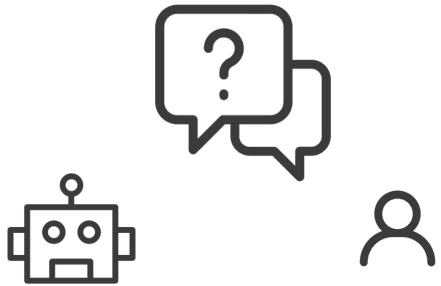
From Text to Multimodal CUI

EN. 601.792.01

Ziang Xiao
Department of Computer Science
Spring 2024



Announcement

- Office Hour: Tue. 4-5pm Malone 309
- Course Website: https://adv-conv-ui.cs.jhu.edu/sp24/index.html
- Piazza: https://piazza.com/class/lrp2hge6oz845x/
- Canvas:https://jhu.instructure.com/courses/63649
- Course Format: In person, Research-focused, Discussion-based

Course Topics

- Conversational User Interface Basics
- LLMs and CUI
- Evaluation
- Human Agency in CUI
- Emerging Topics

Discussion Format

- For each class, you will be randomly assigned into a group w/ 4-5 students.
- The group announcement will be made prior to the class.
    - A lead discussant will be appointed
    - Other group members should coordinate and each choose a different role
- The lead discussant should read everyone's response prior to the class, they will lead and organize the discussion and synthesize for reporting back to the whole class.
- Given the hat you wear, you will participate the discussion from a unique perspective on the class topic and assigned readings.

Roles

 **Scientific Reviewer.** The paper has not been published yet and is currently submitted to a top venue where you've been assigned as a peer reviewer.

 **Archaeologist.** This paper was found buried under ground in the desert. You're an archeologist who must determine where this paper sits in the context of previous and subsequent work. Find and report on one *older* paper cited within the current paper that substantially influenced the current paper and one *newer* paper that cites this current paper.

 **Academic Researcher.** You're a researcher who is working on a new project in this area. Propose an imaginary follow-up project *not just* based on the current but only possible due to the existence and success of the current paper.

Roles

 **Industry Practitioner.** You work at a company or organization developing an application or product of your choice. Bring a convincing pitch for why you should be paid to implement the method in the paper, and discuss at least one positive and negative impact of this application.

 **Private Investigator.** You are a detective who needs to run a background check on one of the paper's authors. Where have they worked? What did they study? What previous projects might have led to working on this one? What motivated them to work on this project? Feel free to contact the authors, but remember to be courteous, polite, and on-topic.

 **Social Impact Assessor.** Identify how this paper self-assesses its (likely positive) impact on the world. Have any additional positive social impacts left out? What are possible negative social impacts that were overlooked or omitted?

Course Project

(More details TBA)

- Group project with 3-4 students
- Topics related to CUI
- Project Scope
  - System
  - Study
  - Analytics
  - Or anything if you think is suitable (we can discuss)

Course Project

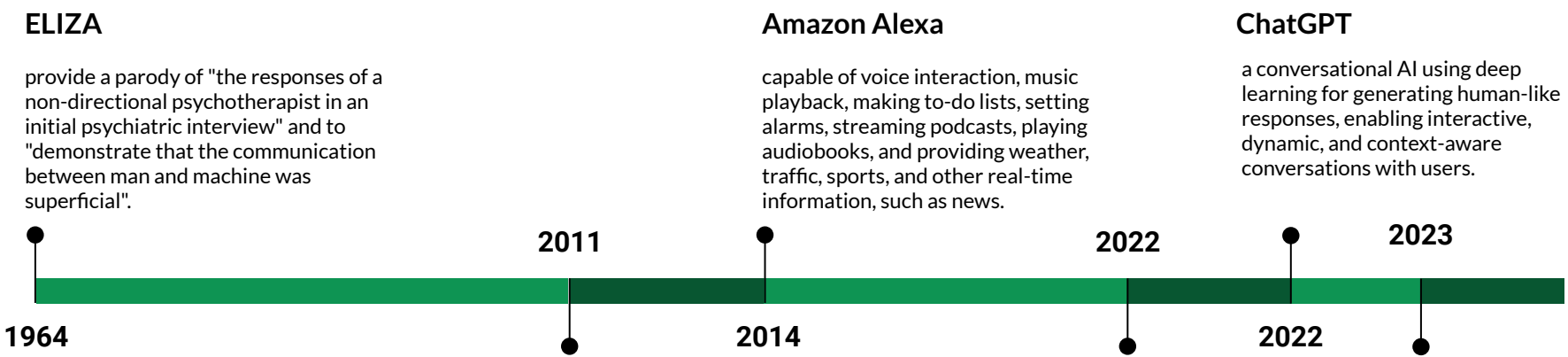
(More details TBA)

Find your team!

