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 Agenda

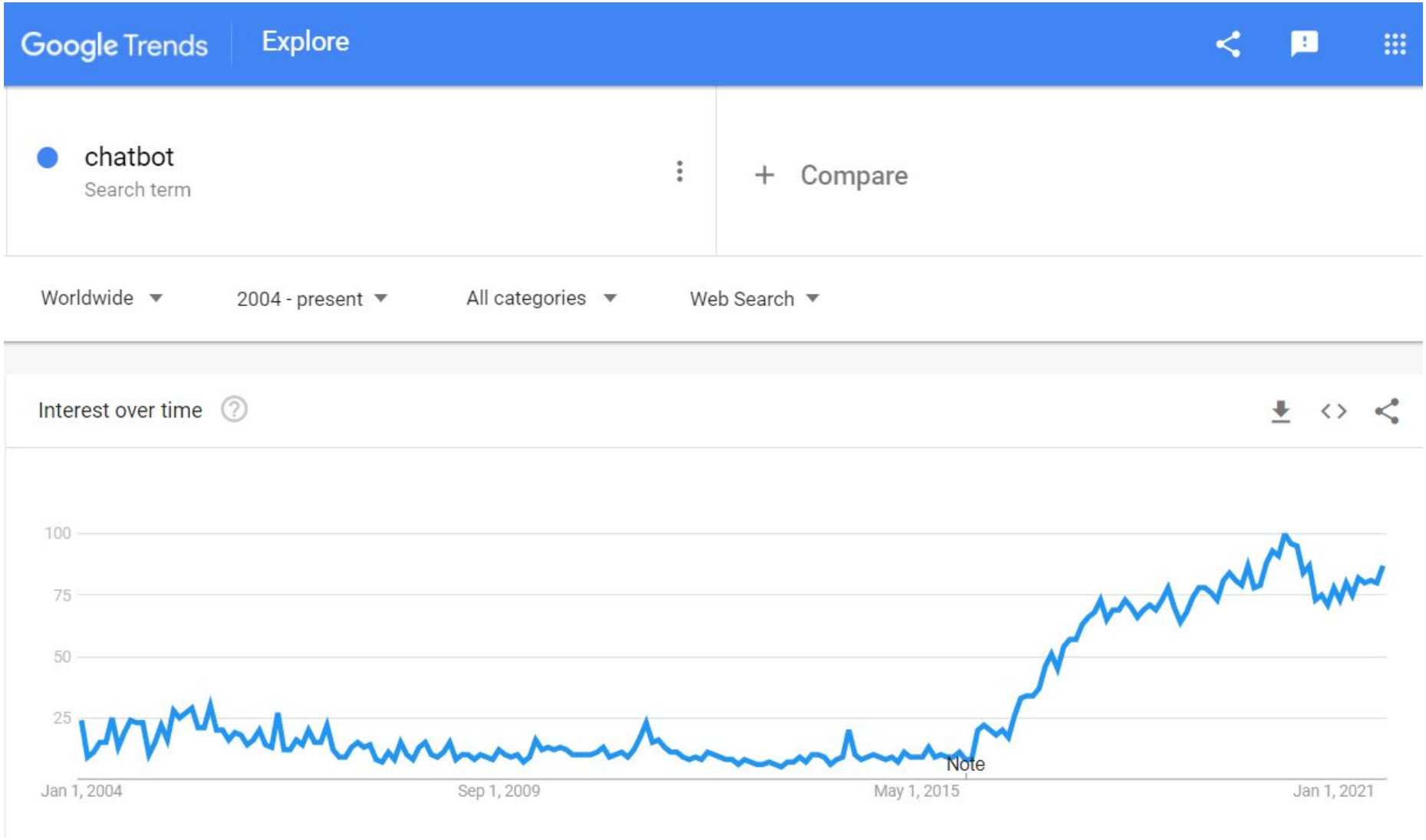
Day 1	Introduction to Conversational UIs Conversation Design Evaluating Conversational UI
Day 2	Understanding the Content of User's Utterances
Day 3	Speech Recognition Text-to-speech Synthesis Spoken language understanding
Day 4	Response Generation Review of speech modules

