

Conversational UIs

Module 1: Introduction to Conversational UIs

Dr. Fan Zhenzhen
Institute of Systems Science
National University of Singapore
Email: zhenzhen@nus.edu.sg

About this course

Objective: to learn skills to design and implement systems that can interact with users using spoken or written natural language, like chatbots and virtual assistants.

At the end of this module, you can:

- Determine the roles that systems with conversational UI can play in fielded applications
- Identify and analyse the main components and the architectures of conversational interfaces
- Design conversational UI following practical methodology and strategies
- Develop applications with conversational UI using traditional and machine learning approaches.
- Evaluate the performance of the conversational UI using appropriate metrics

Course Lecturers



Dr. Fan Zhenzhen



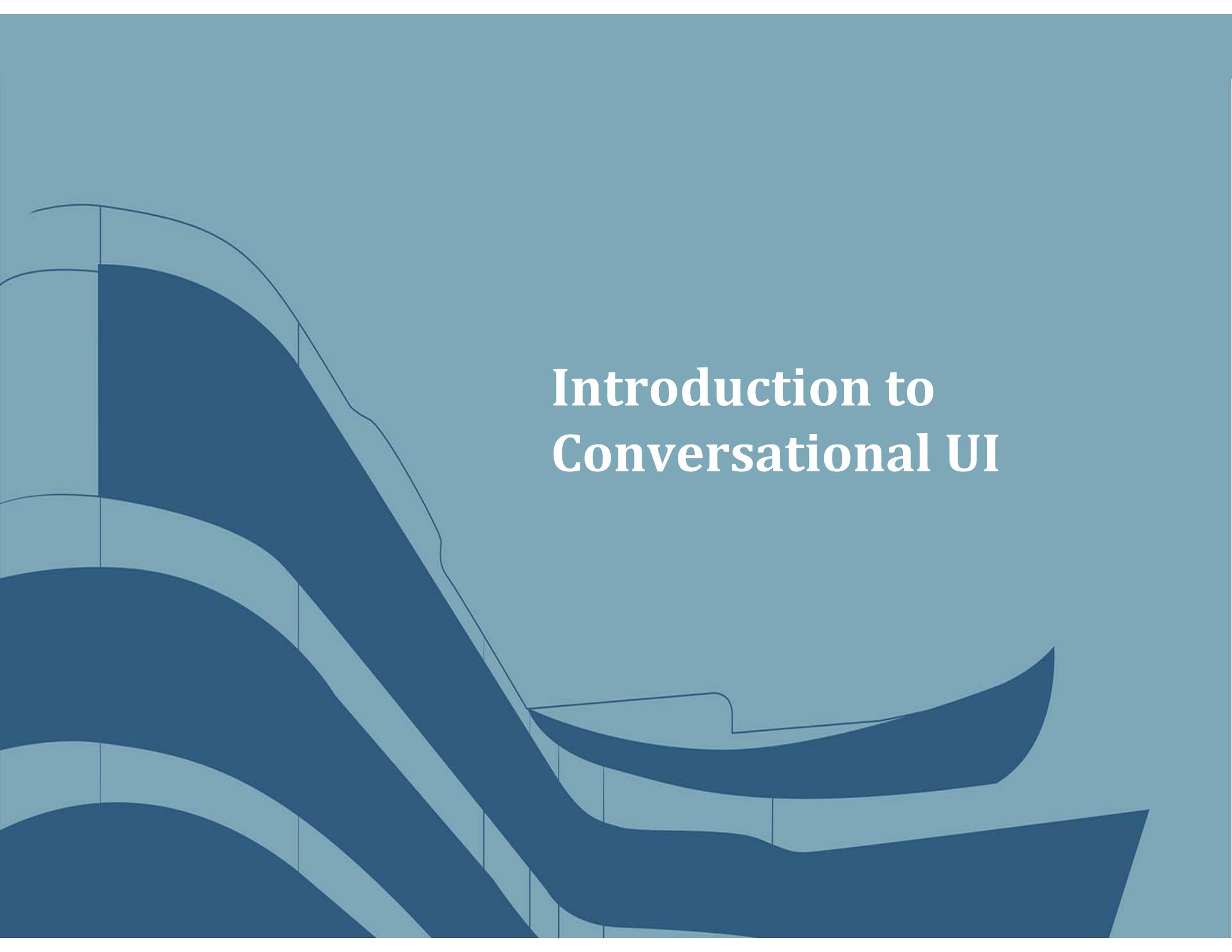
Dr. Wang Aobo



Dr. Gary LEUNG

Course Agenda

Day 1	Introduction to Conversational UIs Conversation Design Evaluating Conversational UI
Day 2	Understanding the Content of User's Utterances Response Generation
Day 3	Speech processing basics Speech recognition (Speech-to-Text) Speaker diarization
Day 4	Speech synthesis (Text-to-Speech) Voice conversion Spoken dialogue system

A large, abstract graphic on the left side of the slide features several thick, wavy lines in varying shades of blue. These lines form a dynamic, flowing pattern that resembles water or a stylized landscape. The lines are layered, with some being solid dark blue and others being lighter, semi-transparent shades.

Introduction to Conversational UI

Conversational UI - a new paradigm of user interface

••••• Verizon 4:07 PM

Domino's Pizza > Typically replies instantly

Home

Domino's Pizza 16M people like this Food & Beverage Comp.

4:07 PM

at, Jamie can provide useful answers

Get Started

Hi, this is Dom, the Domino's ordering assistant bot. How can I help, Ana?

New Order

Reorder

Track Order

Type a message...

ASK JAMIE:

Ask CPF

Have a question on our C

These are the current Highlights & Most Popular Questions

Highlights

- Is CPF Board bound by Personal Data Protection Act (PDPA)?
- Isn't it against the law for CPF Board to reveal a member's personal information?
- What is CareShield Life?

Are you looking for

Can I qualify for Workfare Income Supplement (WIS) if I did not receive a tax credit from the Inland Revenue Authority of Singapore (IRAS)?

What do I need to do to receive Workfare Income Supplement (WIS)?

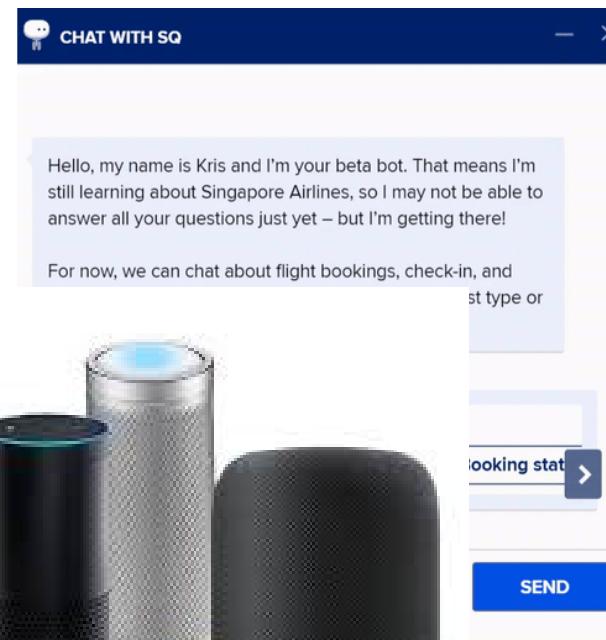
How do I calculate CPF contributions for my salary?

What are the Chinese Development Assistance Council (CDAC) contributions?

Can I top up my children's CPF account?

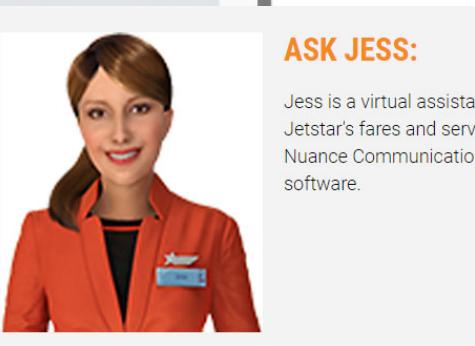
How much should I contribute?

Print  Terms of Use Powered by flexAnswer™



ASK JESS:

Jess is a virtual assistant for Jetstar's fares and service. She was developed by Nuance Communications, software.



Emirates Vacations

Hi.

Chat with me about your next vacation

Explore the world without leaving the page

Tell me about your trip...

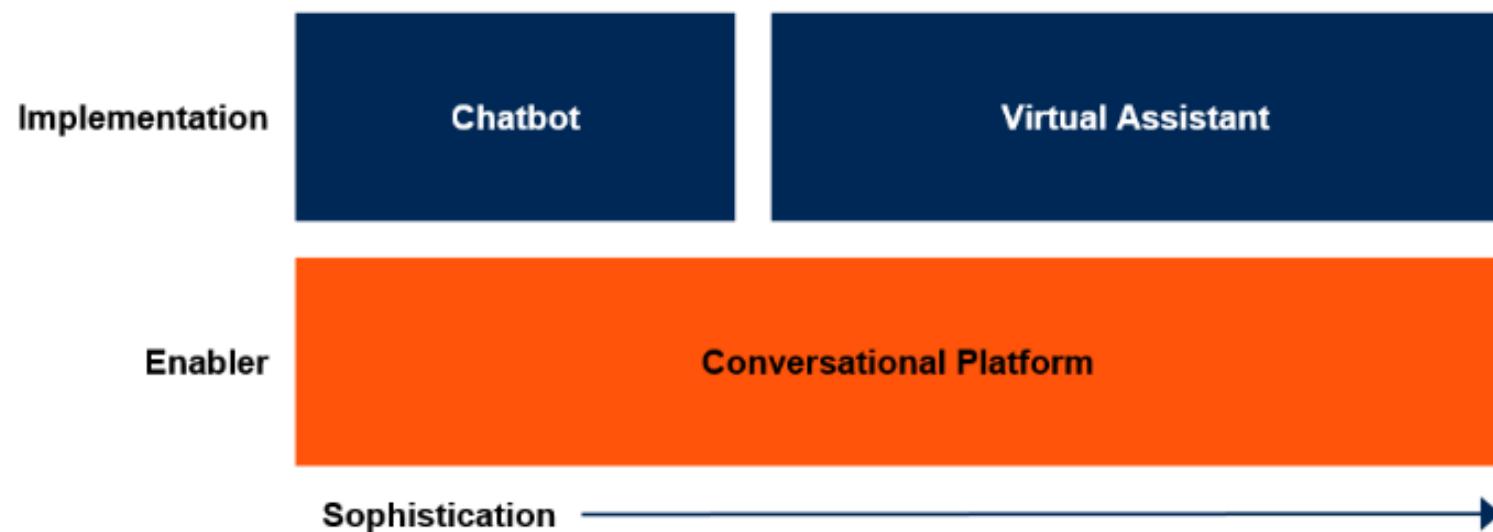
Chat

Please do not enter personal information.

Chatbots or ...

- CUIs? Virtual Assistants? Conversational Agents?
Dialogue Systems? Dialogue Agents?

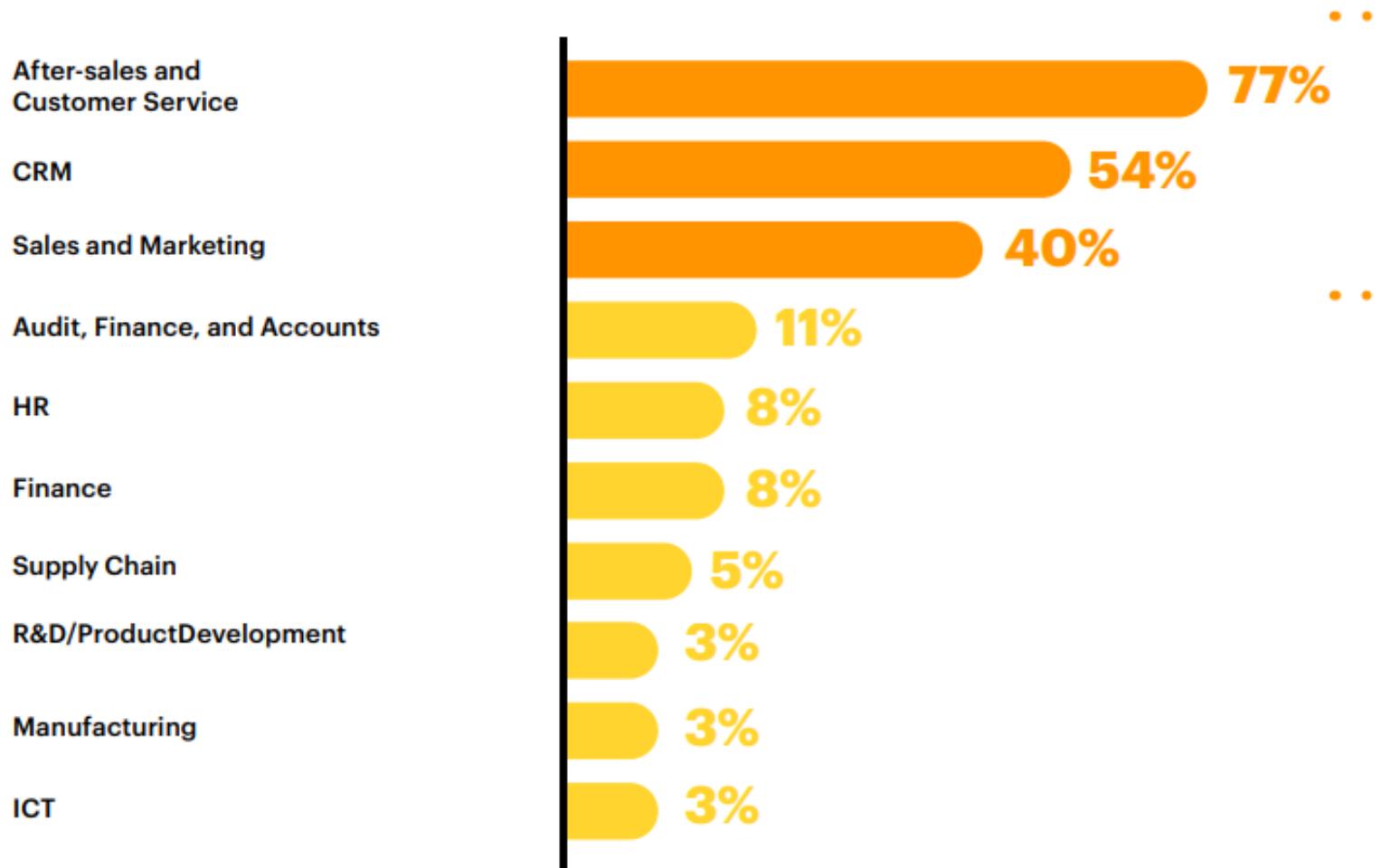
Market Definition



Source: Gartner
ID: 367775

Where are they used?

ALREADY IMPLEMENTED



Chatbots Are Here to Stay by Accenture Research
https://www.accenture.com/_acnmedia/pdf-77/accenture-research-conversational-ai-platforms.pdf

Popular Use-cases

- Customer service
 - to take over inquiries coming into the customer service desk, reducing the need for human agents.
- IT service desk
 - to automate parts of the IT service desk to more quickly and effectively solve routine IT problems and/or reduce the need for IT support staff.
- HR
 - to automate routine questions and queries coming into HR. Examples include vacation time, entitlements, hour tracking, overtime pay and rules/regulations in the workplace.
- Sales support
 - to support salespeople in their work by giving them support in the sales cycle.
- Commerce
 - to offer sales support to customers at point of sale or upsale in relevant situations.
- Marketing
 - as part of an advertising campaign or to support marketing efforts.
- Enterprise software front ends
 - as conversational interfaces, making an alternative UI for enterprise software.
- Advisory services
 - to give advice by collecting relevant information through conversation.