Grading

- Reading responses (15%)
- Class Engagement (30%)
  - Class presentation (10%)
  - Class lead discussant (10%)
  - Discussion participation (10%)
- Project (50%)
  - Project literature review (10%)
  - Project written proposal and presentation (15%)
  - Project final presentation and paper (25%)
- Anonymous course feedback (5%)

Milestones


1964
ELIZA
provide a parody of "the responses of a non-directional psychotherapist in an initial psychiatric interview" and to "demonstrate that the communication between man and machine was superficial".
2011
IBM Watson
a question-answering computer system capable of answering questions posed in natural language
2014
Amazon Alexa
capable of voice interaction, music playback, making to-do lists, setting alarms, streaming podcasts, playing audiobooks, and providing weather, traffic, sports, and other real-time information, such as news.
2022
Cicero
the first AI to play at a human level in Diplomacy, a strategy game that requires building trust, negotiating and cooperating with multiple players.
2022
ChatGPT
a conversational AI using deep learning for generating human-like responses, enabling interactive, dynamic, and context-aware conversations with users.
2023
BingChat
the search engine uses conversational AI technologies to enhance user experience and understand user queries.

A taxonomy Conversational Agent

Type	Focus	Typical sessions	Examples
Virtual companions	Broad, deep	10 to 100's of exchanges	ELIZA, Cleverbot, Tay, Xiaoice, Zo, Hugging Face
Intelligent assistants	Broad, shallow	1-3 exchanges	Siri, Cortana, Alexa, Google Assistant, Bixby
Task-focused chatbots	Narrow, shallow	3-7 exchanges	Dom the Domi-nos Pizza Bot, customer service bots, Russian trolls, non-player characters

Grudin, J., & Jacques, R. (2019, May). Chatbots, humbots, and the quest for artificial general intelligence. In Proceedings of the 2019 CHI Conference on Human Factors in Computing Systems (pp. 1-11).

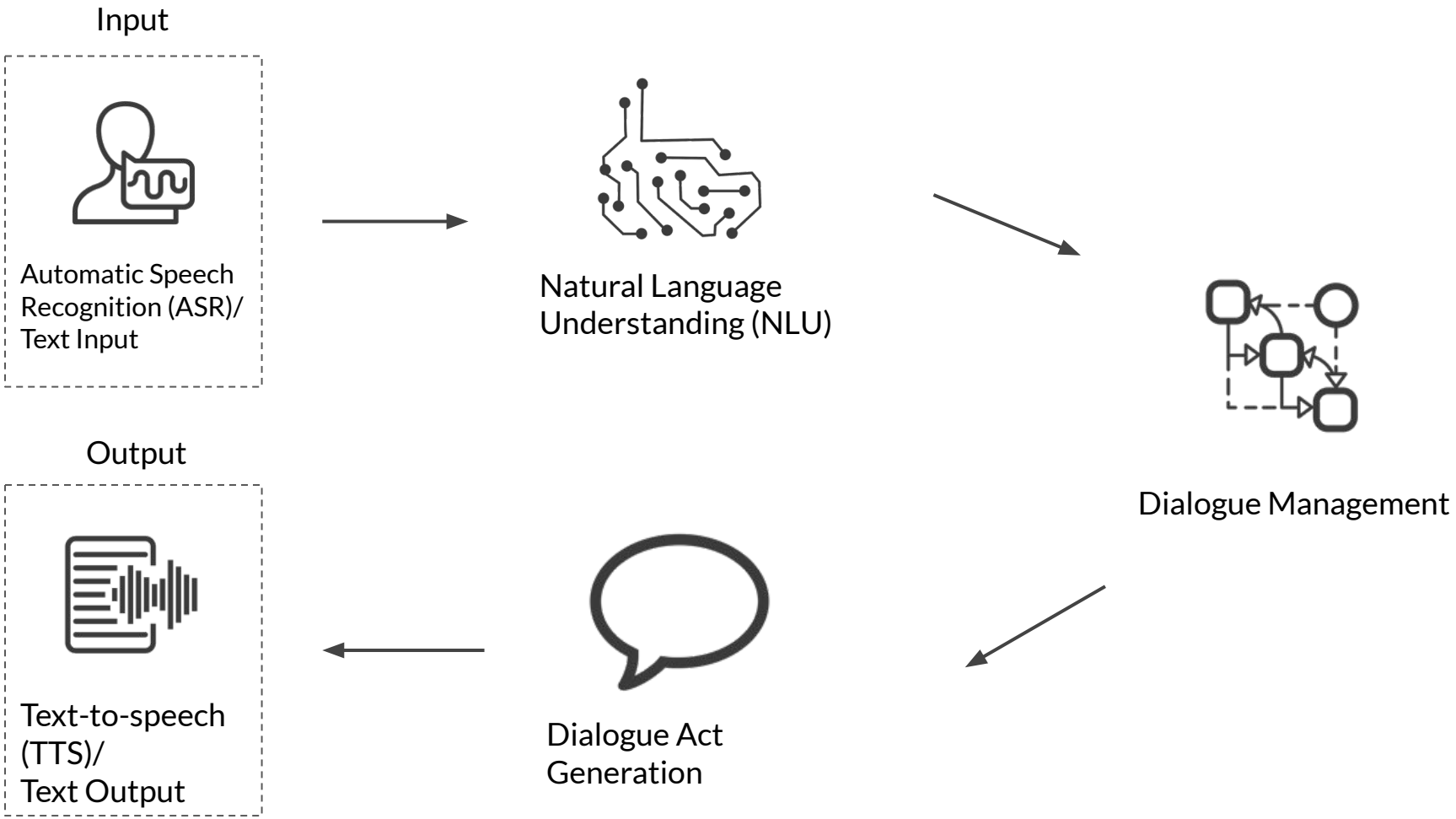
Does ChatGPT fit into any of these categories?