INTRODUCTION TO CONVERSATIONAL UI

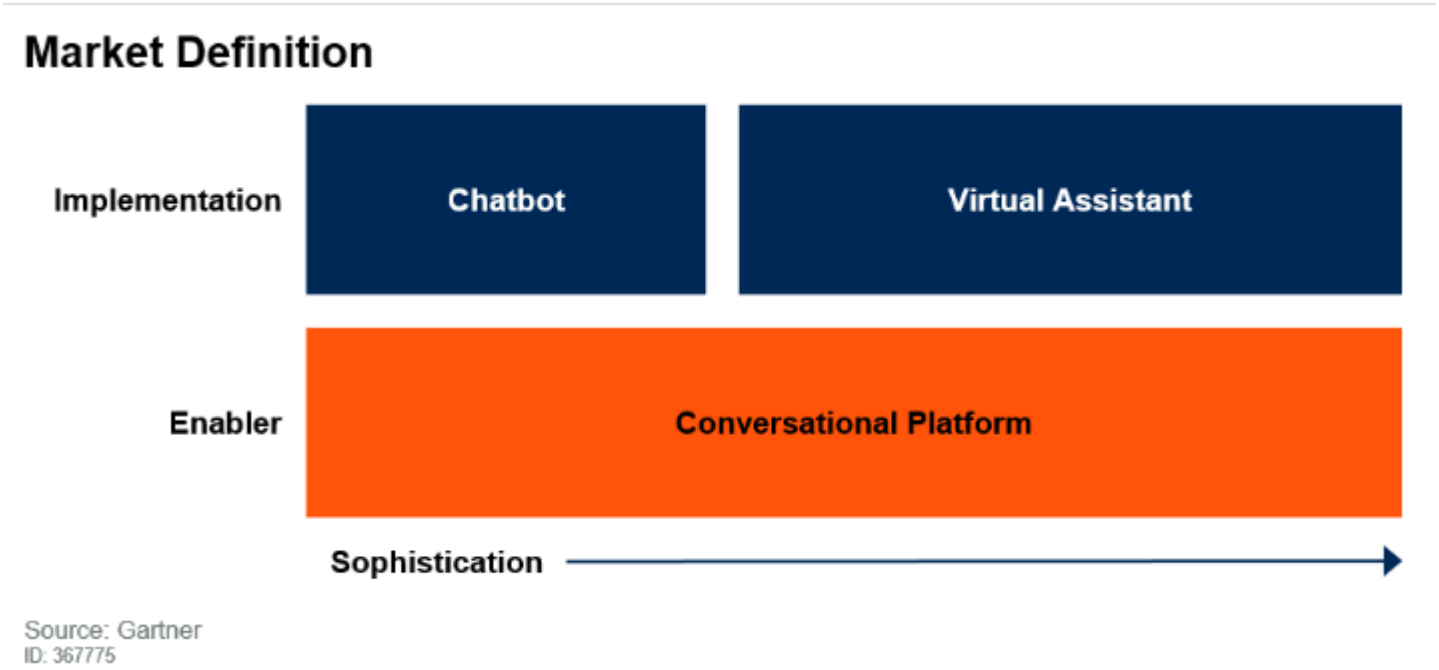
Conversational UI - a new paradigm of user interface

What can a conversational UI do?