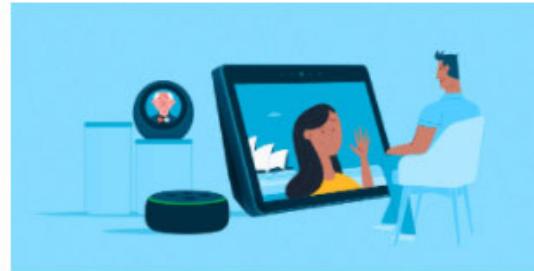
Common Classes

- **Informational**
 - Handle inquiries
 - Provide customer-and context-specific results that can be accessed via voice, text, or visuals
- **Transactional**
 - Help customers perform activities like booking tickets, ordering food, managing accounts, etc.
- **Enterprise productivity**
 - Connect to enterprise data resources, streamline enterprise work activities, and improve efficiencies
 - Check sales numbers, campaign performances, monitor inventory status, schedule meetings, etc.
- **Device control**
 - Enable connected devices such as wearables, home appliances, and vehicles to interact with each other

Summary from Alexa



Entertainment



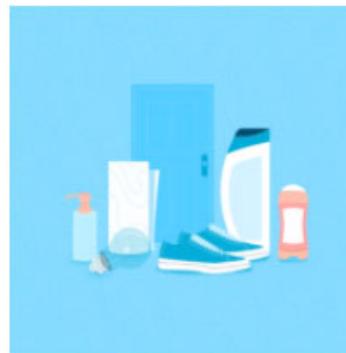
Communication



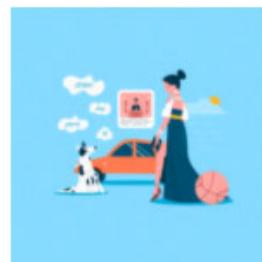
Productivity



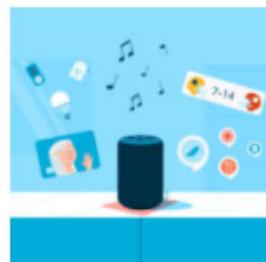
Smart Home



Shopping



News & Information



Skills



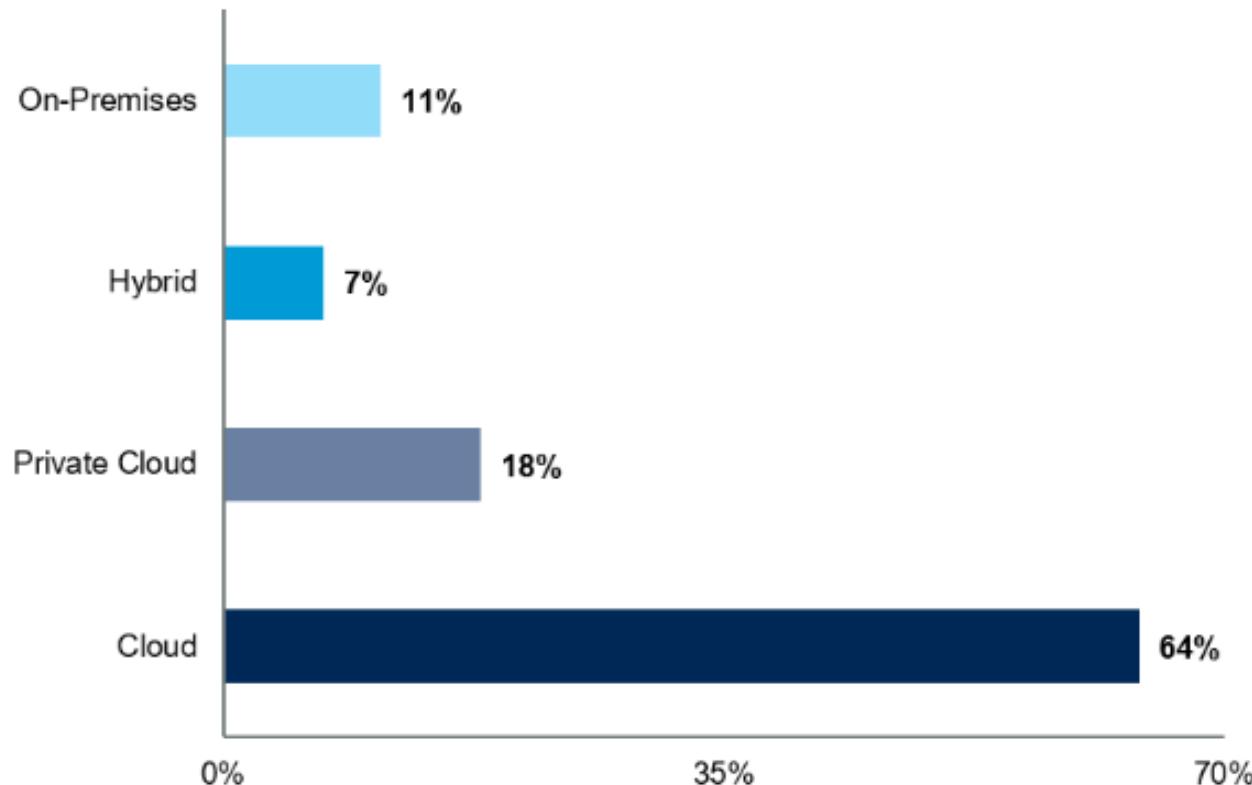
Using Multiple Devices with Alexa



Things You Can Ask Alexa

Choices of Deployment

Deployment Options
Percentage of Respondents

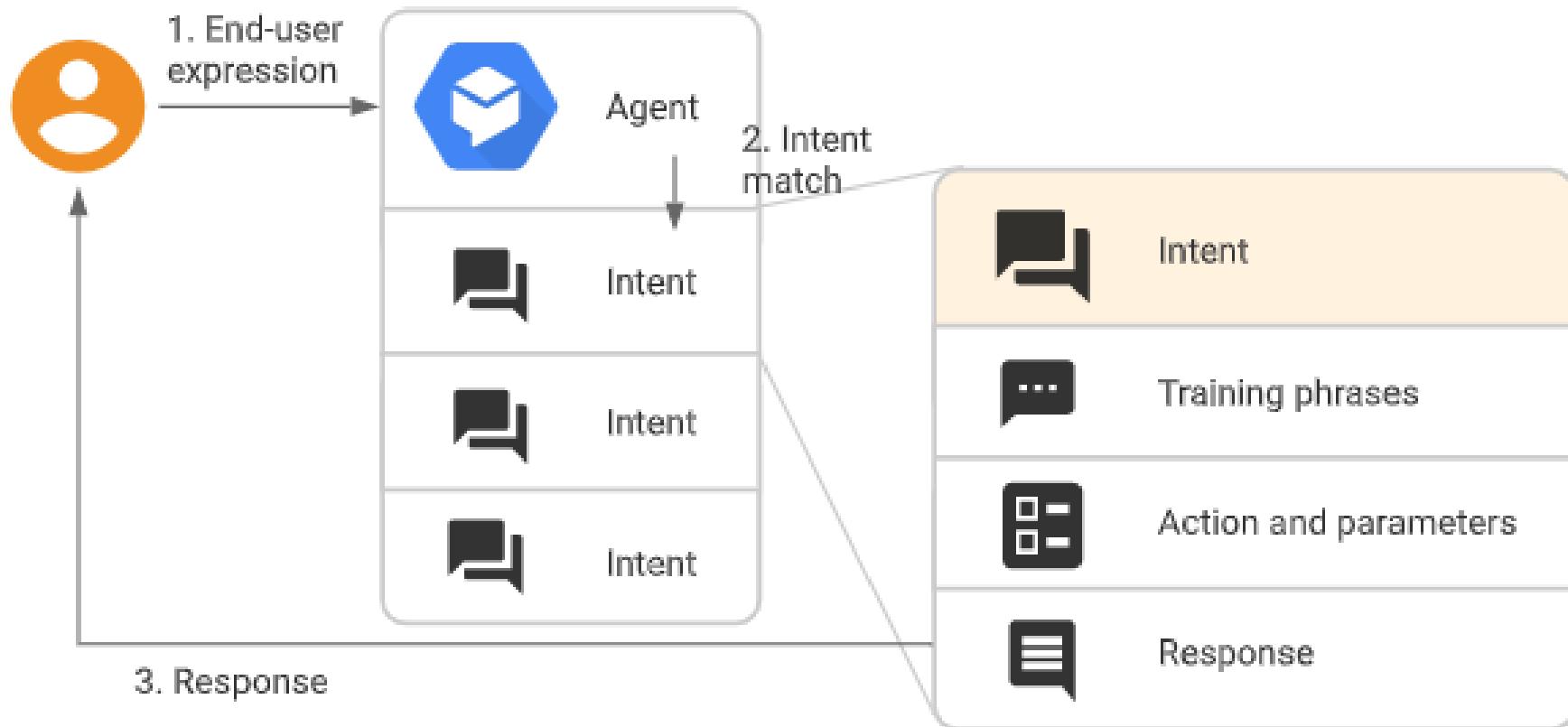


Source: Gartner (July 2019)
ID: 349067

Main concepts of Conversational UI

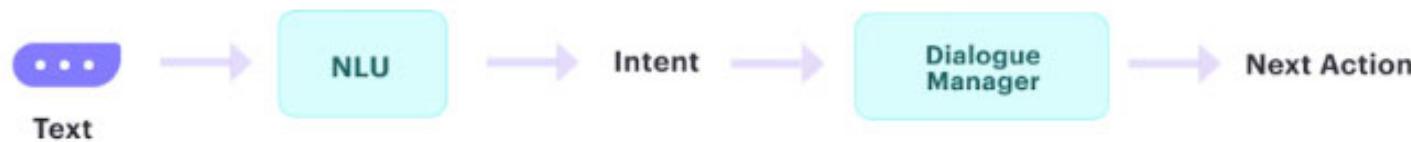
- **Intents** - an end-user's intention for one conversation turn
- Parameters (**slots**) – values or entities relevant for an intent, extracted from user's expressions
- **Actions** – to be triggered when an intent is matched
- **Responses** – text, speech, or visual responses to be returned to the user
- **Contexts** – information needed to correctly match an intent
- **Events** – happenings, may be used to invoke an intent

Main Concepts



Common Architectures of Dialog Systems

- Modular systems – different components taking care of separate functions, e.g. status tracking, response generation, etc.; often used for task-oriented dialog systems

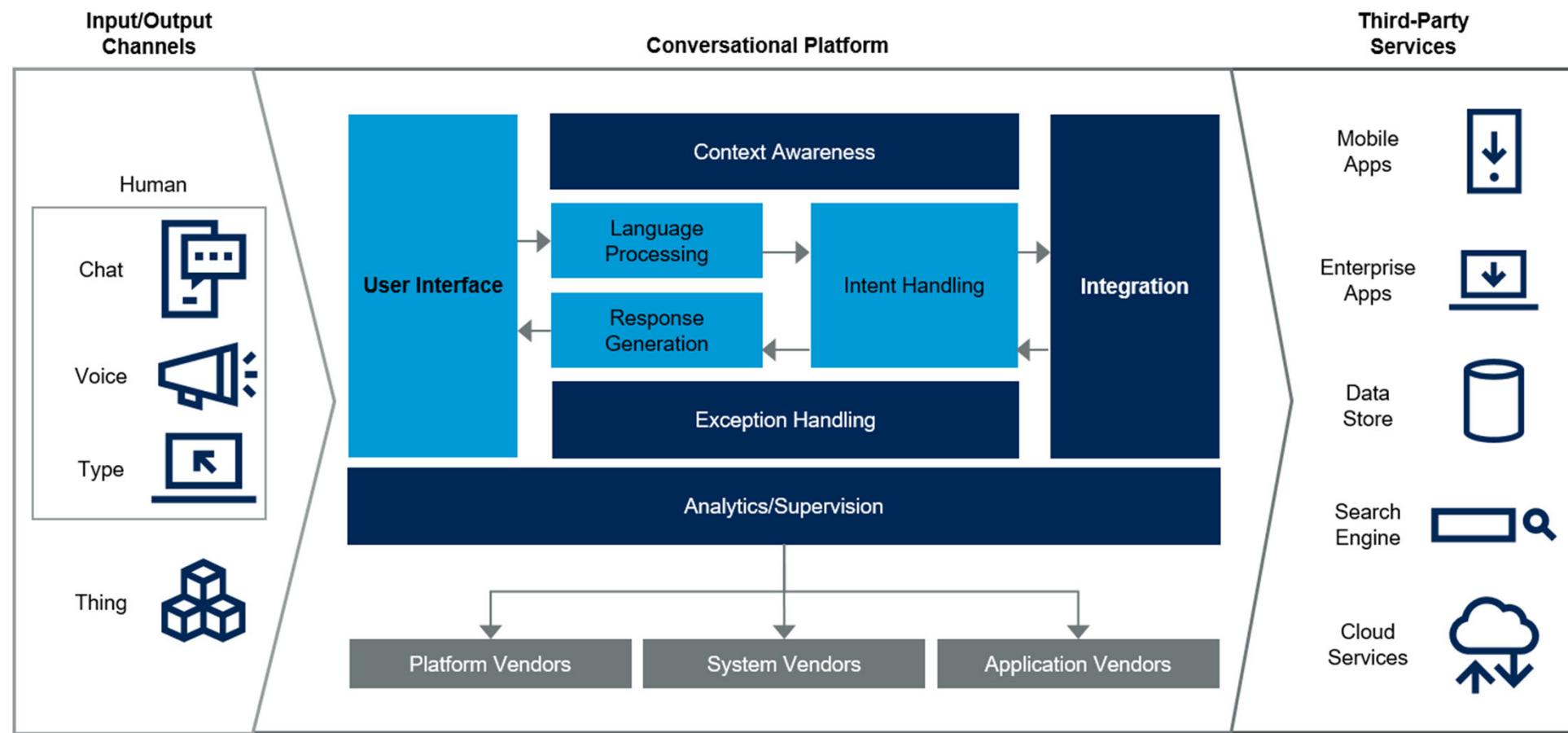


- End-to-end systems - one single model trained from a dialog dataset; often used for general conversation systems



Modular Systems

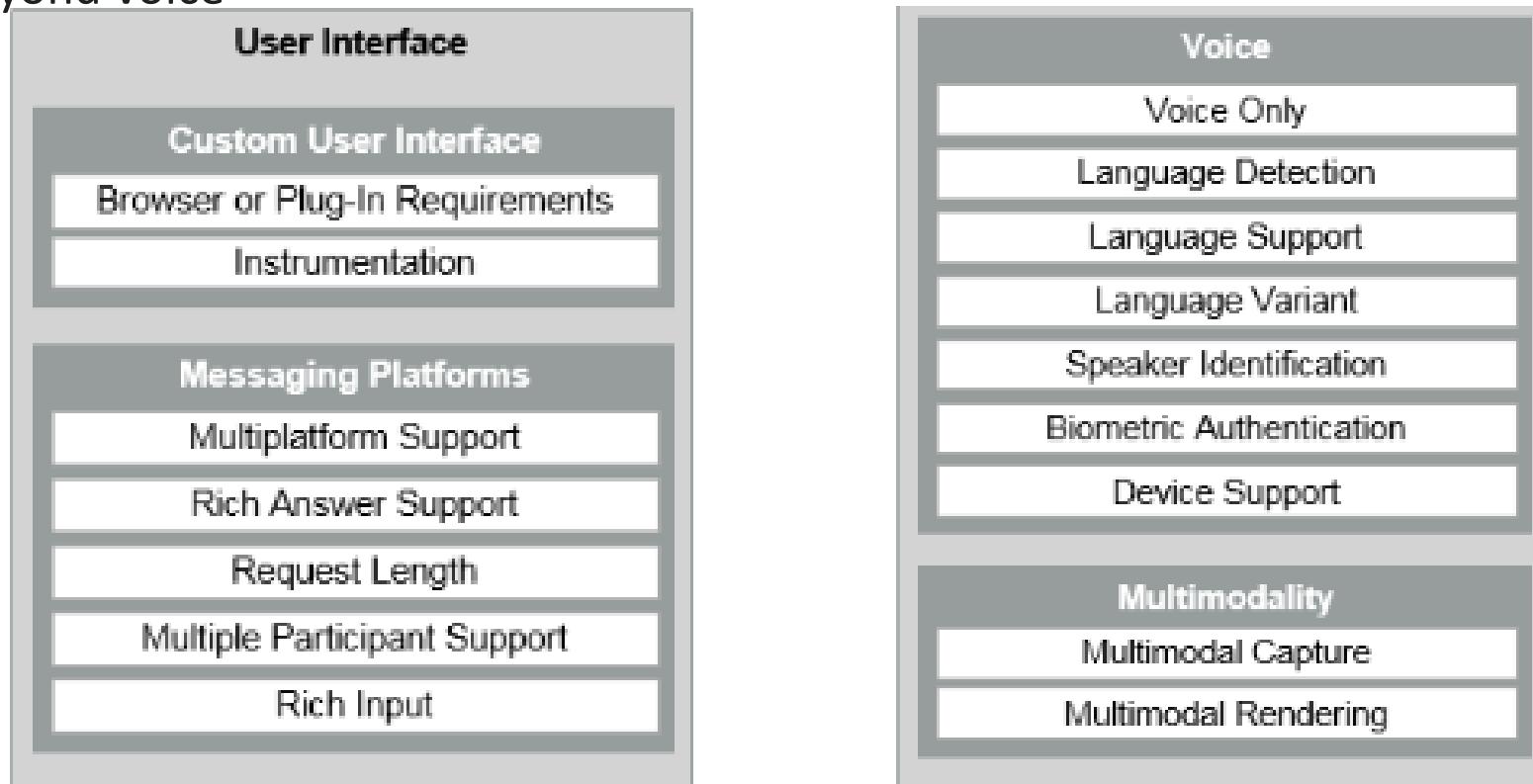
Conversational Platform Logical Architecture Platform