Type	Focus	Typical sessions	Examples
Virtual companions	Broad, deep	10 to 100's of exchanges	ELIZA, Cleverbot, Tay, Xiaoice, Zo, Hugging Face
Intelligent assistants	Broad, shallow	1-3 exchanges	Siri, Cortana, Alexa, Google Assistant, Bixby
Task-focused chatbots	Narrow, shallow	3-7 exchanges	Dom the Domi-nos Pizza Bot, customer service bots, Russian trolls, non-player characters

How do you evaluate Conversational agents like ChatGPT?

Under the hood.

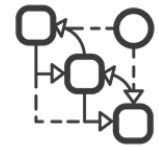
What makes a conversational agent?


Input
Automatic Speech Recognition (ASR)/ Text Input
Natural Language Understanding (NLU)
Dialogue Management
Dialogue Act Generation
Output
Text-to-speech (TTS)/ Text Output

Input

“I want a sandwich with ham and cheese”

TakeOrder{
Type= sandwich,
Topping: [ham],
Extra: [cheese]
}



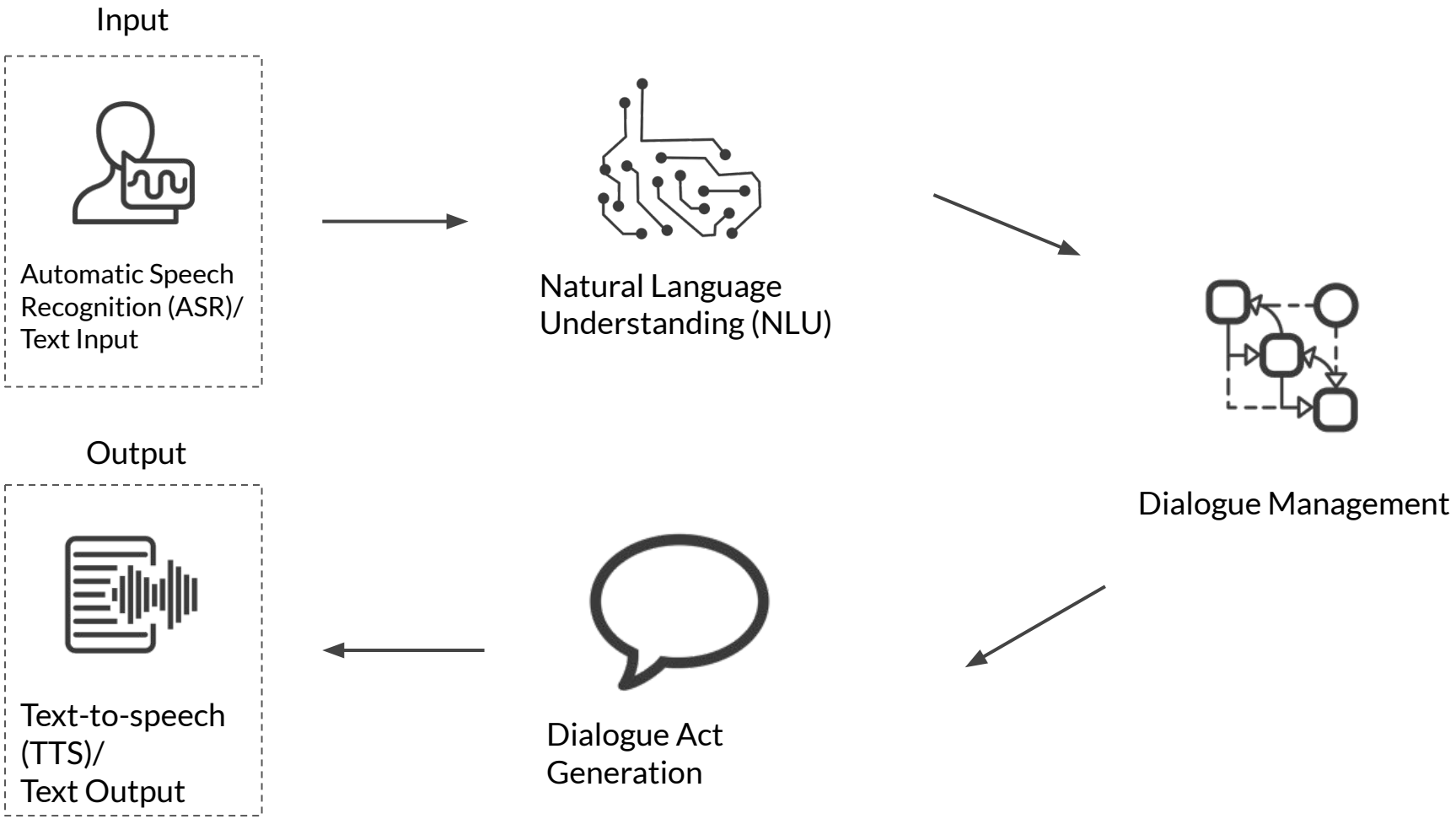
TakeOrder{
Type= sandwich,,
Topping: [ham],
Extra: [cheese],
Toasted: [?]
}