Interest in Chatbots

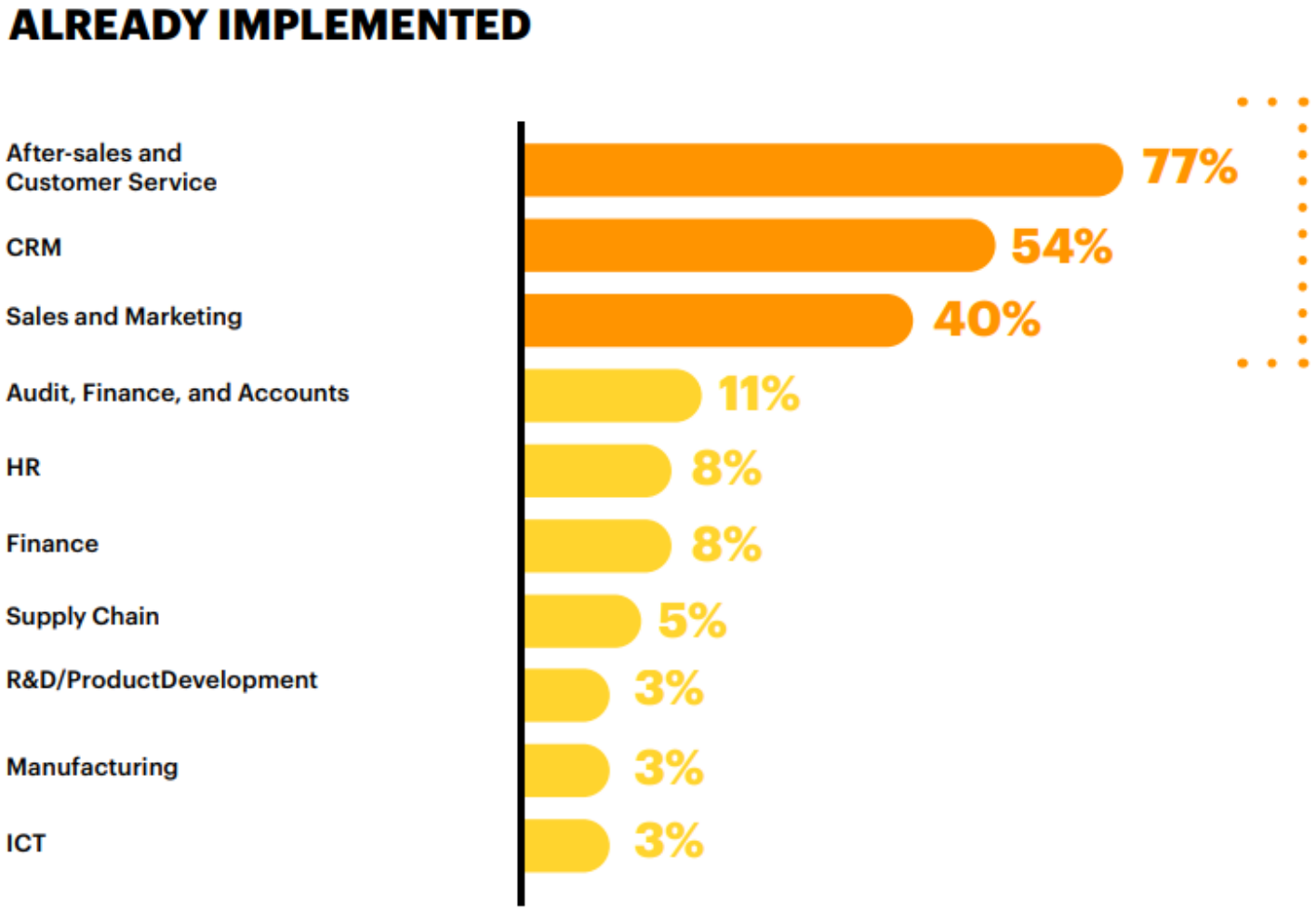


CUIs? Chatbots? Virtual Assistants?



Source: Gartner
ID: 367775

Where are they used?



