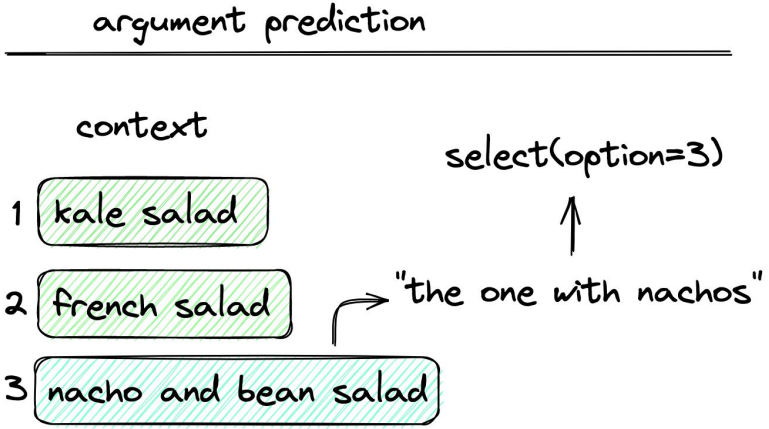
Neural Decision Parser Architecture

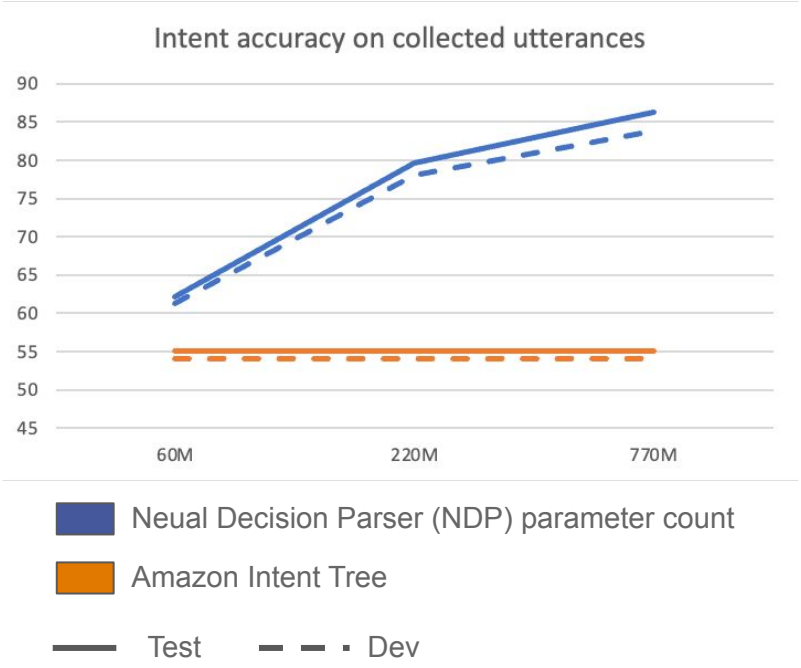
- Encode stateful information independently of current utterance
- Decoder accesses evidence like a database
- Allows pre-computation of stateful information with ad-hoc decoder access



Understanding model performance

- 1.5k curated utterance dataset for testing
- Larger models improve (of course)
- Pre-computation enables large amounts of context with large models



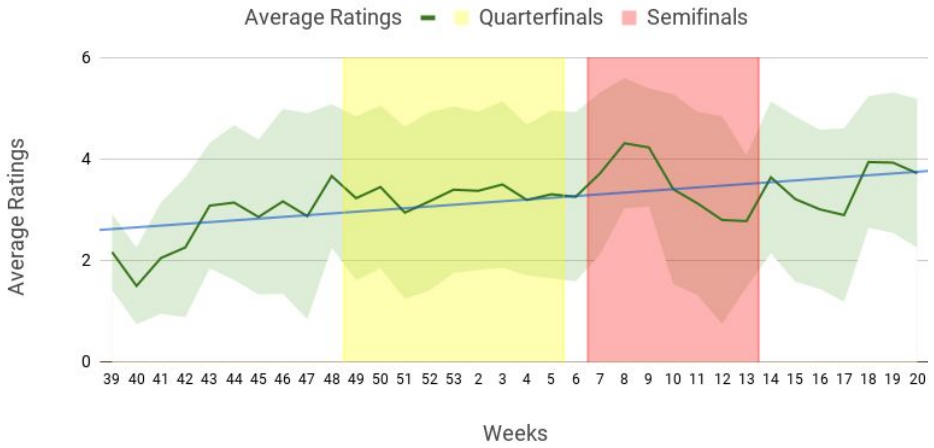


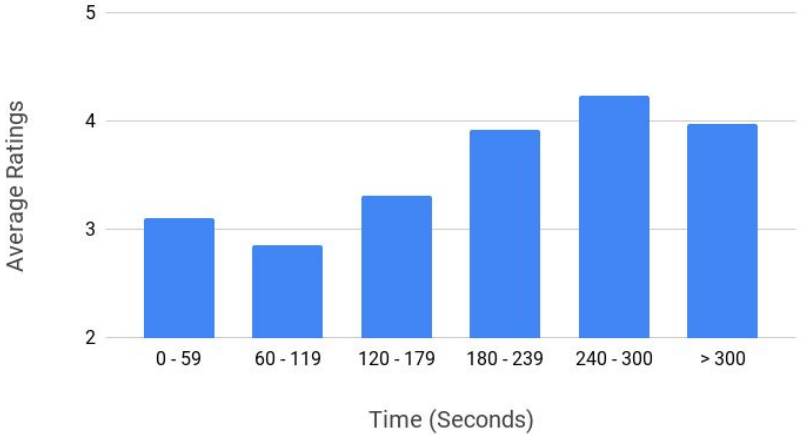
University of Glasgow

Talk Outline

- ~~Conversation Flow System Overview~~
- ~~TaskGraphs~~
- ~~TaskGraph text and image augmentations~~
- ~~Question Answering~~
- ~~Neural Decision Parser~~
- Conclusion & Published Works
- 🔥 Recorded and Live Demo 🔥