Ask if the customer wants the sandwich to be toasted

Output

“I got you. Do you want your sandwich toasted?”

Can you recall moments when the conversational agent fails?

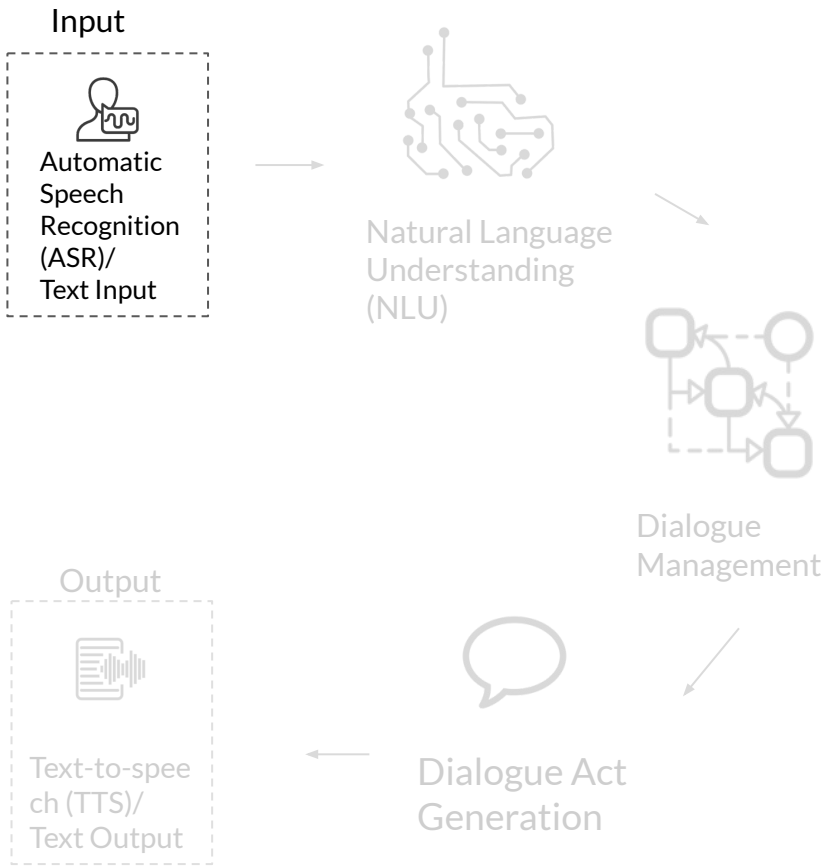

Input
Automatic Speech Recognition (ASR)/ Text Input
Natural Language Understanding (NLU)
Dialogue Management
Dialogue Act Generation
Output
Text-to-speech (TTS)/ Text Output

no perfect conversational UI?

Automatic Speech Recognition

Voice signal to computer recognizable info

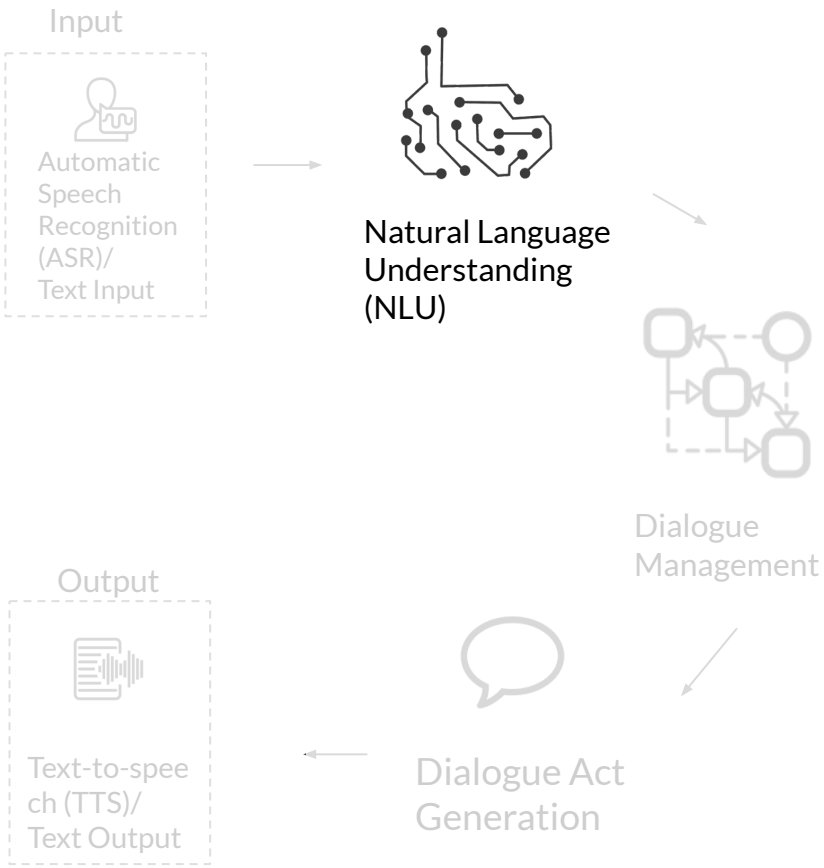
- Lack of knowledge
- Multilingualism and dialects
- Peripheral background sounds
- Dialogue with multiple speakers