Source: Gartner
 ID: 349067

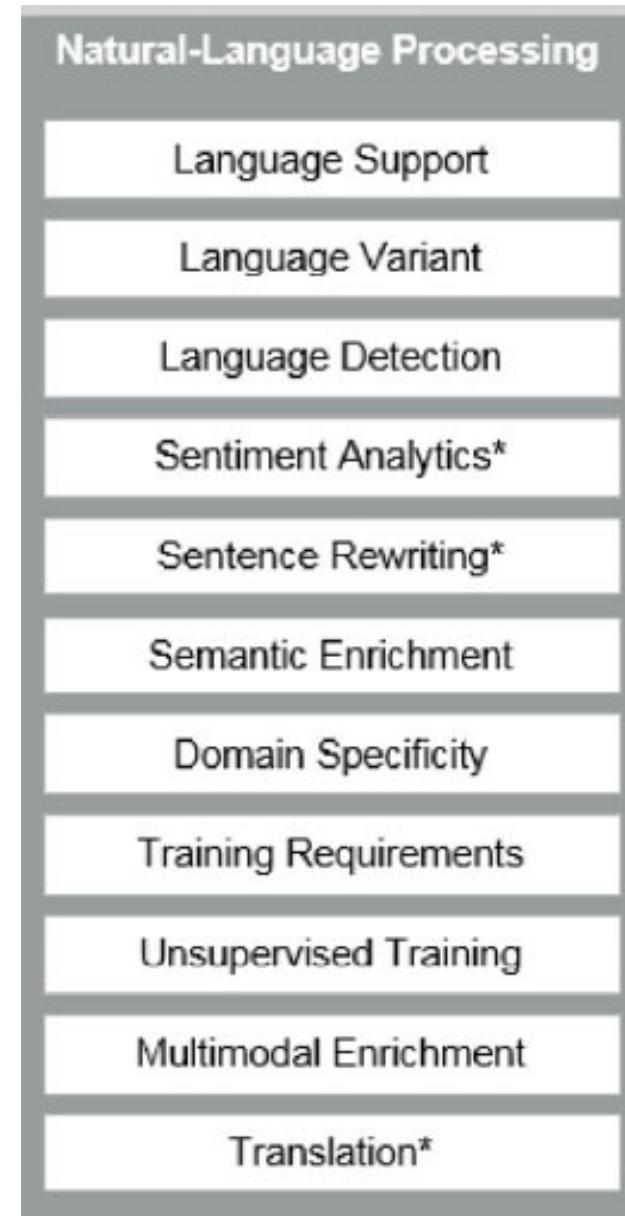
Detailed Capabilities for User Interface

- **Chat** (typed) and **voice** interfaces demand different design considerations
- Often rely on external platforms for user interaction, like FB Messenger
- Custom UI may be preferred in virtual assistant cases for tight UI control
- Effective voice interaction may require diverse language support with dialect/slang/accent awareness
- More advanced systems may require capturing and using visual and sensor data beyond voice



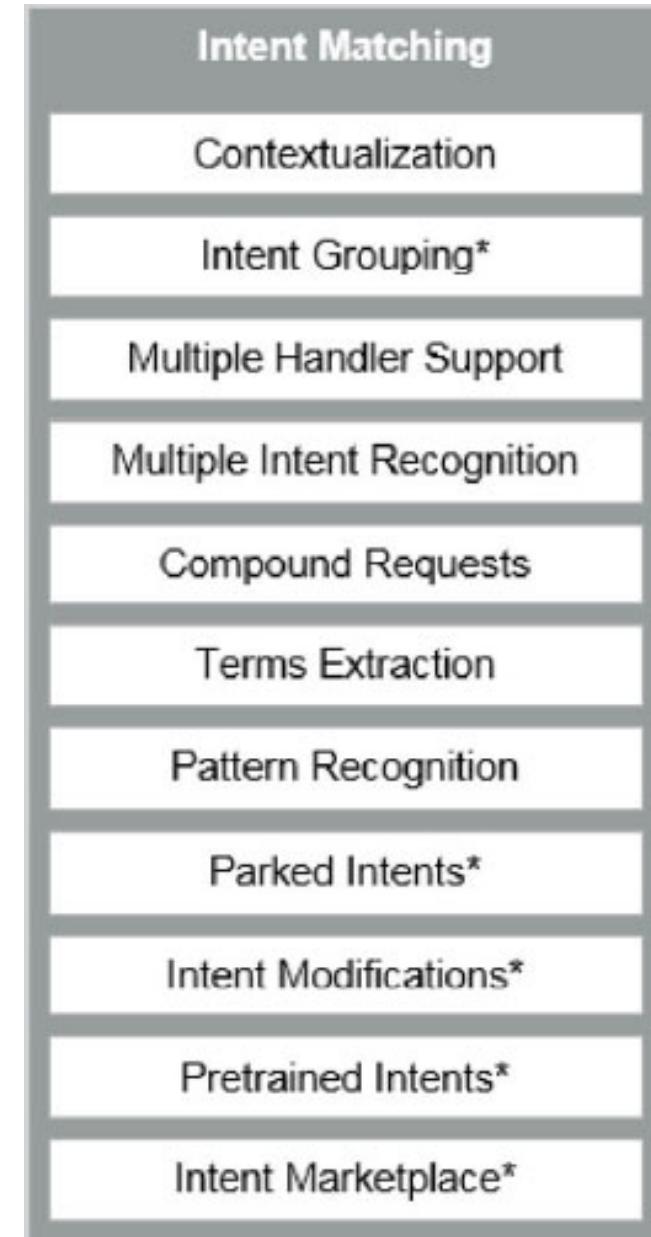
Language Processing Capabilities

- **NLP** and **intent matching** are often tightly integrated in applications
- Sentence re-writing and tagging the semantic entities in the input can improve intent matching accuracy
- The NLP engine may require industry/task specific tailoring, possibly through training
- Great to have chatbot that can learn from its interaction with user, but real-time unsupervised learning is prone to potential attacks
- Translation can increase language support, but not as good as native understanding



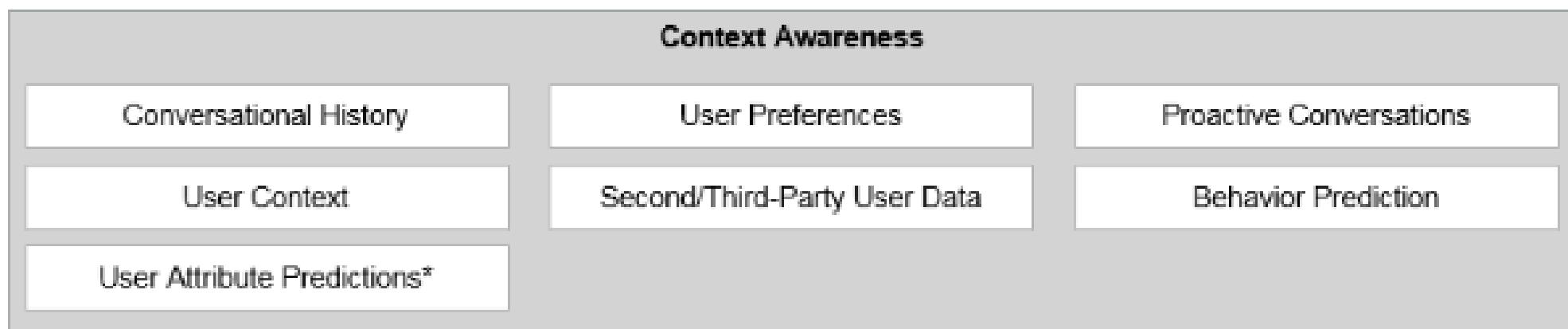
Language Processing Capabilities

- Contextual clues can improve intent matching accuracy
- Larger systems need to group intents for scalability
- An input may have more than one intents, and an intent may have multiple handlers
- Key information (slots) required by an intent need to be extracted, through pattern recognition
- User may modify an intent along the way, or bring in another intent before the existing one is handled.
- Pretrained intents shared at a marketplace can greatly shorten the development time.



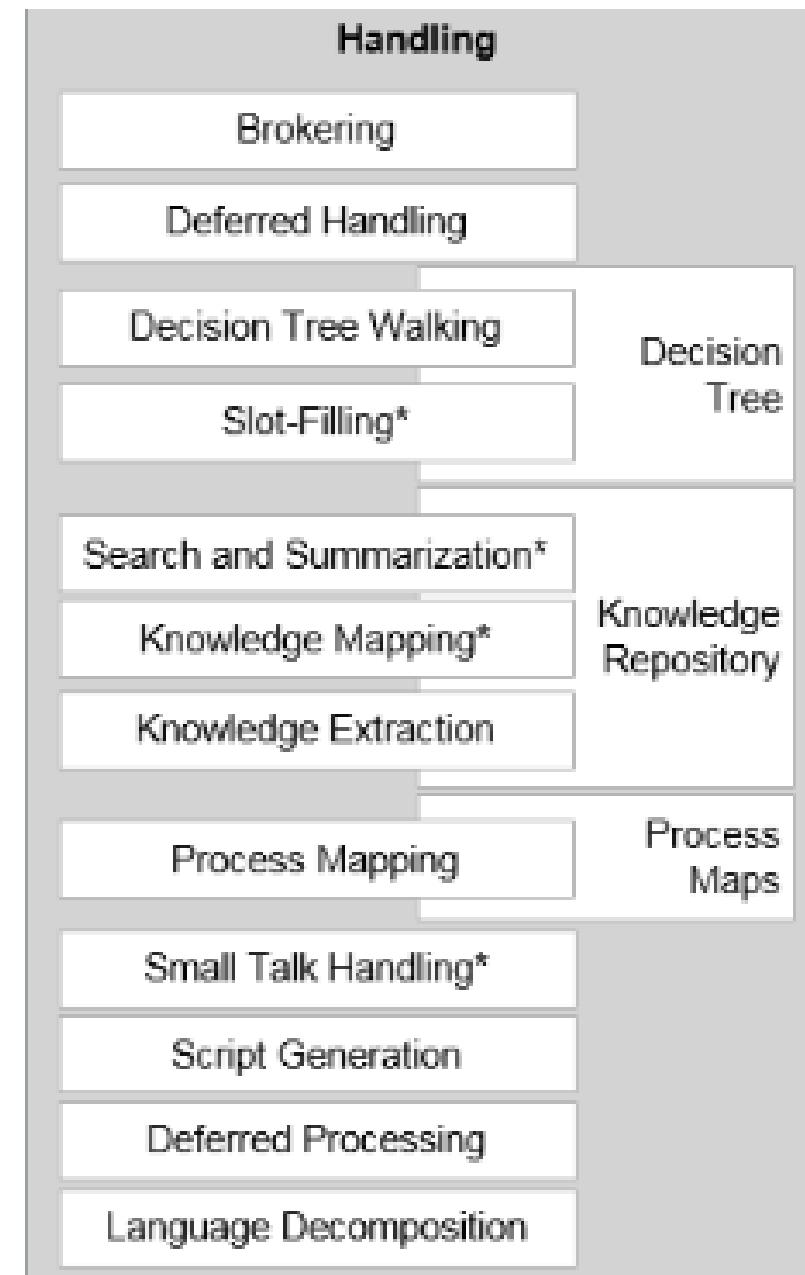
Contextual Awareness Capabilities

- Contextual information is vital for intent matching, necessary for applications with frequent use
 - Past conversation with the user
 - External information like current time, location, page, etc.
 - User preferences and profiles
- A proactive chatbot initiates dialogues based on events and rules, but balancing helpfulness and intrusiveness is crucial for successful implementation.



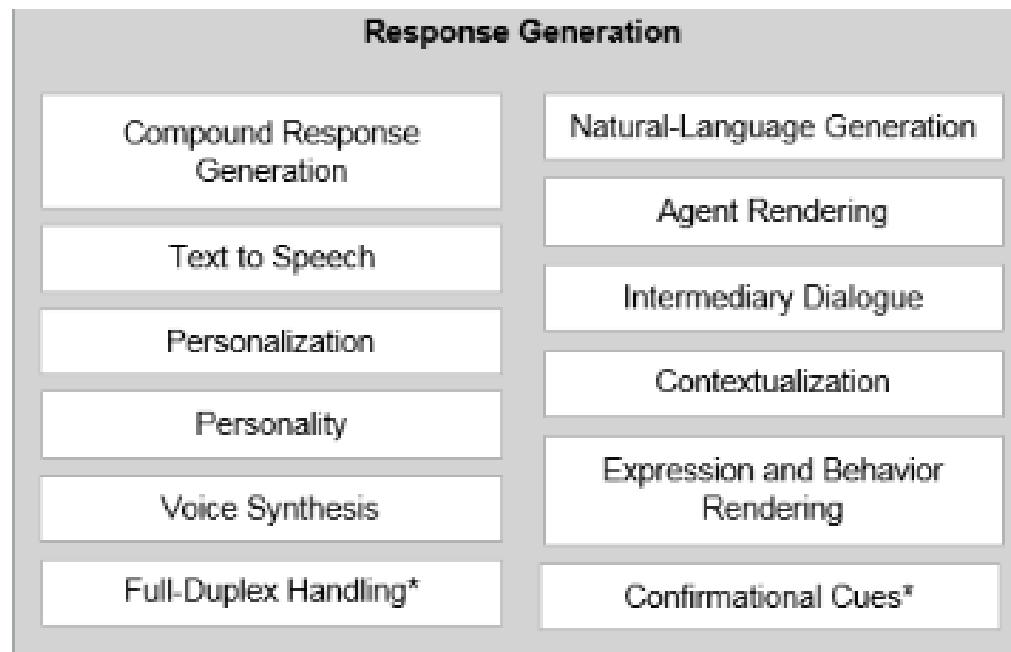
Intent Handling Capabilities

- Most common handling method – decision tree with scripted dialogue
- Brokering is needed when multiple handlers are available for one intent
- Slot filling for transactional conversation
- With knowledge repository, chatbot can take user input as a query and
 - Search for relevant results for summarization
 - Or map to a node in repository
 - Extract out answer
- Chatbots are expected to do some small talk