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

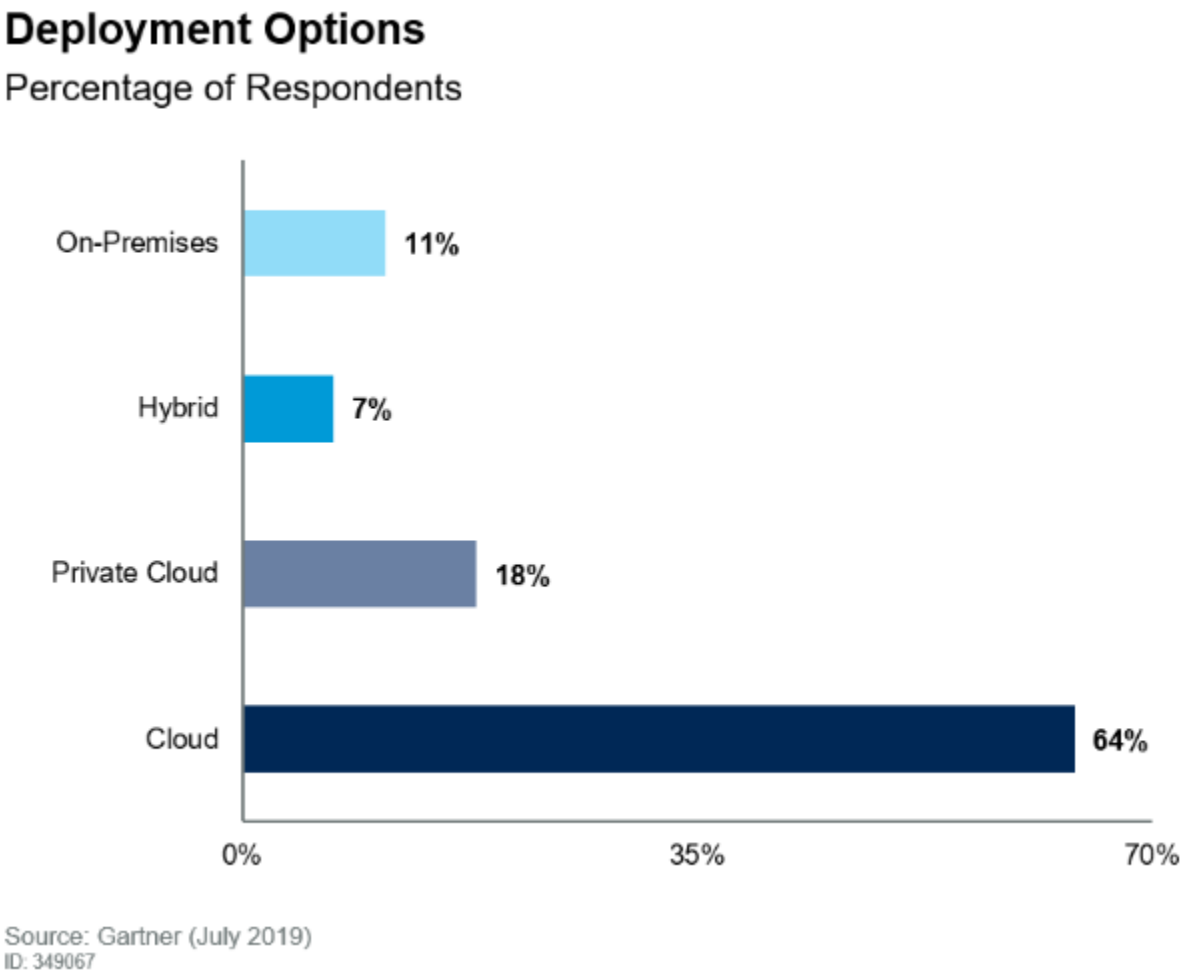
Popular Use-cases

- **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

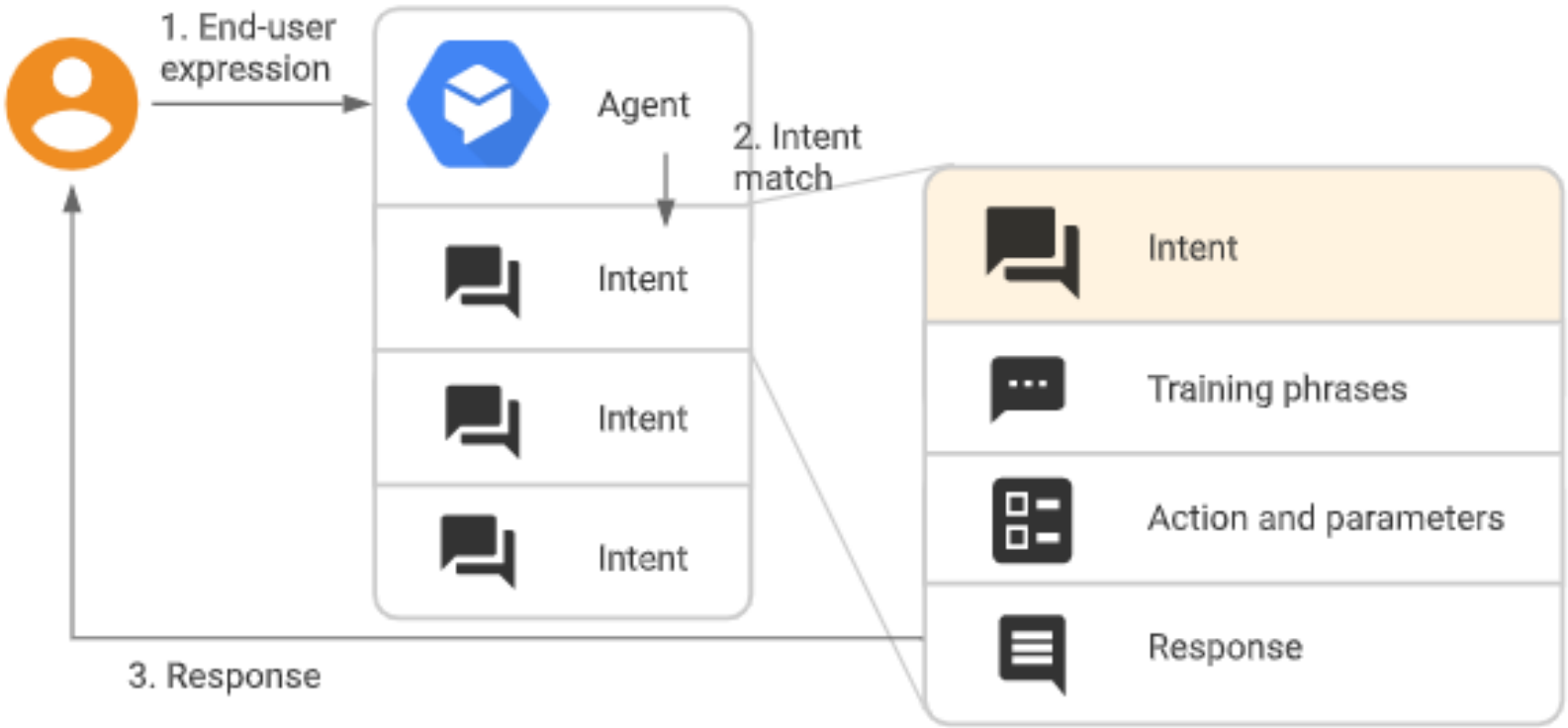
Choices of Deployment



Main concepts of Conversational UI

- **Intents -** an end-user's intention for one conversation turn
- **Parameters –** 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