Natural Language Understanding

Understand human language

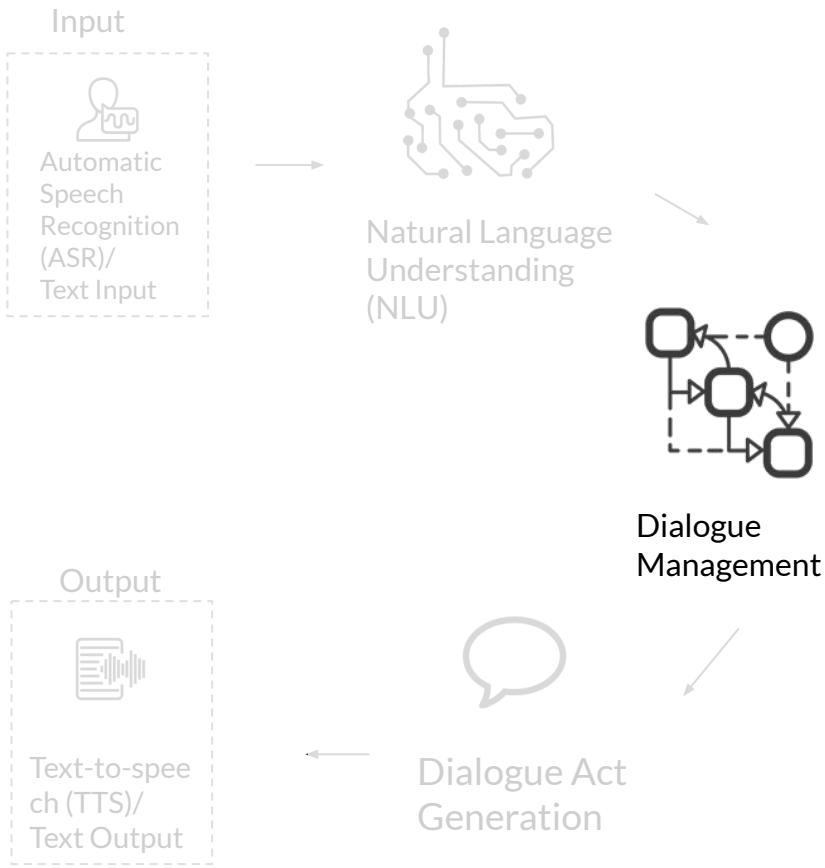
- Diverse and complicated inputs
- Vague human language
- Data are expensive
- Difficult to scale



Dialogue Management

Manage conversation states and Inform next action

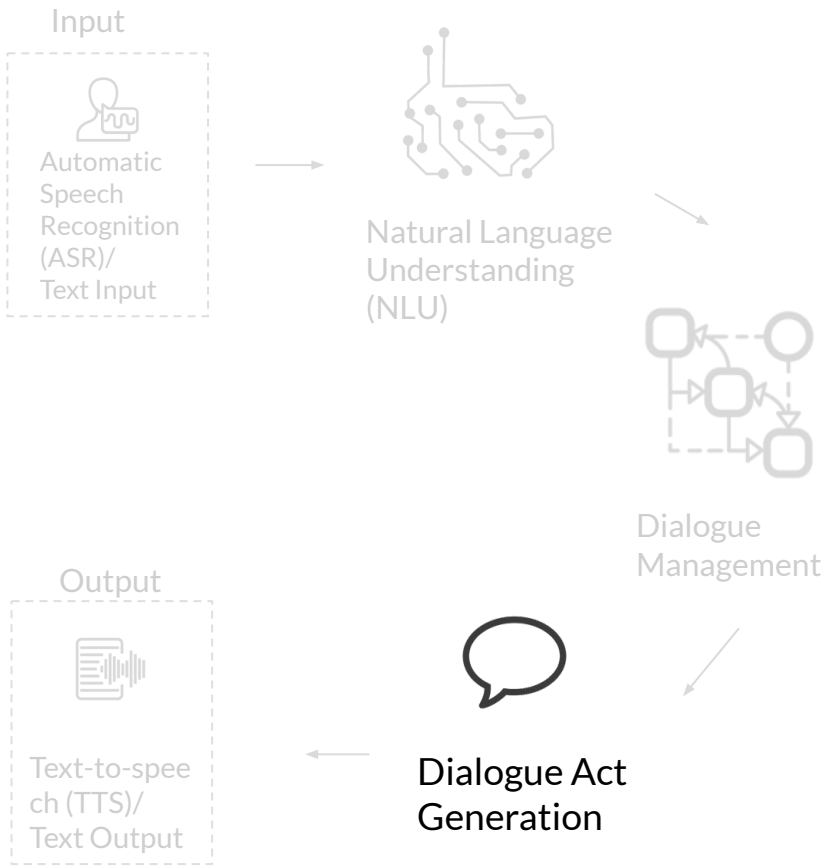
- Nonlinear conversations
- Context matters
- Open-domain
- Conversation strategy