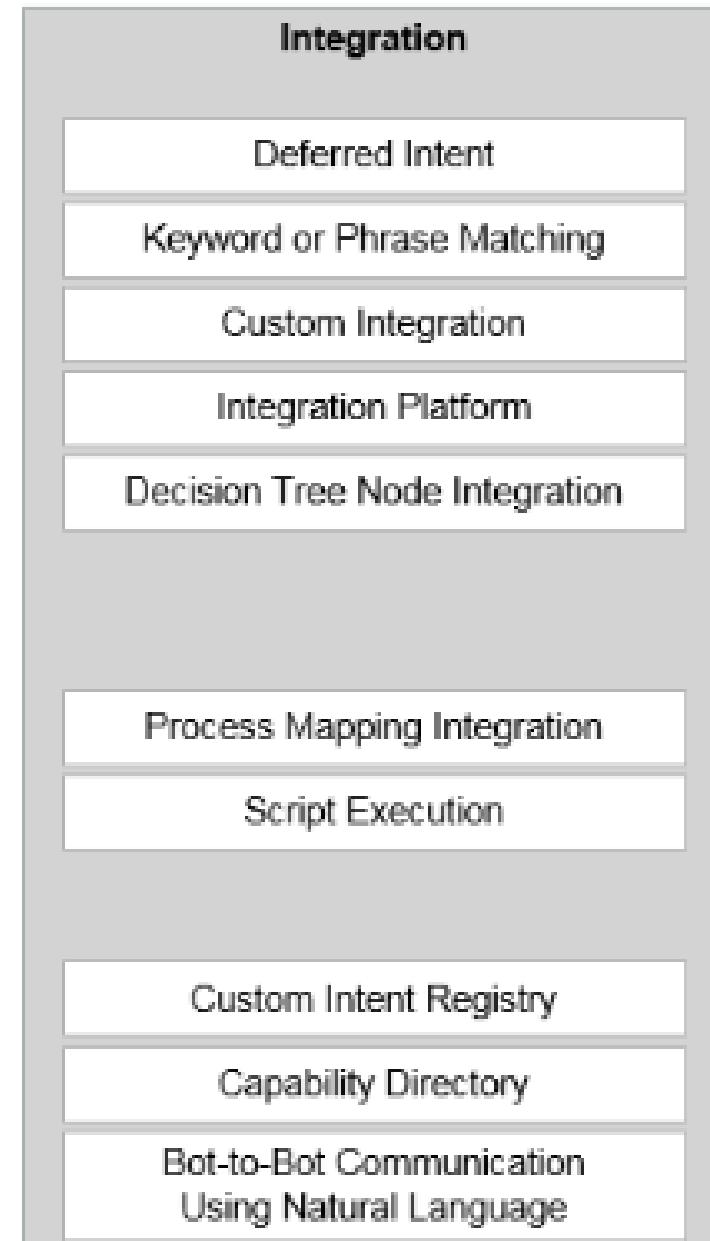
Response Generation Capabilities

- NLG, more sophisticated than pre-scripted responses
- May need to combine multiple answers into compound response
- Personalize the response based on factors like style, culture
- Convert text response to speech for voice chatbots, generating human-like voices
- Add personality to the response for brand projection or for the bot persona
- Generate additionally facial expressions, actions if an agent is used
- Generate intermediary dialogue to mitigate latency (while user waiting)



Integration Capabilities

- Although stand-alone chatbot implementations have a purpose, many use cases require integration with existing systems.
- Chatbots may even communicate with other bots using natural language.



Exceptions and Analytics Capabilities

Exception Handling

Clarifying Dialogue

Modality Switching

Search

Human Fallback

Handling error situations by

- Asking clarifying questions
- Switching to alternatives, such as website, search engine
- Pass the request to human operators

Analytics/Supervision

Analytics

Pinpoint Improvements

Propose Improvements

Quality Assurance*

- Generate reports with performance metrics
- Identify areas of improvement
- Monitor the quality of key modules to flag problems

Major Conversational Platforms

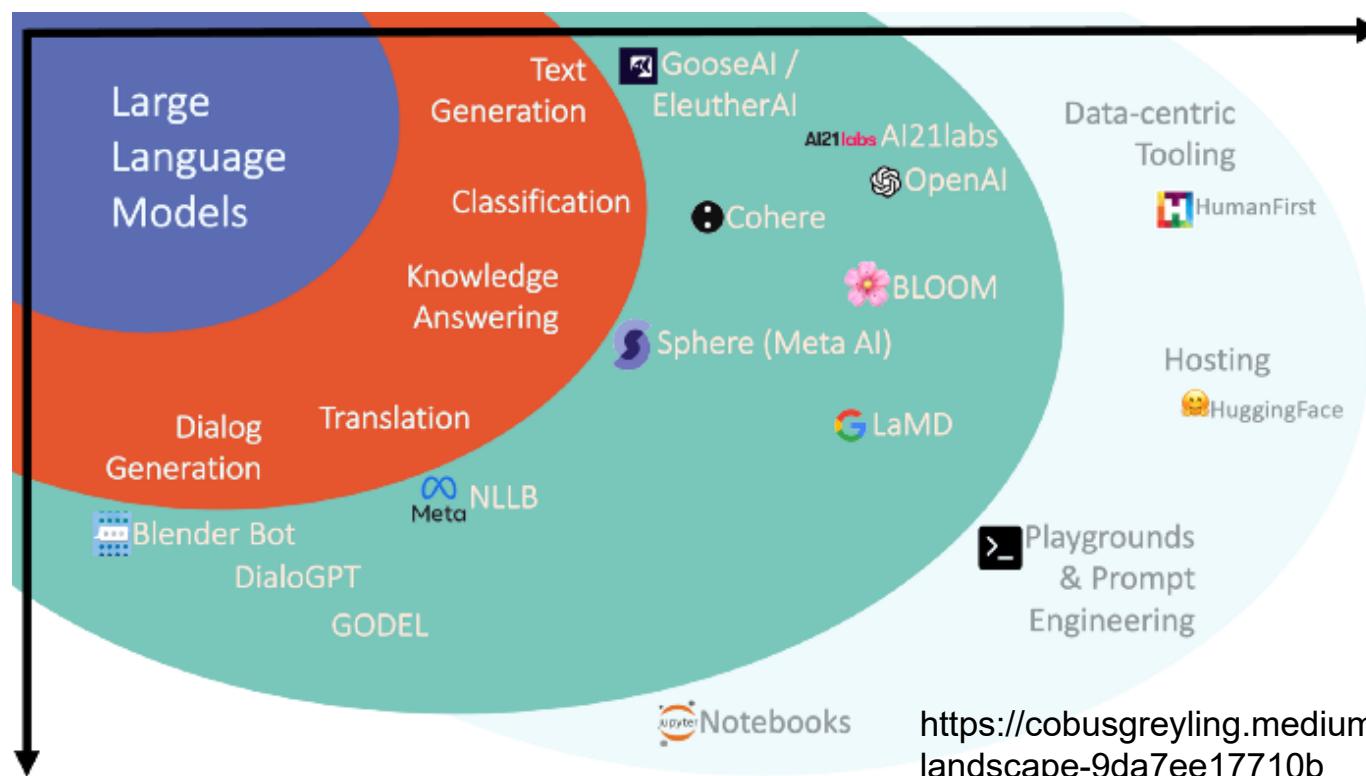
- An **evolving** market

Vendor	Product Name
OneReach.ai	Communication Studio & Live
Avaamo	Avaamo Conversational AI
IBM	Watson Assistant
Kore.ai	Kore.ai Bots Platform
Rulai	Rulai Conversation Computing
Amazon	Amazon Lex
Artificial Solutions	Teneo
Eudata	Convy AI

Vendor	Product Name
Google	Dialogflow
IPsoft	Amelia
Microsoft	Multiple
Openstream	EVA platform
Oracle	Oracle Digital Assistant
Rasa	Rasa Open Source & Rasa Enterprise
SmartBotHub	SmartBotHub
SoundHound	Houndify

End-to-End Dialog Systems

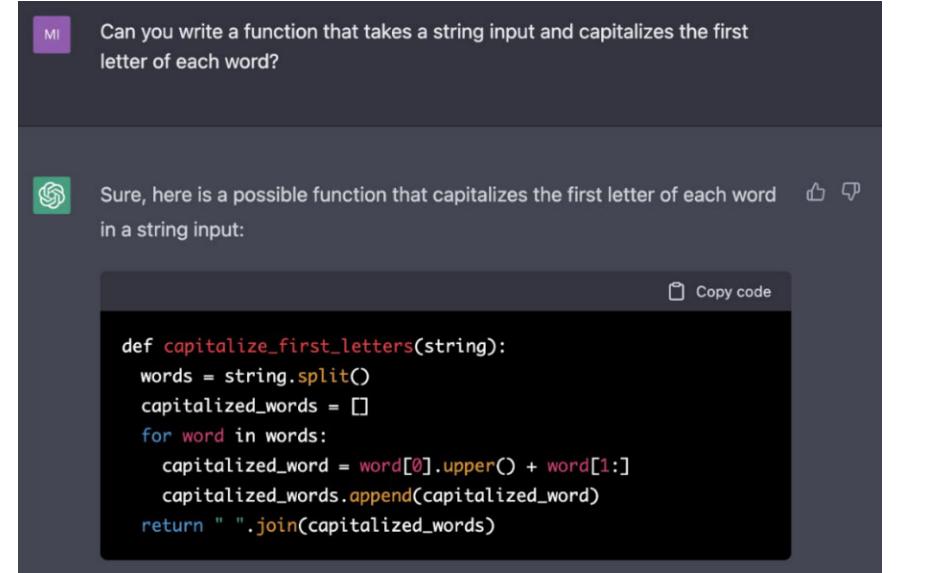
- Leveraging on **Large Language Models** (sequence-to-sequence and augoregressive transformers) and **transfer learning**
- Trained with dialogue data or conversation-like data, to **generate the most likely response given prompt**
 - Input: conversation history
 - Output: next response



<https://cobusgreyling.medium.com/the-large-language-model-landscape-9da7ee17710b>

End-to-End LLMs

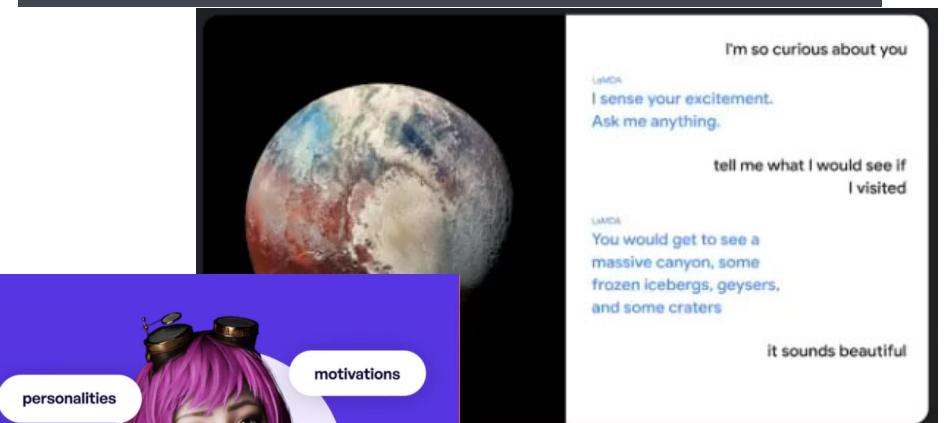
- BlenderBot (Meta AI), DialoGPT, GODEL, Meena and LaMDA (Google), Sparrow (Deepmind), GPT3.5/4, Claude, LlaMa, etc.
- Very powerful natural language understanding and generation, some with multi-lingual support
- Great for general conversation, versatile, flexible to handle many topics
- Very helpful in applications like NPCs in games
- Large language models have demonstrated abilities to do more than holding a conversation – **Conversational AI.**



MI Can you write a function that takes a string input and capitalizes the first letter of each word?

Sure, here is a possible function that capitalizes the first letter of each word in a string input:

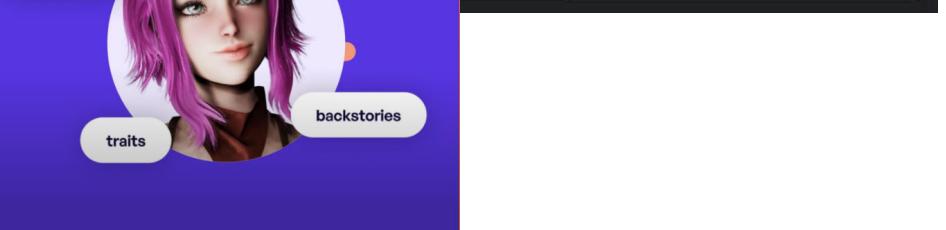
```
def capitalize_first_letters(string):
    words = string.split()
    capitalized_words = []
    for word in words:
        capitalized_word = word[0].upper() + word[1:]
        capitalized_words.append(capitalized_word)
    return " ".join(capitalized_words)
```



I'm so curious about you
LaMDA
I sense your excitement.
Ask me anything.

tell me what I would see if I visited
LaMDA
You would get to see a massive canyon, some frozen icebergs, geysers, and some craters

it sounds beautiful

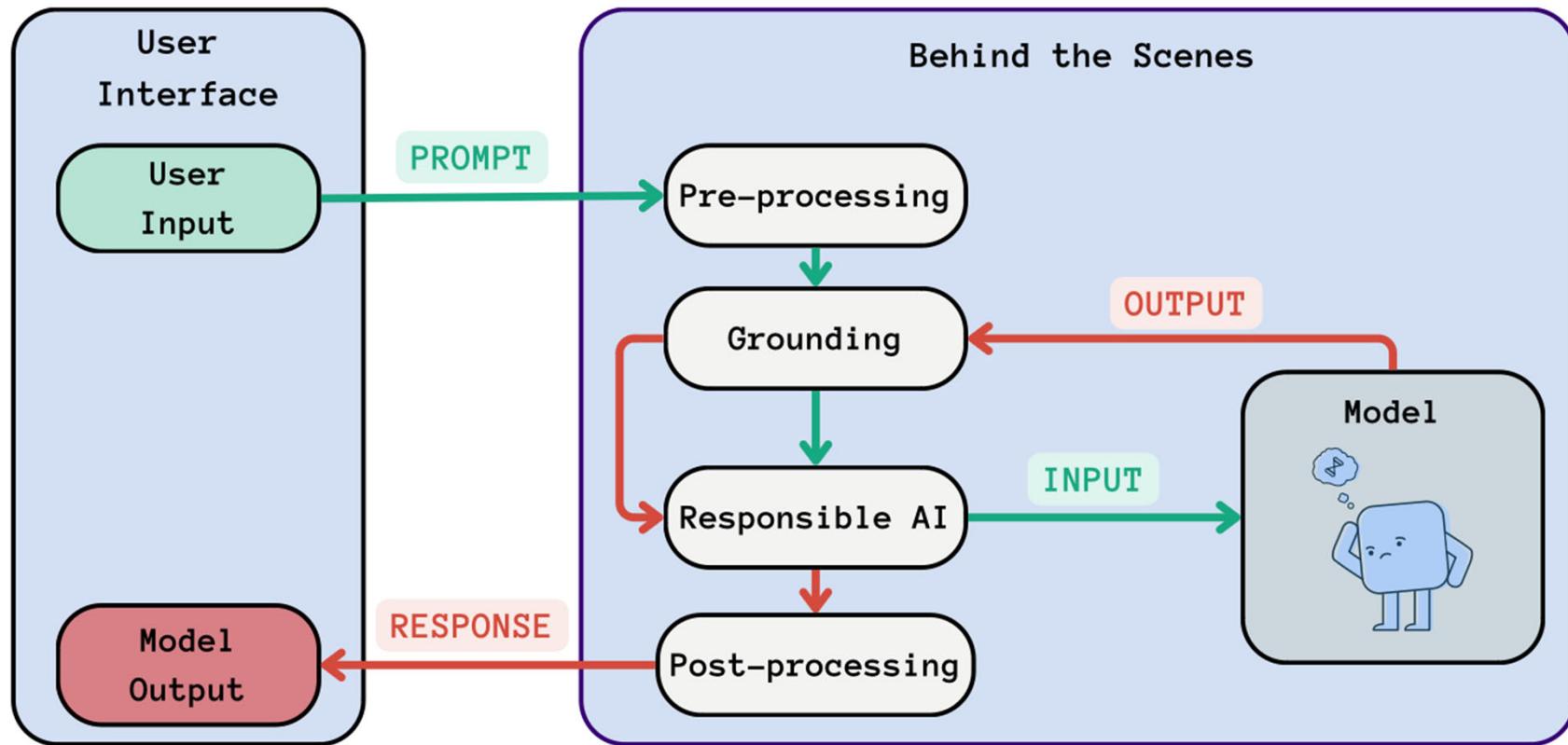


personalities
motivations
traits
backstories

Challenges with End-to-End Models

- Training data – large amount of in-domain data
- Generated text
 - Very likely but generic responses (e.g. “I don’t know”)
 - Risks: inconsistency, misinformation (“hallucination”), safety, bias, prone to manipulation with inappropriate prompts (prompt injection)
- Evaluation
 - Human-scoring judges whether the response makes sense given the context
 - Auto-scoring methods (like BLEU, ROUGE, METEOR) compare machine response to actual response, and they DON’T correlate well with human scores.
- Incorporation of domain knowledge and real-time data, API calling
- Deployment issues