Architectures of Dialog Systems

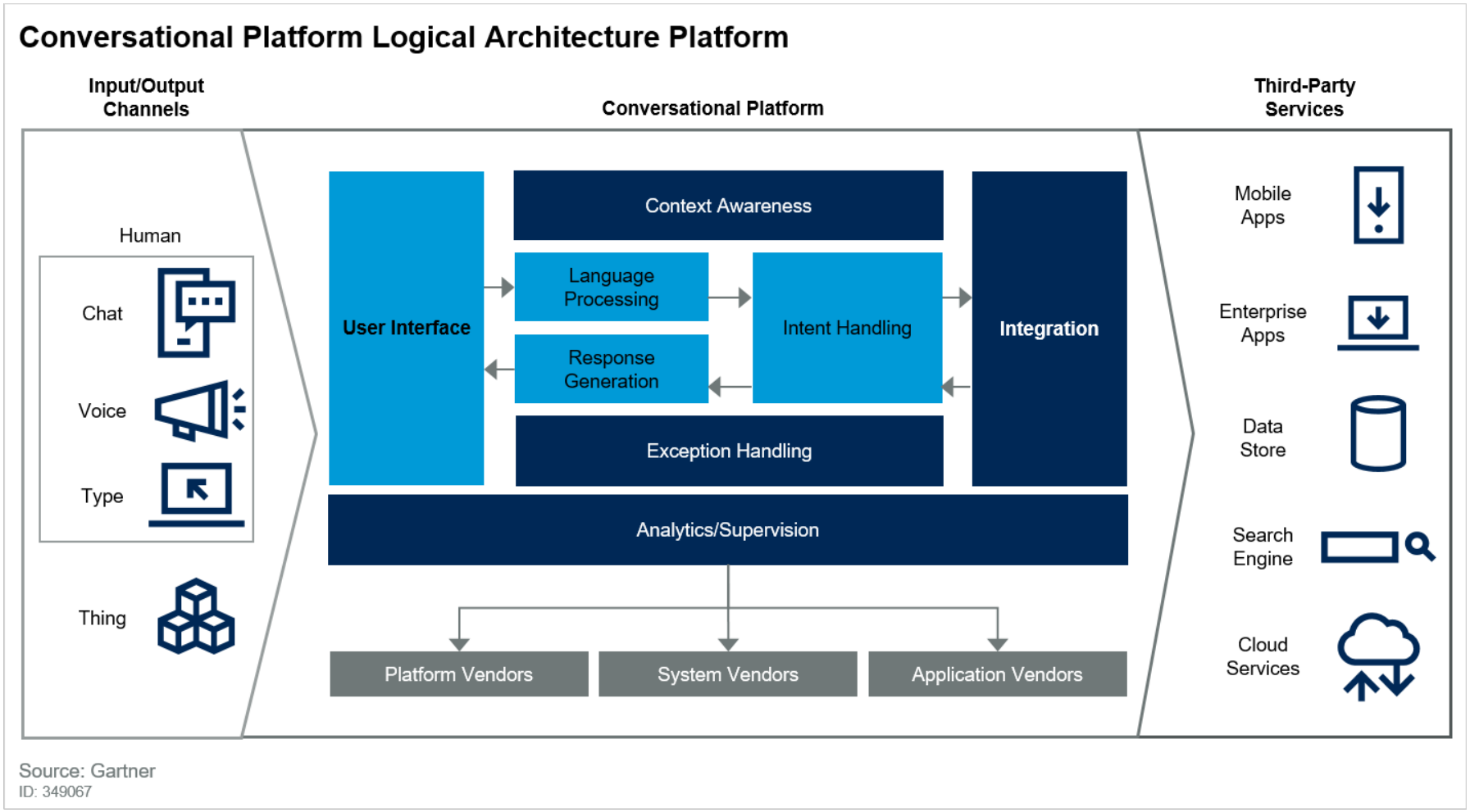
- **End-to-end systems -** one single model trained from a dialog dataset

- **Modular systems –** different components taking care of separate functions, e.g. status tracking, response generation, etc