Performance over the year



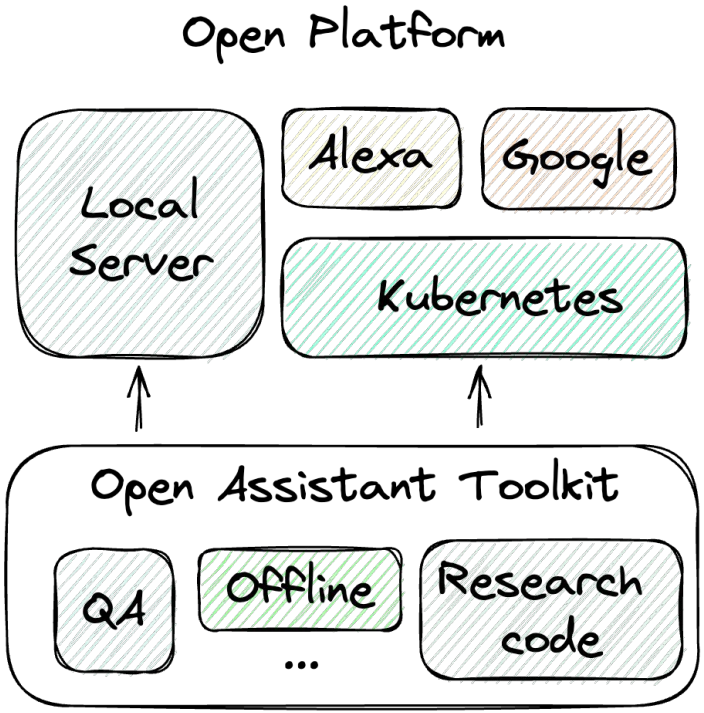


Closing Remarks

- Systems with real users show vastly different interaction skill levels
  - New users require robust intent handling and navigation
  - Experienced users almost exclusively ask questions
- Successful balance between product and research
  - Deadlines almost every 2 weeks: quarter, semi-finals…
  - Strong team dynamic is critical to build stable system
  - Stable system enables experimentation

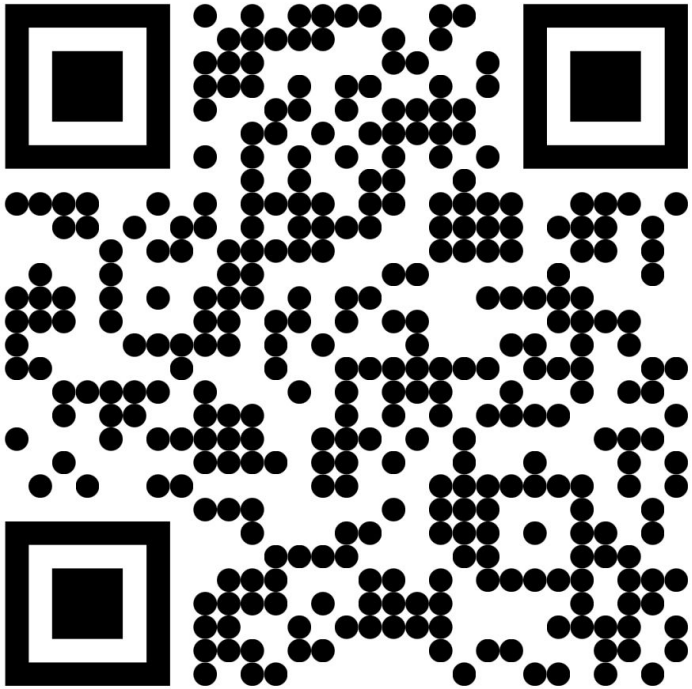
GRILLBot => Open Assistant Toolkit (OAT)

- Alexa Proceedings Overview (Gemmell et al. 2022)
- GRILLBot system paper (Gemmell et al. SIGDial 2022)
- VILT: Video Instructions Linking for Complex Tasks (Fischer et al. IMuIR 2022)
- Conversational assistance directions (SIGIR 2022 Tutorial)



GRILLBot => Open Assistant Toolkit (OAT)

- Alexa Proceedings Overview (Gemmell et al. 2022)
- GRILLBot system paper (Gemmell et al. SIGDial 2022)
- VILT: Video Instructions Linking for Complex Tasks (Fischer et al. IMuIR 2022)
- Conversational assistance directions (SIGIR 2022 Tutorial)



(GitHub - Open Assistant Toolkit)

University of Glasgow

Talk Outline

- ~~Conversation Flow System Overview~~
- ~~TaskGraphs~~
- ~~TaskGraph text and image augmentations~~
- ~~Question Answering~~
- ~~Neural Decision Parser~~
- ~~Conclusion & Published Works~~
- 🔥 Recorded and Live Demo 🔥

Demo: What to look out for?

- Fluency of transitions between user and system
- Robustness to non-standard contextual questions: “who are these experts?”
- System initiative with generated questions and fun facts at the end of system utterances
- Joke intonation by separating premise and punchline