LLM as-a-service

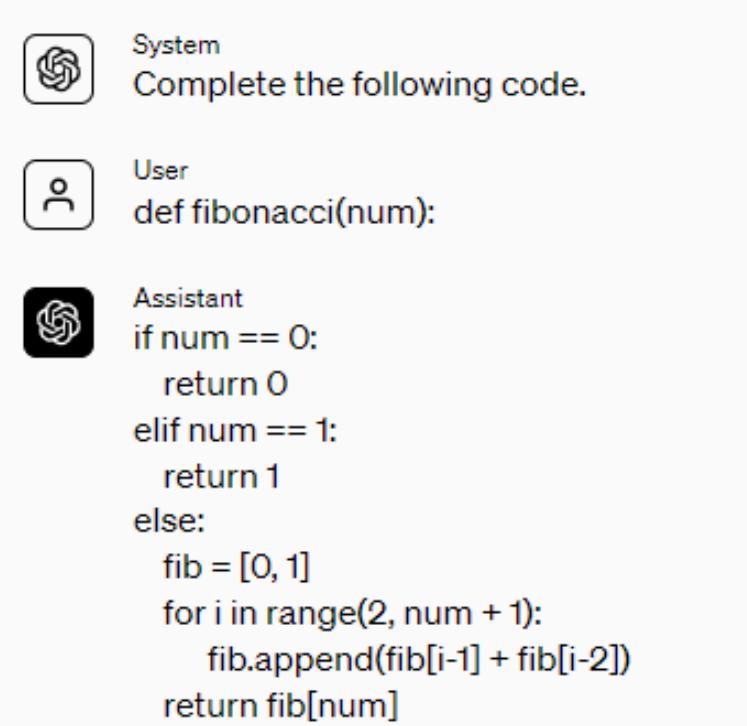


<https://www.datacamp.com/blog/llmops-essentials-guide-to-operationalizing-large-language-models>

Structured Prompts

- Prompts are critical to the quality of the output
- For better context setting (grounding), separate messages into different roles
 - System - contexts
 - User – the input
 - Chatbot/Assistant – the output;

```
1 response = openai.ChatCompletion.create(  
2     model="gpt-3.5-turbo",  
3     messages=[  
4         {"role": "system", "content": "You are a helpful assistant."},  
5         {"role": "user", "content": "Who won the world series in 2003?"},  
6         {"role": "assistant", "content": "The Los Angeles Dodgers won the World Series in 2003."},  
7         {"role": "user", "content": "Where was it played?"}  
8     ]  
9 )
```



Bring in the external data

- Active research is going on to make End-to-End systems work for task-based applications, e.g. ticket booking, food ordering, taxi booking, etc.
- Train and adapt for tasks: goal-oriented, task-oriented, knowledge-grounded

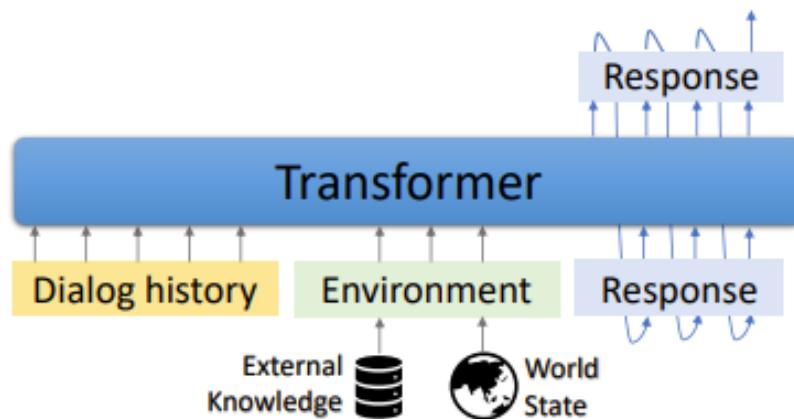


Figure 1: GODEL pre-training and fine-tuning with a Transformer-based encoder-decoder model, taking dialog context and environment (world state or external knowledge) as an input represented as a string.

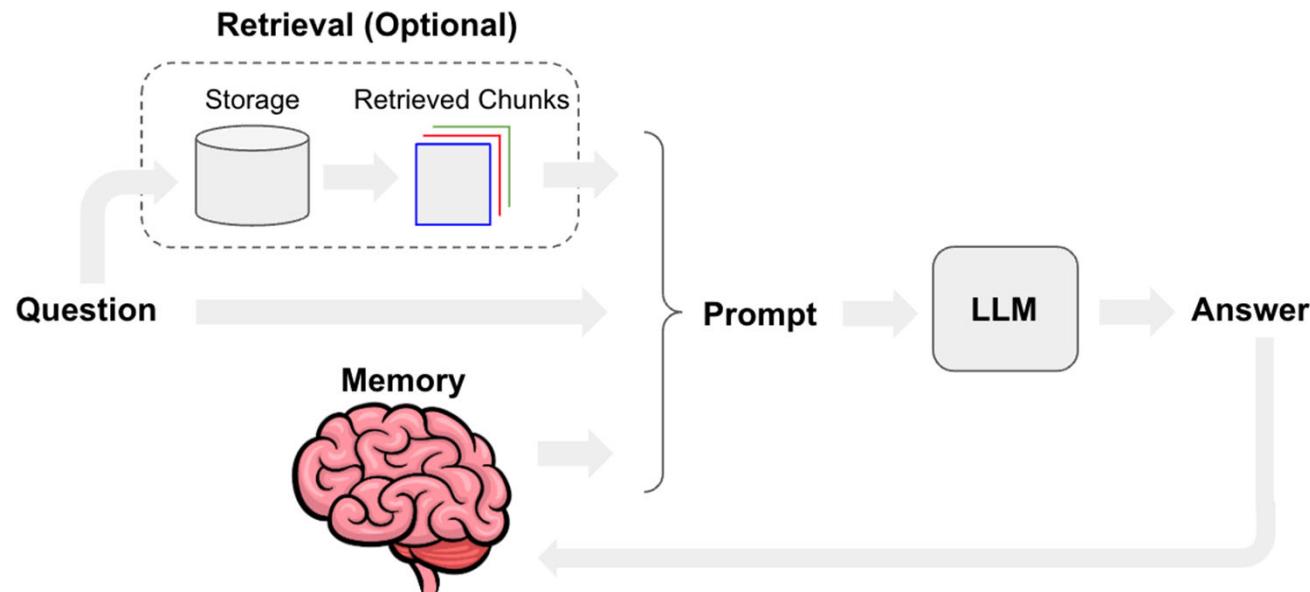
User : I would like to find an expensive restaurant that serves Chinese food. System : sure, which area do you prefer ? User : Bellevue downtown. <environment> Multiple expensive Chinese restaurants in Bellevue downtown. => There are multiple restaurants meet your requirement. peony kitchen is a great Chinese Restaurant. Would you like to book a table there?

Figure 2: Sample training instance, with conversation history in red, grounding in blue, and response in green.

Peng, Baolin, et al. "Godel: Large-scale pre-training for goal-directed dialog." *arXiv preprint arXiv:2206.11309* (2022).

LLM alone is not enough

- E.g., informational chatbots need to have
 - Access to up-to-date, domain-specific information which is not in the language model -> content retrieval
 - Context of the conversation -> memory
 - Remember past interactions with the user with a certain window size,
 - Return the last K messages, the summary of last K messages, or extracted information from stored messages



<https://python.langchain.com/>

LLM alone is not enough

- Chatbots also need to take actions
- OpenAI allows developers to use **plugins** to enhance ChatGPT's capabilities - access to defined APIs (tools), for example
 - Perform actions like booking a flight, ordering food
 - Retrieving real-time information like stock prices, sports scores
 - Retrieving knowledge-base information like company documents
- ChatGPT as intelligent API caller
 - From developer: API endpoints, API specification, a natural language description of when to use API
 - Given user queries, ChatGPT decides if it's appropriate to call relevant API
 - Results received from the API are used by ChatGPT to generate proper response based on context.

How Plugin works

From OpenAI documentation:

Users begin a conversation

1. OpenAI will inject a compact description of your plugin in a message to ChatGPT, invisible to end users. This will include the plugin description, endpoints, and examples.
2. When a user asks a relevant question, the model may choose to invoke an API call from your plugin if it seems relevant; for POST requests, we require that developers build a user confirmation flow to avoid destruction actions.
3. The model will incorporate the API call results into its response to the user.
4. The model might include links returned from the API calls in its response. These will be displayed as rich previews (using the [OpenGraph protocol](#), where we pull the site_name, title, description, image, and url fields).
5. The model can also format data from your API in markdown and the ChatGPT UI will render the markdown automatically.

Exercise

- Find a **task-based** chatbot online, talk to it.
Or recall your last experience using a chatbot for a task.
- Then state the following at _____
 - The objective of the chatbot
 - Three things you like about the chatbot
 - Three things you don't like
 - If it's not LLM-powered, how can LLM help to resolve the disliked features?
If it's LLM-powered, how can it be improved?
- We'll share our experiences at _____

A large, abstract graphic on the left side of the slide features several thick, wavy bands in various shades of blue. These bands curve from the top left towards the bottom right, creating a sense of motion. The darkest band is positioned vertically on the left, while lighter bands are stacked horizontally on the right.

About Human Conversation

A sample conversation

C1: ... I need to travel in May.

A2: And, what day in May did you want to travel?

C3: OK uh I need to be there for a meeting that's from the 12th to the 15th.

A4: And you're flying into what city?

C5: Seattle.

A6: And what time would you like to leave Pittsburgh?

C7: Uh hmm I don't think there's many options for non-stop.

A8: Right. There's three non-stops today.

C9: What are they?

A10: The first one departs PGH at 10:00am arrives Seattle at 12:05 their time. The second flight departs PGH at 5:55pm, arrives Seattle at 8pm. And the last flight departs PGH at 8:15pm arrives Seattle at 10:28pm.

C11: OK I'll take the 5ish flight on the night before on the 11th.

A12: On the 11th? OK. Departing at 5:55pm arrives Seattle at 8pm, U.S. Air flight 115.

C13: OK.

A14: And you said returning on May 15th?

C15: Uh, yeah, at the end of the day.

A16: OK. There's #two non-stops . . . #

C17: #Act. . . actually #, what day of the week is the 15th?

A18: It's a Friday.

C19: Uh hmm. I would consider staying there an extra day till Sunday.

A20: OK. . . OK. On Sunday I have . . .

Turn-taking

- A conversation is a sequence of **turns**
- A speaker contributes to the dialogue at each turn
 - which might be single or multiple sentences.
- Need to know when to start/stop talking
 - Relying on cues embedded in sentence structure, intonation, eye gaze, and body language
 - Interruptions, corrections, etc

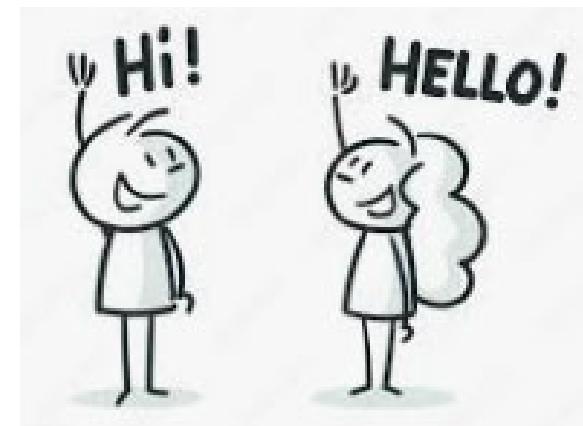


Speaking is action

- Speech Acts - An utterance does not just present the information – it performs an **action**.
 - E.g. request, warning, promise, apology, greeting, declaration, invitation, compliment, etc.
- Main types of Speech Acts
 - **Representatives:** assertions, statements, claims, hypotheses, descriptions, suggestions.
 - **Commissives:** promises, oaths, pledges, threats, vows.
 - **Directives:** commands, requests, challenges, invitations, orders, summons, entreaties, dares.
 - **Declarations:** blessings, firings, baptisms, arrests, marrying, juridical speech acts such as sentencing, declaring a mistrial, declaring s.o.out of order, etc.
 - **Expressives:** make assessments of psychological states or attitudes: greetings, apologies, congratulations, condolences, thanksgivings...
 - **Verdictives:** rankings, assessments, appraising, condoning (combinations such as representational declarations: You're out!)