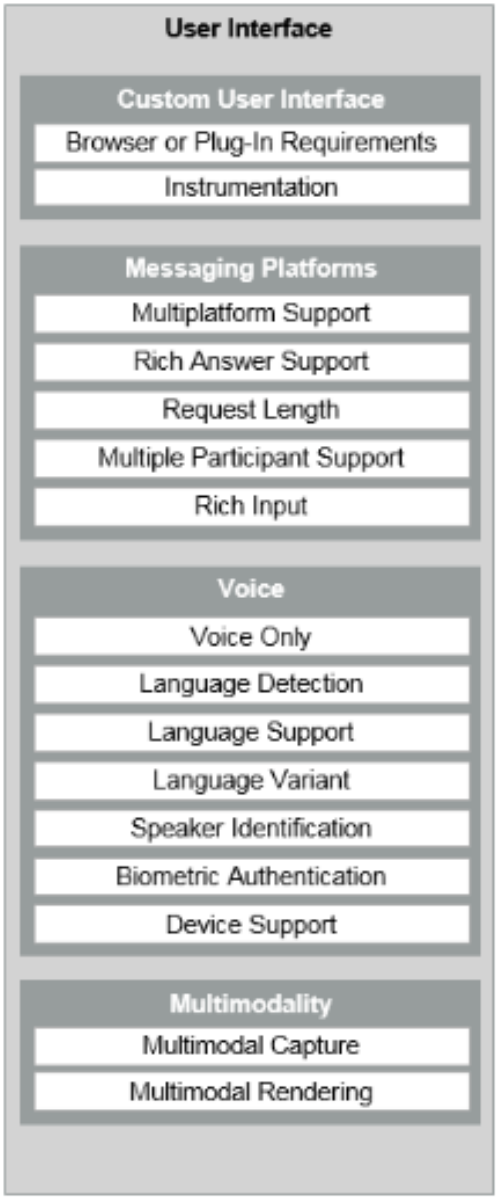
End-to-End Dialog Systems

- Leveraging on sequence-to-sequence learning (transformers), transfer learning
  - Input: conversation history
  - Output: next response
- Works going on for task-based applications, e.g. ticket booking, food ordering, taxi booking, etc.
- Challenges:
  - Training data – large amount of in-domain data
  - Generic responses are also very likely responses (e.g. “I don’t know”)
  - Evaluation
    - Human-scoring judges whether the response makes sense given 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, API calling

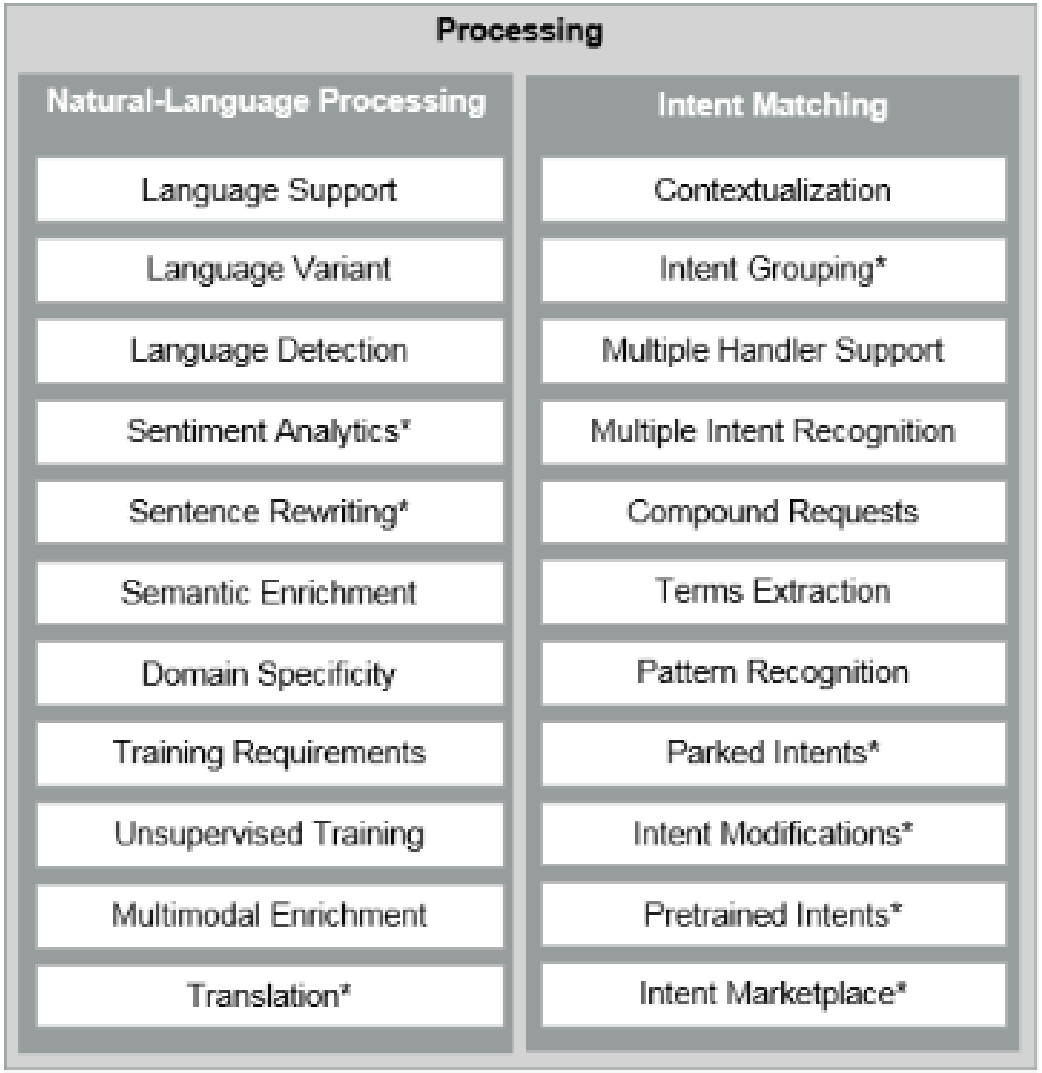
Architectures and components of Conversational Platforms