Dialogue Act Generation

Convert action to natural language

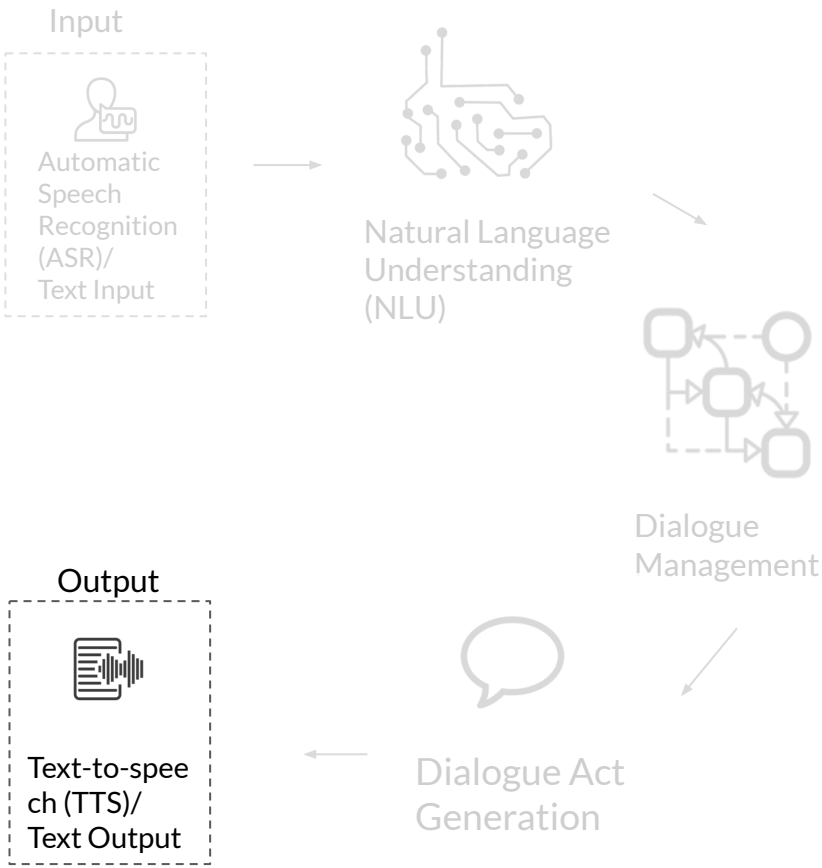
- Large action space
- Personalization
- Coherent experience
- Difficult to evaluate