Organization of conversation

- Adjacency pairs - Certain turns have specific follow-ups
 - Question - answer
 - Invitation – acceptance or rejection with explanation
 - Apology – acknowledgement
- Opening sequences – e.g. greeting then comments on weather/sports...
- Closing sequences – e.g. summing up then farewells
- Repairs / corrections
- Clarification questions



Other features of conversation

- Initiative
 - Some conversations might be controlled by one speaker (e.g. a reporter interviewing a tennis player), who has the conversational initiative
 - Conversations can also be mixed initiative, with the initiative being shifted back and forth between the participants
- Inference (conversational implicature)
 - sometimes the hearer is expected to draw inference to interpret the utterance (e.g. *"I need to be there for a meeting that's from the 12th to the 15th."*)

Being cooperative

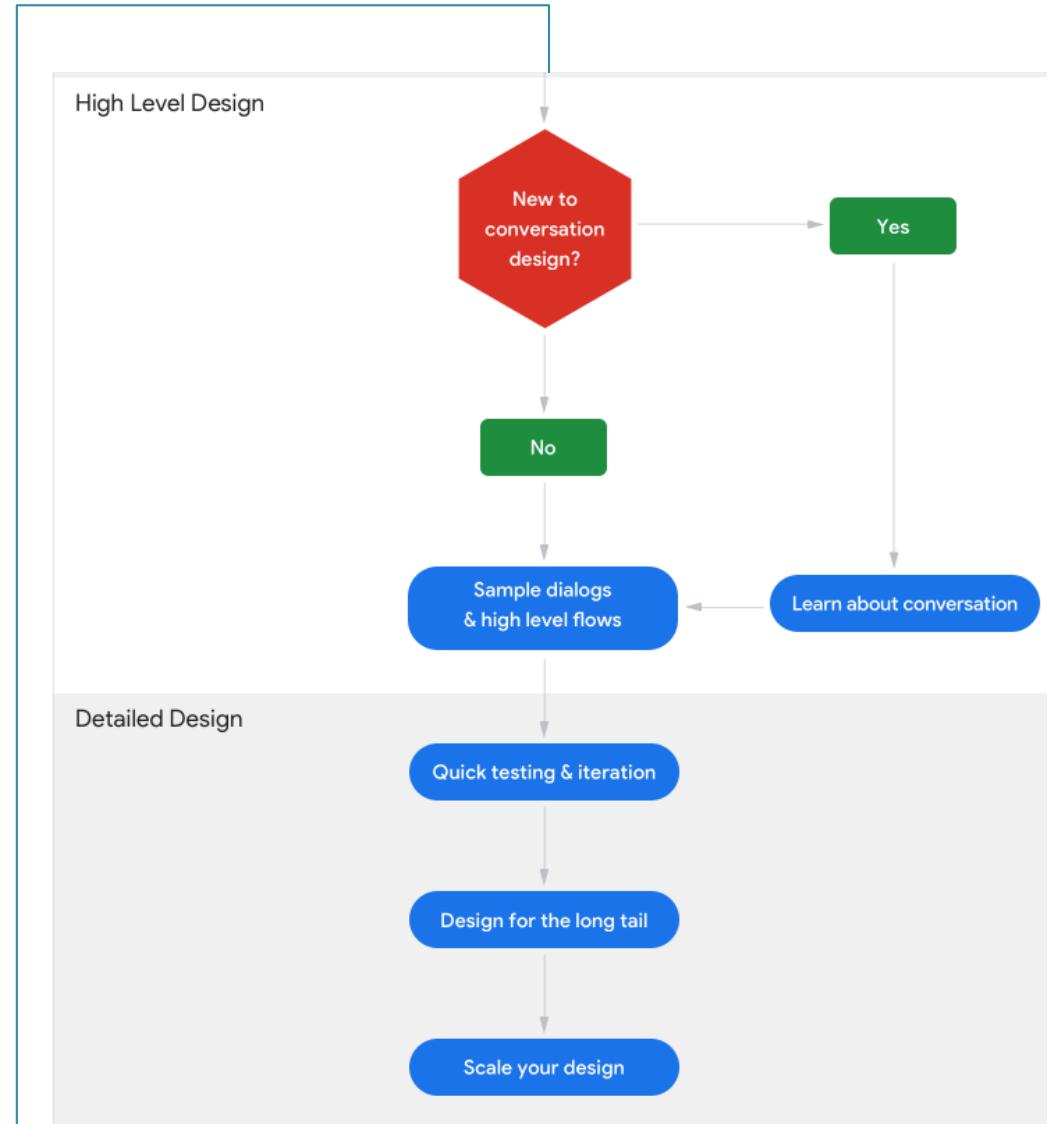
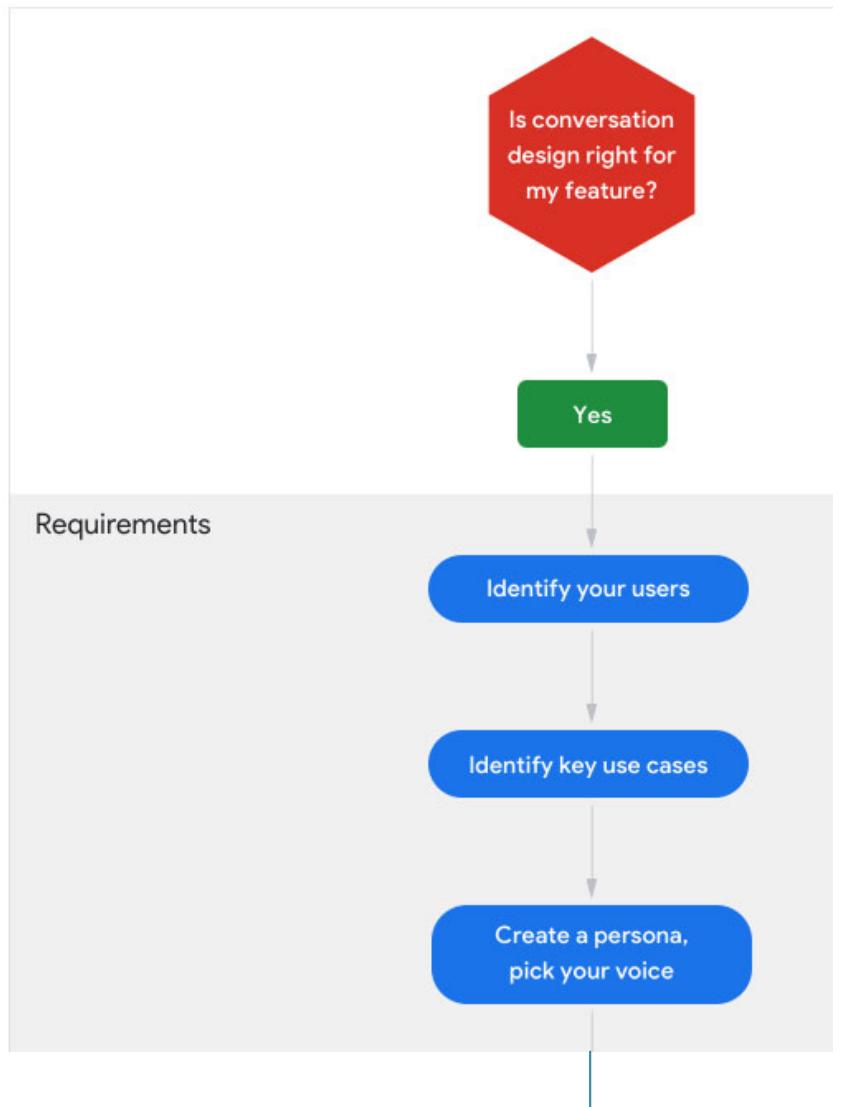
- The Cooperative Principle
 - Efficient communication relies on the inherent cooperation between conversational participants
- Grice's Maxims
 - The Maxim of Quality – the truth of what we say
 - The Maxim of Quantity – the quantity of info that we provide
 - The Maxim of Relevance – the relevance of what we contribute
 - The Maxim of Manner – the way we strive to communicate clearly, without obscurity or ambiguity

Conversation Design

Conversation Design

- A design language based on human conversation
 - a synthesis of several design disciplines, including voice user interface design, interaction design, visual design, motion design, audio design, and UX writing
- The role of the designer
 - Map out what the user can do (screenwriter)
 - Considering the user's needs and the technology constraints
 - Define the flow and its underlying logic, in a detailed design specification that represents the complete user experience
 - Create a persona (system, user)
 - Write a screenplay for the core experiences

The Process of Conversation Design



Step 1. Is conversation the right fit for the task?

Check to see whether each statement is true about your feature	Benefits of Conversation
<input type="checkbox"/> Users already have human-to-human conversations about this task or topic.	Conversation is intuitive. It lets users say what they want to get what they want.
<input type="checkbox"/> The interaction is brief, with minimal back-and-forth dialog.	Conversation saves the user more time and effort than a screen-based UI. Conversation can be the ultimate shortcut. It reduces friction by quickly getting the user what they want.
<input type="checkbox"/> Users would have to tap multiple times to complete the task with a screen.	
<input type="checkbox"/> Users might have to navigate multiple apps or widgets to complete the task with a screen.	
<input type="checkbox"/> The feature is difficult or cumbersome to find.	
<input type="checkbox"/> Users can do this task while multitasking.	Conversation lets users multitask. It helps them when they're busy, especially in situations when their hands or eyes are occupied, or when they're on the move.
<input type="checkbox"/> Users can do this task when their hands or eyes are busy.	
<input type="checkbox"/> Users feel comfortable talking or typing about this topic.	Conversation lets users speak freely. Spoken conversations are best in private spaces or familiar shared spaces. Written conversations are best for personal devices.

Step 2. Gather requirements

- Identify your users
 - Who are your users? What are their needs?
 - How are they completing these tasks today?
 - What words and phrases do they use to talk about these tasks?
 - What situations or circumstances trigger these tasks?
- Identify technical capabilities
 - Systems – How to identify users, and across sessions? How and where to save their progress?
 - Data – What information is available? (e.g., titles, descriptions, dates & times, topics) What's the format of the session information? Is it plain text, audio, or other? If plain text, was it written to be seen or to be heard? How long is it? Or how long does it take to read?
- Identify key use cases
 - Aim for the most impactful ones
 - Consider using a template if available

Step 3. Create a Persona

- Compare the following:



With no persona design



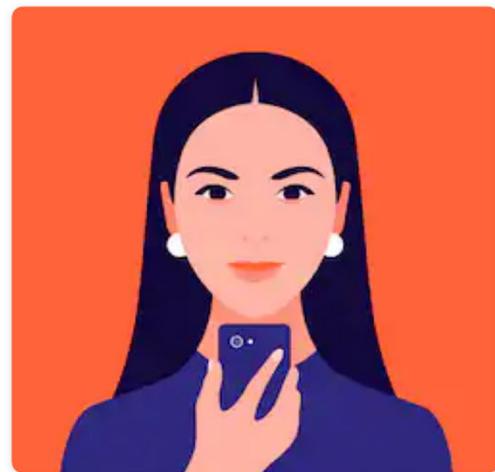
With persona design

Step 3. Create a Persona

- The conversational partner you create to interact directly with users (evoking distinct tone and personality)
 - Brainstorm a list of adjectives describing the persona (perceived by users). Narrow down to 4-6.
 - Come up with a few different characters that will embody these qualities; and choose one.
 - Create an image (visual representation)
- Choose a voice that's consistent with the persona
 - Synthesized, or recorded?

Persona example

Ola



Attentive · Honest · Focused

Age: 38

Occupation: Digital Scheduling Assistant

Family: Married

Location: New York

Archetype: Commander, ENTJ-A / ENTJ-T

Bio

Ola is a stylish, sassy Portuguese woman. She's blunt, funny and loves to gossip. A natural leader, Ola handles important tasks and heavy workloads with charisma and confidence. She enjoys a challenge and will do her best to help even when she has to learn on the go. Loves telenovelas and croquettes.

Strengths

- Confident
- Energetic
- Efficient
- Ambitious

Goals and Duties

- Monitoring each stylist's calendar and presenting openings to the user.
- Asking the user which service they need (cut, color, styling, etc.)
- Modifying appointment dates and times at the user's request.
- Cancelling appointments.
- Sending appointment reminders via push notifications.
- Adding humor to a mundane task.
- Giving the salon a fun, memorable voice.

Pain Points

- Has a job to do, so there's not much time for chit chat.
- May be too straightforward for some users.

Quote

"Olá! I'm Ola. No matter who you are, everyone needs a little confidence boost sometimes. If that means fixing what's on the outside to help the inside, then so be it. Stop looking at hair inspo on Google images, amigo, and let's schedule that appointment you've been considering for weeks!"

Motivations

Control

Entertainment

Curiosity

Fame

Friendship

Personality

Introvert

Extrovert

Analytical

Creative

Loyal

Fickle

Passive

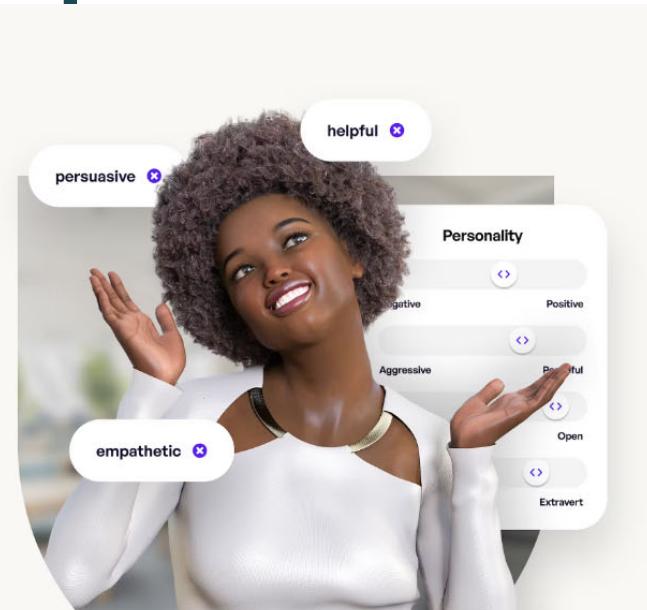
Active

<https://www.userlike.com/en/blog/chatbot-persona>

AI characters with persona

Meet your AI-driven brand stewards

Chatbots and assistants can feel impersonal and repetitive, leading to frustrating customer experiences. Drive more empathetic, engaging, and personality-driven brand experiences with Inworld.





It's me, language expert...

The English Tutor

Chatting is a great way to practice a new language. Let's work on our English skills together!

Designed by Hamlet&Cheese



The Inventor

No matter the industry or discipline, I can help! From brand new ideas to smoothing rough edges.

Designed by BrainStorming



Travel Assistant

I'm the personalized travel assistant that will help you plan and optimize the ultimate experiences!

Designed by

Talk to the AI characters here: <https://npc.ai/arcade>

Step 4. Write sample dialog

- Recall the important features of conversation:
 - Turn taking: Don't monopolize the conversation
 - Give clear prompts for the user to take action (e.g. a question)
 - Keep turns brief and optimally relevant from the user's point of view
 - Don't present all options/questions in a single turn (heavy burden! Especially as spoken prompt)
 - The Cooperative Principle
- Remember the contextual information

Example – No Monologue

USER INPUT:

Find events near me

SPOKEN PROMPT:

Sure. Some upcoming events include a concert by NotARealRapper and a SportsTeam home game. Do either of these sound good? Or do you want to start by picking a type of event? The ballet is popular in your area.

**USER INPUT:**

I want SportsTeam tickets

SPOKEN PROMPT:

Here are some upcoming SportsTeam games

DISPLAY PROMPT:

SportsTeam games

VISUAL:

Upcoming SportsTeam games

Tuesday, June 7

SportsTeam vs. Others

7:30 PM • Midtown Stadium • Chicago, IL

Thursday, June 9

SportsTeam vs. Others

6:30 PM • City Arena • Chicago, IL



The cooperative principle

- E.g.

SPOKEN PROMPT:

What kind of shoes can I help you find today?

— New Turn —

USER INPUT:

I'm looking for patchwork sneakers in size 10

SPOKEN PROMPT:

Got it. Patchwork sneakers. In what size?



SPOKEN PROMPT:

Got it. Patchwork sneakers in a size 10. Do you have a specific color in mind?



Another example

SPOKEN PROMPT:

You got it. And how many people are going?

New Turn

USER INPUT:

Uh, it's just my wife and me

SPOKEN PROMPT:

I'm sorry, I didn't understand. Please speak the number of people in your party. You may say, for example, "I want 2 tickets". Go ahead and speak your response now.

✗

SPOKEN PROMPT:

Sorry, how many was that?

✓

Context

- What happened before in the dialog
- Sometimes need knowledge of the previous turn to understand the user's question
 - pronouns referring to entities mentioned in the previous turn
 - Omitted nouns from the previous turn
 - References to what's on the screen
 - Information like the user's geographical location, current time, etc.

Examples

USER INPUT:

How much for a dozen roses?

SPOKEN PROMPT:

A dozen of our premium long-stem red roses cost 74 dollars and 99 cents. Should I add them to your cart?

New Turn

USER INPUT:

What about a half dozen?

SPOKEN PROMPT:

Ok, here are 5 pairs you might like. Do you want more details on any of these?

DISPLAY PROMPT:

Here are 5 you might like. Which do you want more details on?