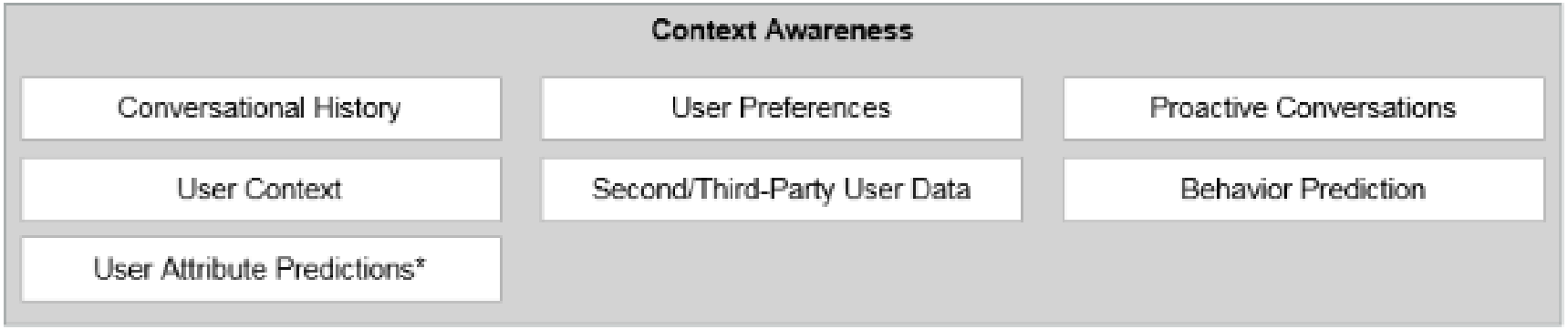
Detailed Capabilities for User Interface