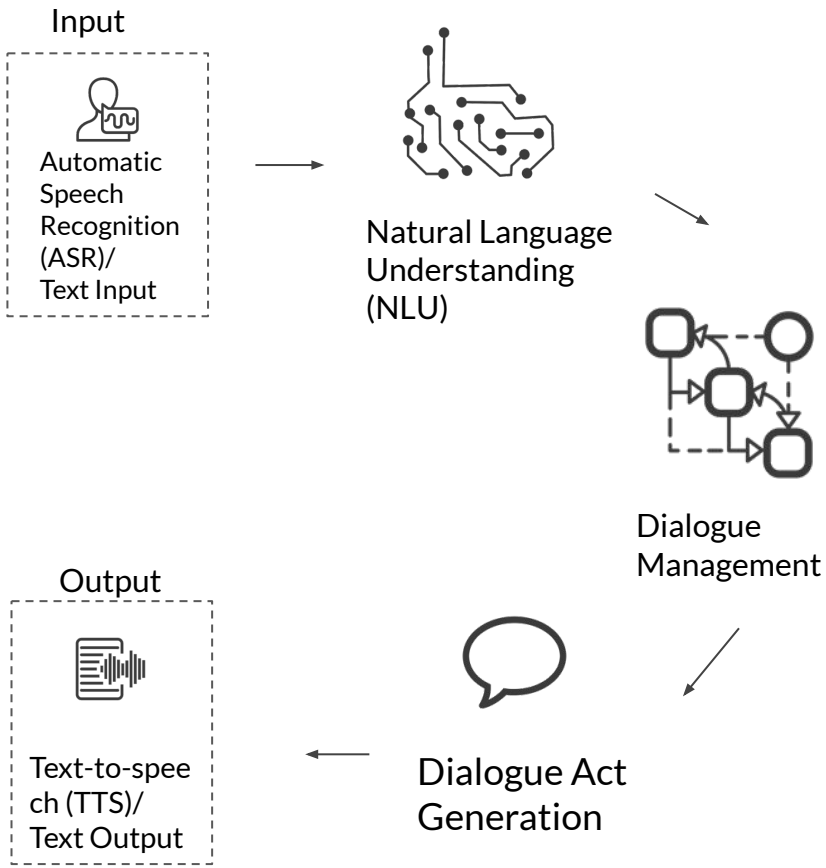
Text-to-Speech

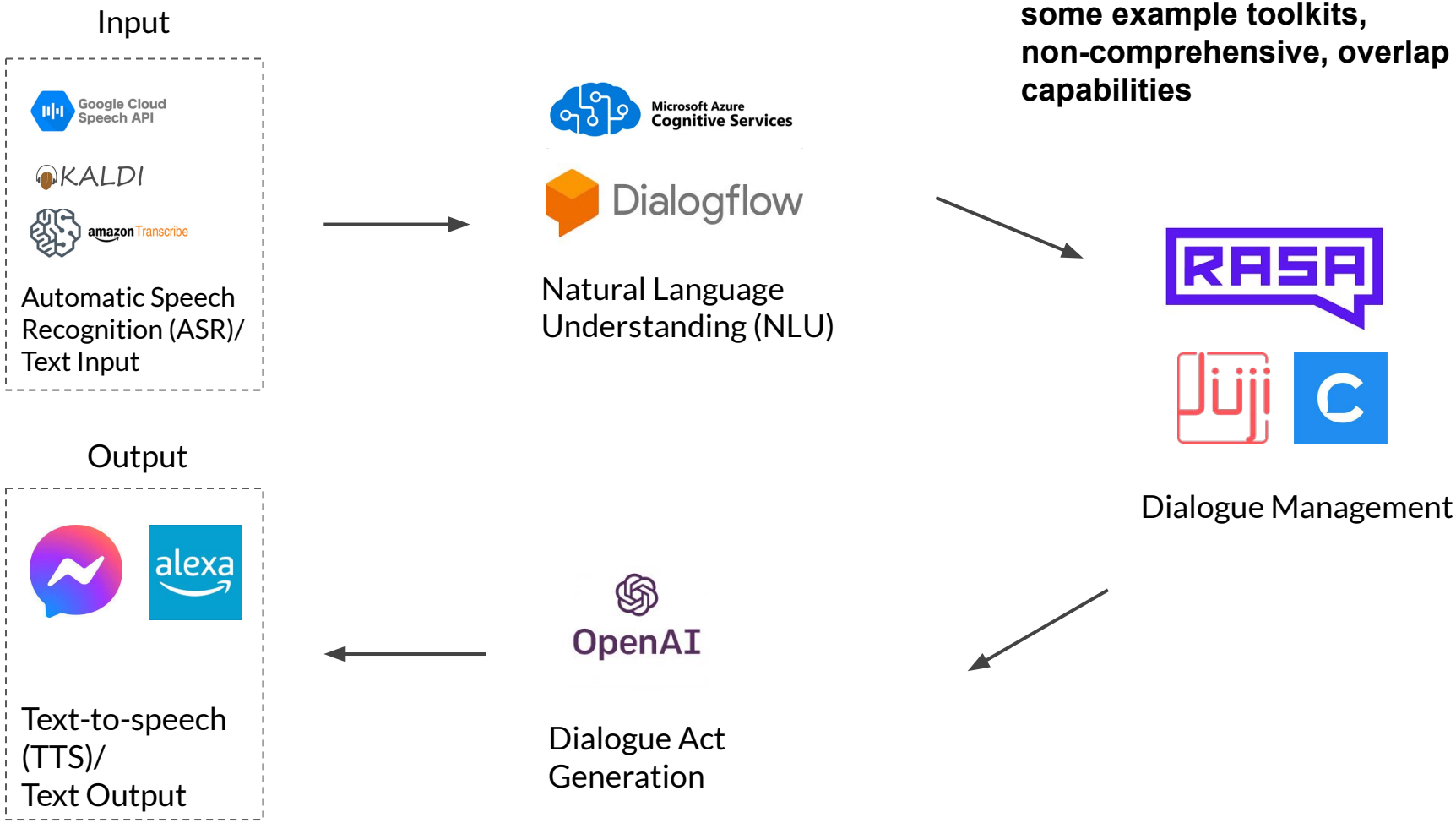
Deliver the final utterance

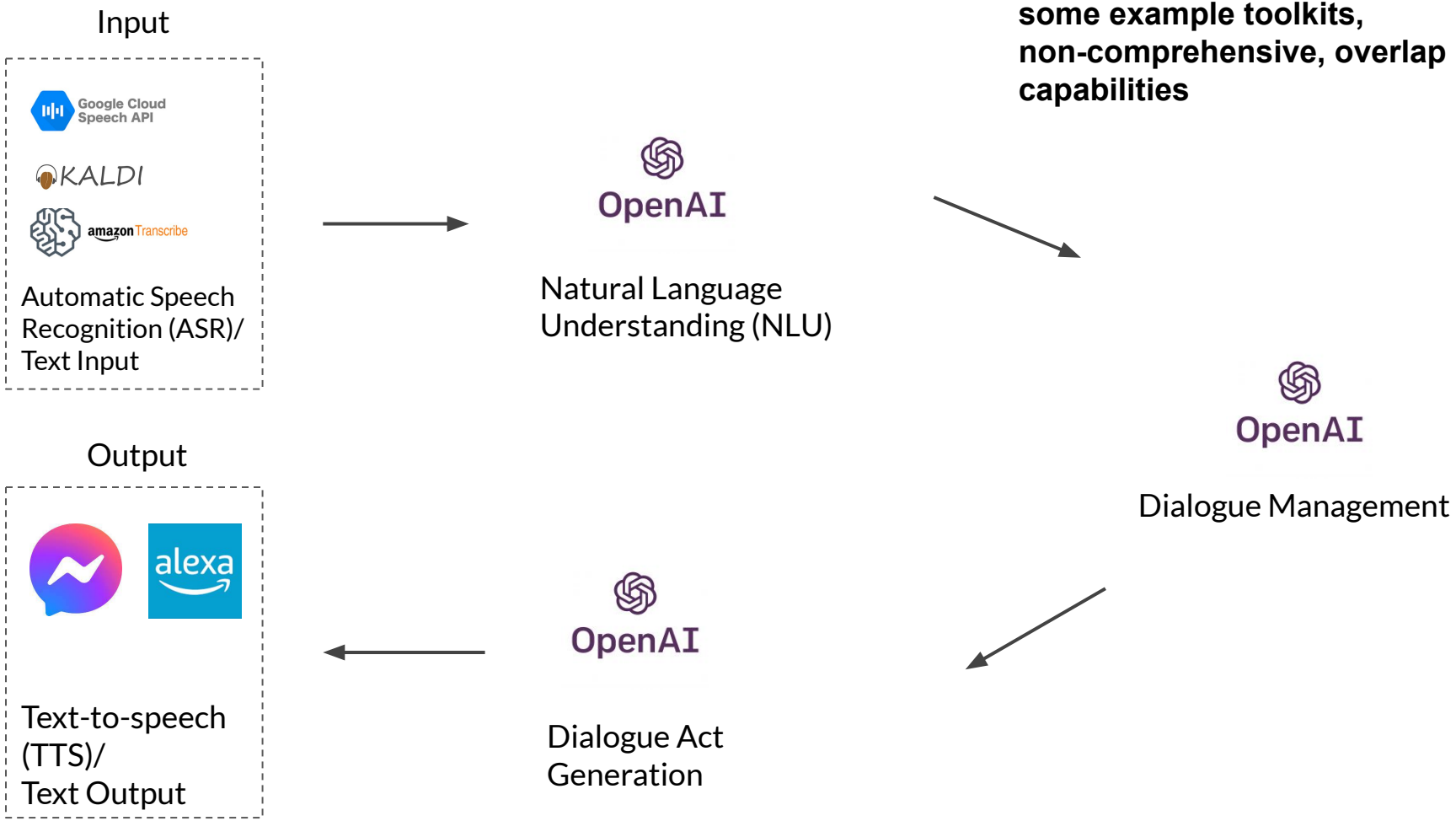
- Design Choice
- Data
- Prosody issues
- Uncanny valley



Error Cascades!




Input
Google Cloud Speech API
KALDI
amazon Transcribe
Automatic Speech Recognition (ASR)/ Text Input
Microsoft Azure Cognitive Services
Dialogflow
Natural Language Understanding (NLU)
some example toolkits, non-comprehensive, overlap capabilities
RASA
Dialogue Management
Output
alexa
Text-to-speech (TTS)/ Text Output
OpenAI
Dialogue Act Generation


Input
Google Cloud Speech API
KALDI
amazon Transcribe
Automatic Speech Recognition (ASR)/ Text Input
OpenAI
Natural Language Understanding (NLU)
some example toolkits, non-comprehensive, overlap capabilities
OpenAI
Dialogue Management
OpenAI
Dialogue Act Generation
Output
alexa
Text-to-speech (TTS)/ Text Output

Beyond technical components

Language Design

Intent, language style, dialect

Voice Design

Prosody, Tone, and Intonation

Multimodal Interaction Design

Actions, vision,
gestures, expressions
….


What can I help you with?
Book a train ticket to Beijing on National Day.
What's your departure city?
Hefei.