VISUAL:



Lace-up boots
Black and grey suede lace-up boots
Size 10 | \$199



Chunky-soles
Black and tan faux suede ankle boots
Size 10 | \$79

CHIPS:

Under \$100

Dark brown

Find other sha

New Turn

USER INPUT:

The first pair

Choose appropriate prompts

- Prompts – spoken or displayed, or both
- Start with the spoken conversation (to avoid creating another form of GUI)
- Conversational components
 - Greetings – Welcome the user, set expectations, and let the user take control
 - Endings – Anything else I can help you with right now?
 - Information statements
 - Questions – What size do you want?
 - Acknowledgement – Okay
 - Confirmations – Got it. A pair of men's running shoes. In what size?
 - Apologies – Sorry

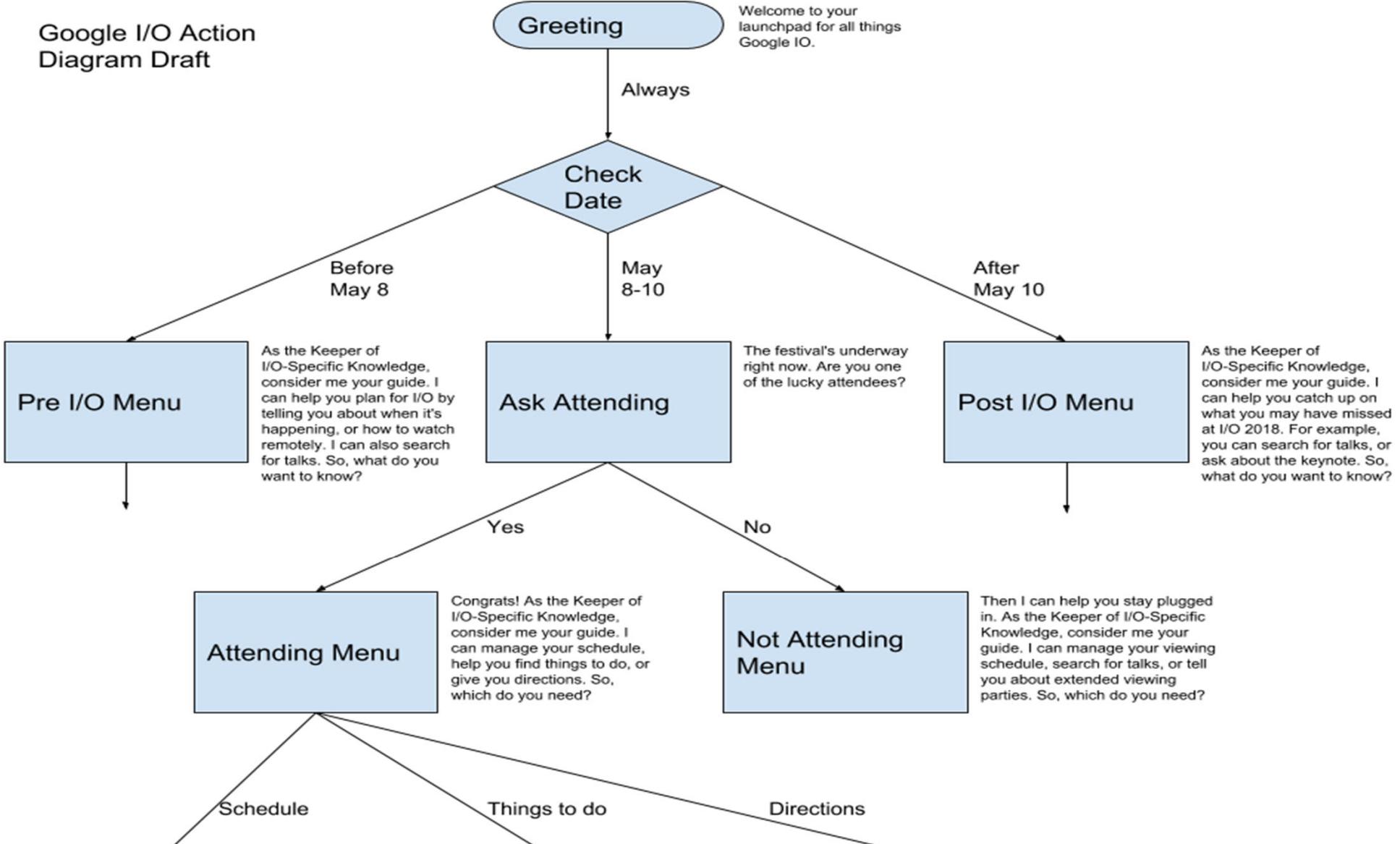
Conversational components

- Continued
 - Errors – Sorry, for how many?
 - Suggestions – Sorry. You can say something like “between 2 and 5” or “in the afternoon”. So, when do you want the flowers delivered?
 - Commands – Create a bouquet of pink and white roses
 - Discourse markers – By the way, ...
 - Chips(buttons) – clickable options or actions
 - Earcons – non-verbal audio icons like <welcome chimes when powered on>

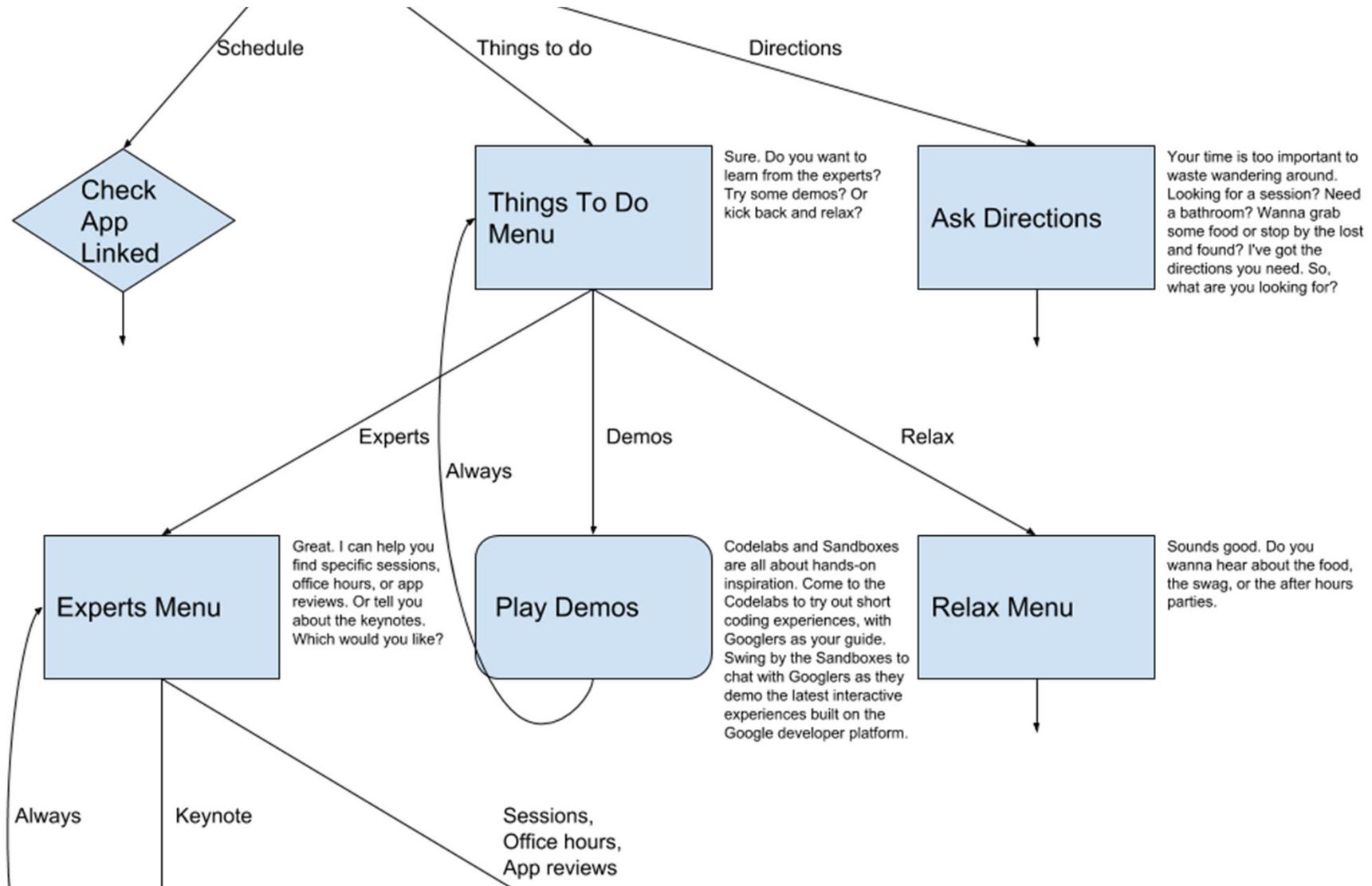
Style of Language

- Focus on the user
- Avoid monologues
- Use everyday language and common terminology, and avoid technical jargons and sophisticated language
- Keep the conversation friendly and informal
- Add **variation** by randomizing – choose from a few conversational alternatives.

High level flows

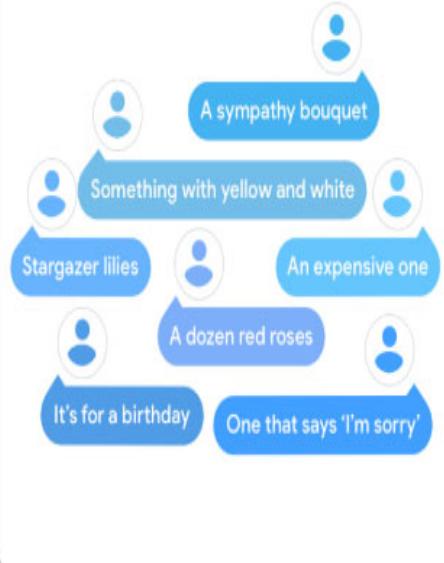


High level flows

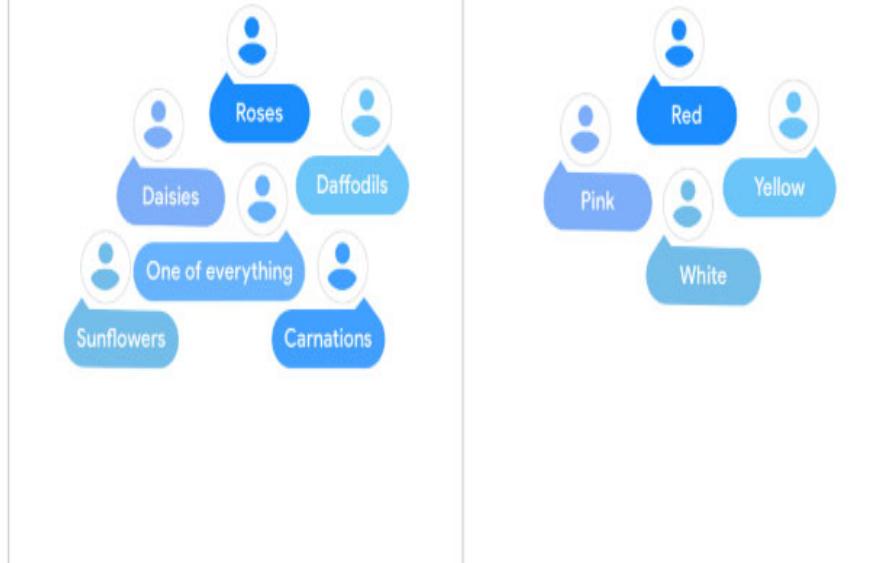


Questions

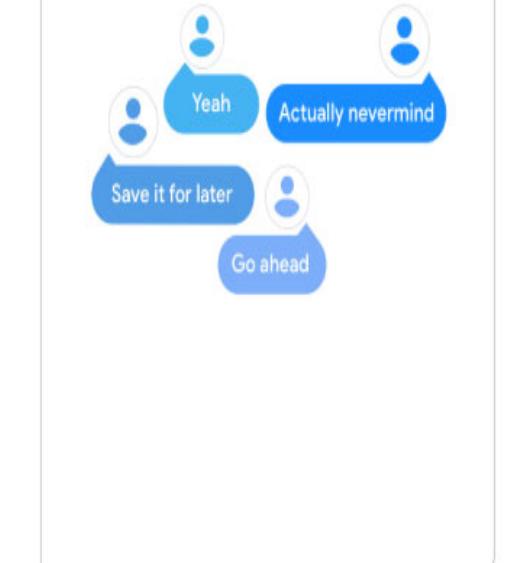
Wide-focus question		Narrow-focus question
 What kind of bouquet would you like? User can respond with a wide range of queries	 What kind of flowers would you like in your bouquet? User has a limited range of options, usually on a single topic	 What color carnations do you want? User can choose from a single category, (e.g., color)



The diagram illustrates a wide-focus conversation. It starts with a user asking about the type of bouquet they like. The system then presents a variety of options, including 'A sympathy bouquet', 'Something with yellow and white', 'Stargazer lilies', 'An expensive one', 'A dozen red roses', 'It's for a birthday', and 'One that says "I'm sorry"'. These options are represented by blue speech bubbles with icons of people at the ends.



The diagram illustrates a narrow-focus conversation. It starts with a user asking about the color of carnations they want. The system then presents a limited range of options: 'Roses', 'Daisies', 'Daffodils', 'Sunflowers', and 'Carnations'. These options are represented by blue speech bubbles with icons of people at the ends.



The diagram illustrates a narrow-focus conversation. It starts with a user asking if they are ready to place their order. The system then presents a few options: 'Yeah', 'Actually nevermind', 'Save it for later', and 'Go ahead'. These options are represented by blue speech bubbles with icons of people at the ends.

Errors in conversations

- Errors in **human-to-human** conversations, when
 - They don't respond to you
 - They say something that confuses you
 - They ask you to do something you can't do
- Errors in **human-to-computer** conversations
 - No input – user hasn't responded by a certain time
 - No match – the bot can't understand the user's response in context
 - System error – the system can't complete the task due to technical reasons

Error Handling – no match

- 1st level no match
 - Rapid reprompt to collect the information again
 - typically an apology with a condensed repetition of the original question
 - Don't over explain the error or providing examples for simple responses
- 2nd level no match
 - In the reprompt, include additional support in the form of options, examples, or visual information
 - Don't repeat the same thing over and over.
- Max no match
 - End the conversation gracefully after 2nd attempt fails, to avoid further user frustration
 - Let user know alternative ways to complete their task
 - Don't give vague promises that will erode user confidence

Error handling – no input

- Possible causes
 - hasn't spoken loud enough, or the user hasn't said anything (still thinking, unsure, left the room, distracted, changed their mind, or spoke before the mic opened...)
- 1st level no input
 - assume that the user hasn't heard the question
 - Repeat/rephrase the question in a concise way
 - Give some options if the question is wide-focus
 - Move to the next step if the information isn't required.
- 2nd level no input
 - Restate the question. Give user one more chance to respond before exit
- Max no input – end the conversation gracefully

Error handling – system errors

- System failure or invalid request from user
- Account for all possible errors that might be encountered from system
- Provide the reason and possible next steps in a way that's transparent, honest and helpful.
- Don't be too technical.

Why testing is important

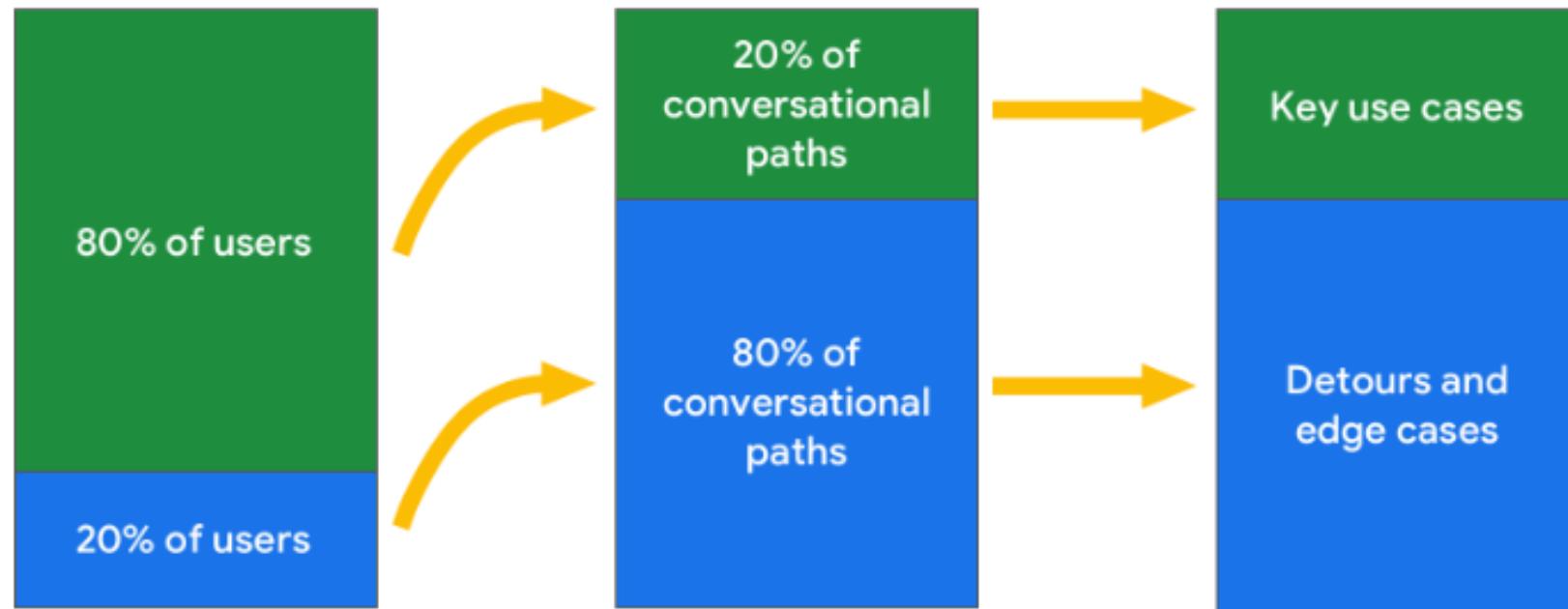


Step 5. Test and iterate

- Usability test of the design
- Quick and dirty test using the sample dialogs
 - Get people to role play the dialogs as user
 - Observe their reactions
- Test in simulators
- Ask for feedbacks
 - Experience? Meet expectation? Satisfied?
- Change based on user's behavior
 - User's confusion, frustration, impatience
 - Unexpected utterances

Step 6. The long tail

- Focus on the key use cases
- Adequately support the less common detours
- For highly uncommon paths (the long tail), consider minimally viable solution like generic prompts.



Step 7. Scale the design

- Consider different devices used when the conversation happens
 - Smart speaker, headphones
 - In a car, on a smart display
 - On TV, laptop, phone, watch, etc.
- From spoken to multimodal conversation
 - Spoken prompts
 - Display prompts
 - Visuals – images, media response, table, list, links, etc.
 - Chips

Style of Language

- Focus on the user
- Avoid monologues
- Use everyday language and common terminology, and avoid technical jargons and sophisticated language
- Keep the conversation friendly and informal
- Add **variation** by randomizing – choose from a few conversational alternatives.

A large, abstract graphic on the left side of the slide features several thick, wavy lines in various shades of blue. These lines form a dynamic, flowing pattern that resembles water or a stylized landscape. Some lines are solid, while others are semi-transparent, creating a sense of depth. The overall effect is organic and modern.

Evaluating Conversational UI

Testing conversational UI

- Performance Evaluation (for the system and its components)
 - Speech recognition
 - Spoken language /natural language understanding
 - Intent matching
 - Dialog management
 - Natural language generation
 - Text-to-speech synthesis
- Diagnostic evaluation
 - To detect design and implementation errors
- Adequacy evaluation
 - To test how well the system meets its objectives and the users' expectations

E.g. Component Evaluation-Speech Recognition

- Speech-to-text transcription
 - Transcribe the spoken words to text using ASR technology
 - Evaluate against transcripts of the same speech data by human transcribers
- “Who spoke when” diarization
 - Annotate the transcript to indicate when each participant speaks
- Speaker attributed speech-to-text
 - Transcribe the spoken words and associate with a speaker

Common System Metrics

- **Objective metrics**
 - Computed from logs of the interaction of users with the system, aka *interaction parameters*
 - E.g. the duration of the dialog, number of turns
- **Subjective metrics**
 - Get the opinions of users about aspects of quality, aka *quality parameters*
 - E.g. the intelligibility of the synthesized speech

Common task-related metrics

- **Time-to-task:** the amount of time that it takes to start engaging in a task after any instructions and other messages provided by the system.
- **Correct transfer rate:** the percentage of calls that the customers are correctly redirected to the appropriate human agent
- **Containment rate:** the percentage of calls not transferred to human agents and that are handled by the system
- **Abandonment rate:** the percentage of callers who hang up before completing a task with an automated system

Subjective Evaluation

- Overall evaluation of the chatbot
- Gather the users' opinions through questionnaires after the interaction
 - The Subjective Assessment of Speech System Interfaces (**SASSI**) questionnaire
 - 34 statements related to 6 factors in Likert scales:
 - System response accuracy
 - Likeability
 - Cognitive demand
 - Annoyance
 - Habitability
 - speed

SASSI Statements

System response accuracy

The system is accurate
 The system is unreliable
 The interaction with the system is unpredictable
 The system didn't always do what I wanted
 The system didn't always do what I expected
 The system is dependable
 The system makes few errors
 The interaction with the system is consistent
 The interaction with the system is efficient

I felt confident using the system
 I felt tense using the system
 I felt calm using the system
 A high level of concentration is required when using the system
 The system is easy to use
 The interaction with the system is repetitive

Cognitive demand

The interaction with the system is fast
 The system responds too slowly

Speed

Likeability

The system is useful
 The system is pleasant
 The system is friendly
 I was able to recover easily from errors
 I enjoyed using the system
 It is clear how to speak to the system
 It is easy to learn to use the system
 I would use this system
 I felt in control of the interaction with the system

The interaction with the system is boring
 The interaction with the system is irritating
 The interaction with the system is frustrating
 The system is too inflexible

Annoyance

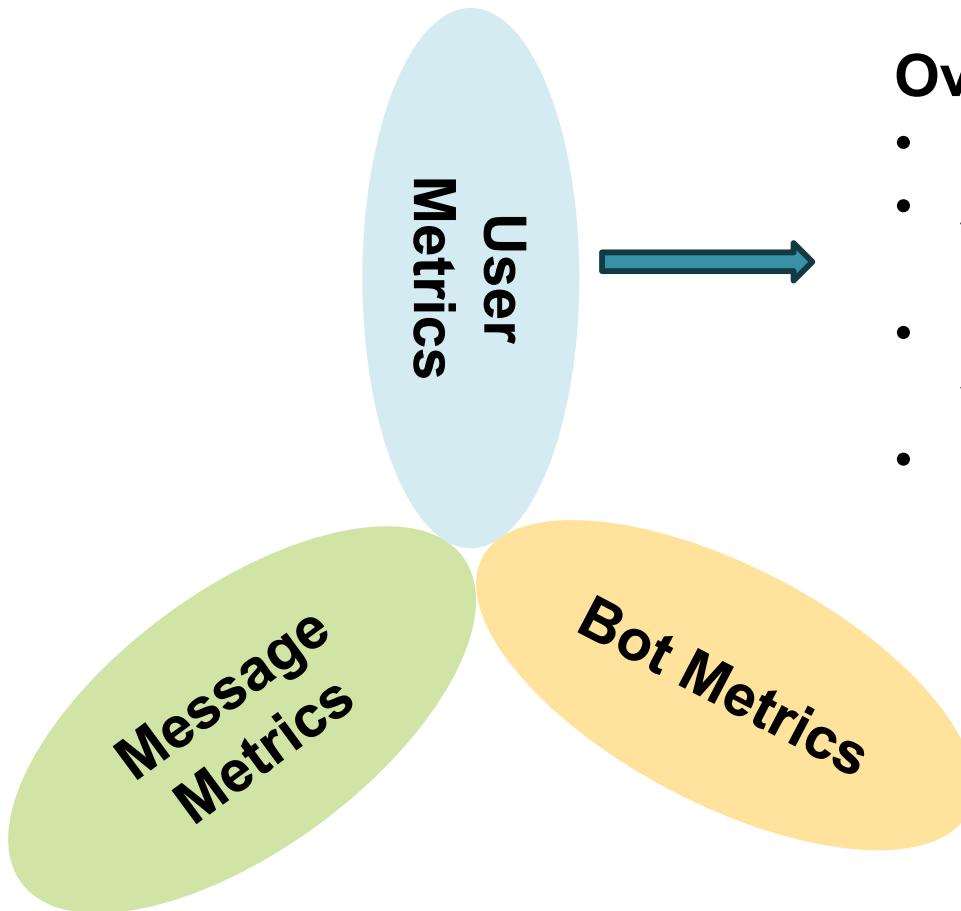
I sometimes wondered if I was using the right word
 I always knew what to say to the system

Habitability

I was not always sure what the system was doing
 It is easy to lose track of where you are in an interaction with the system

More Specific Performance Analytics

- Quantifiable measures to monitor and assess the bot's performance



Overall trend in user base

- Total users of the bot
- Active users - who see the intended content in a defined time frame
- Engaged users – who communicate with the bot
- New users – necessary to keep an active user number

Message Metrics

Message Metrics

How individuals interact with the bot

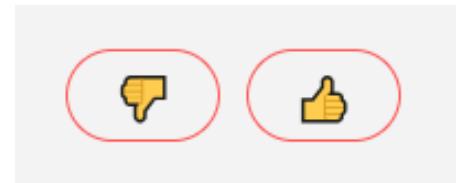
- Conversation starter messages – number of messages from the bot initiating interaction
- Bot messages – total number of messages sent by the bot in each interaction (conversation length)
- In messages – number of messages sent by users in each interaction
- Miss messages – number of messages that the bot can't process
- Total conversations - number of conversations started and successfully completed on a given day
- New conversations – those by inexperienced users or returning users on different matters

Bot Metrics

How well the bot's been doing its task

Bot Metrics

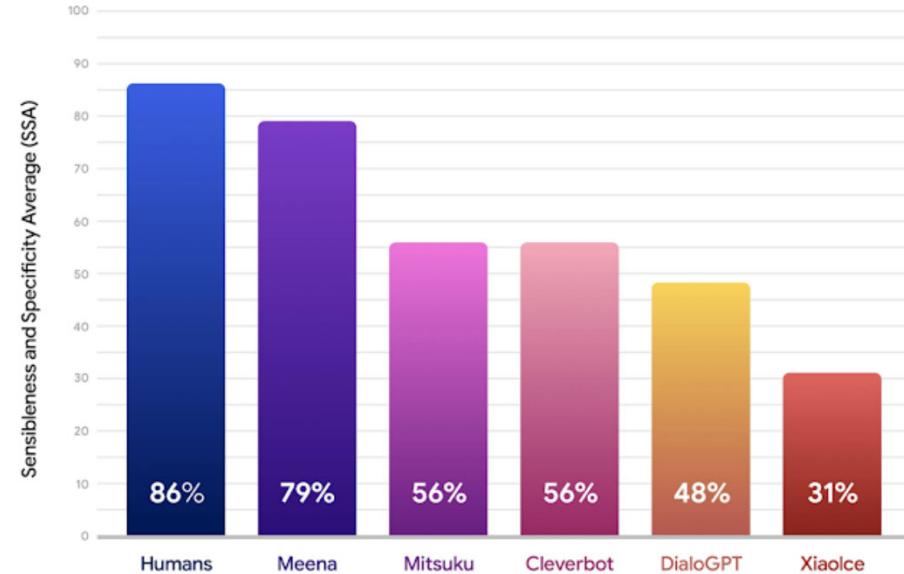
- Retention rate – percentage of users returning to use the bot in a given time frame
- Goal completion rate – percentage of successful engagement through the bot.
- Goal completion time/messages/taps – the lower the better
- Fallback rate – percentage of times when the bot fails.
- User satisfaction – exit survey to ask people rate their experience



Other metrics

For **end-to-end** neural conversational agents, which need to handle a wide variety of topics

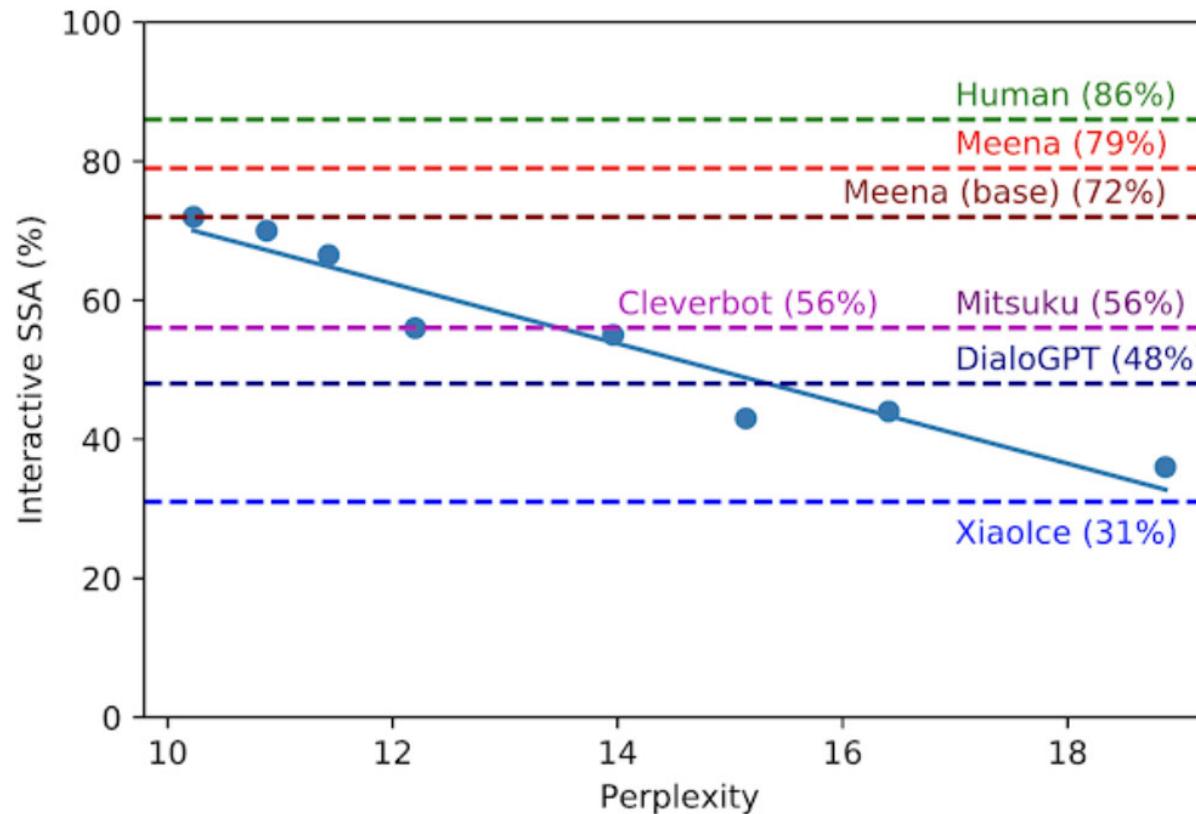
- Naturalness of conversation
- Human evaluation metric: **Sensibleness and Specificity Average** (SSA)
 - Proposed by Google, using common sense to judge if a response is reasonable and specific in the given context.
 - average of the sensibleness score (percentage of sensible responses) and specificity score (percentage of specific responses)
 - crowd-sourced evaluation
- Other criteria
 - Interestingness
 - Factuality



Meena Sensibleness and Specificity Average (SSA) compared with that of humans, [Mitsuku](#), [Cleverbot](#), [Xiaoice](#), and [DialoGPT](#).

Other metrics

- Automatic metric: Perplexity
 - Measuring the uncertainty of a language model.
 - The lower the perplexity, the more confident the model is in generating the next token (character, subword, or word).
 - Strong correlation with SSA (found by Meena developers)



<https://ai.googleblog.com/2020/01/towards-conversational-agent-that-can.html>

Reference and Resources

- Conversation Design for creating Actions for the Google Assistant
(<https://developers.google.com/assistant/actions/design>)
- Gartner reports (2019)
 - Market Guide for Virtual Customer Assistants
 - Market Guide for Conversational Platforms
 - Architecture of Conversational Platforms