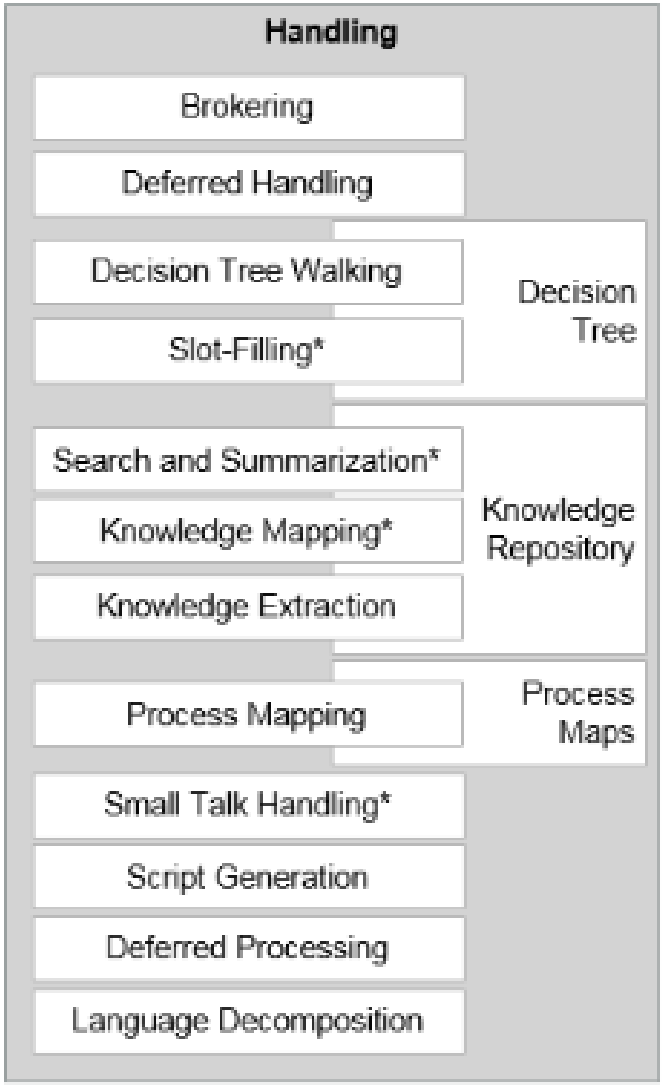
Language Processing



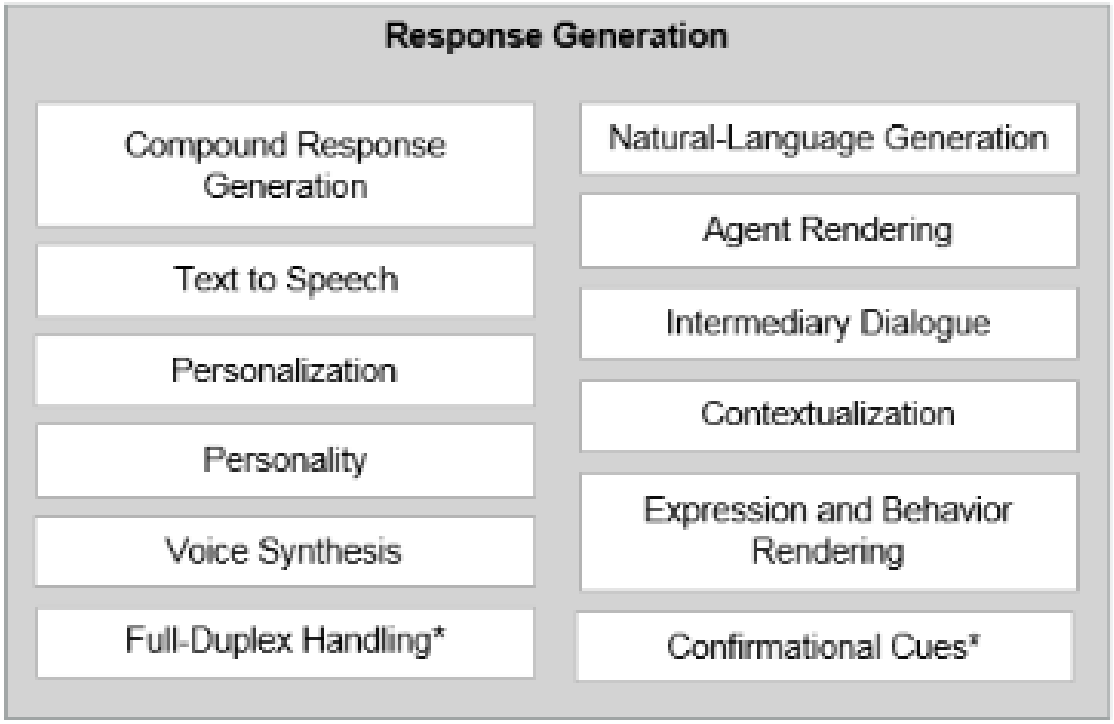
Contextual Awareness Capabilities



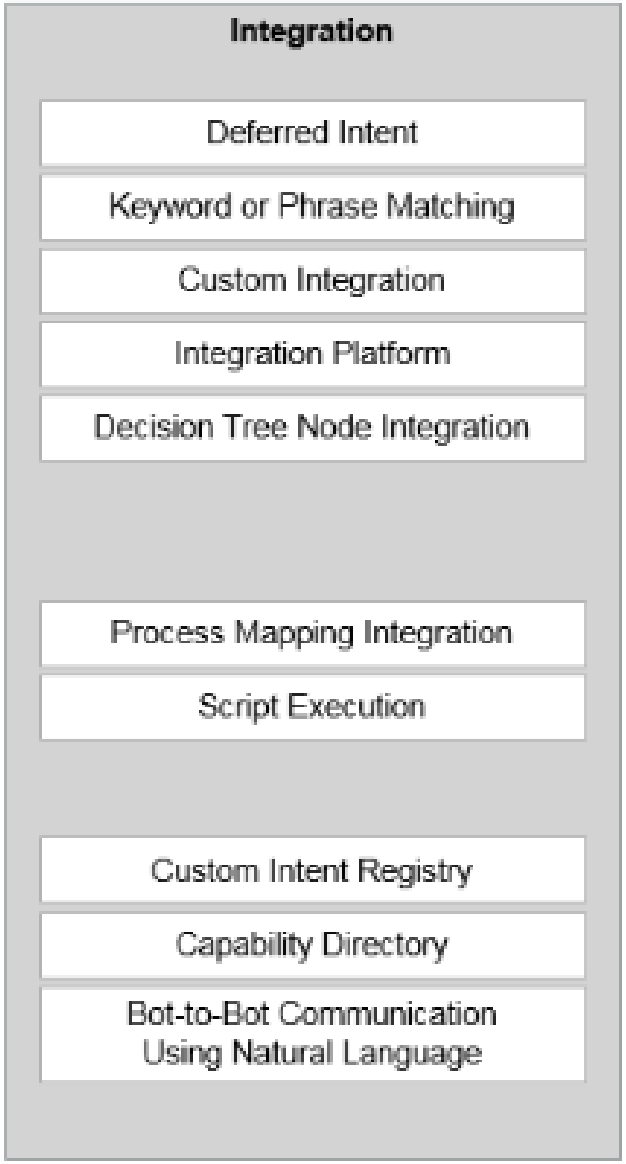
Intent Handling Capabilities



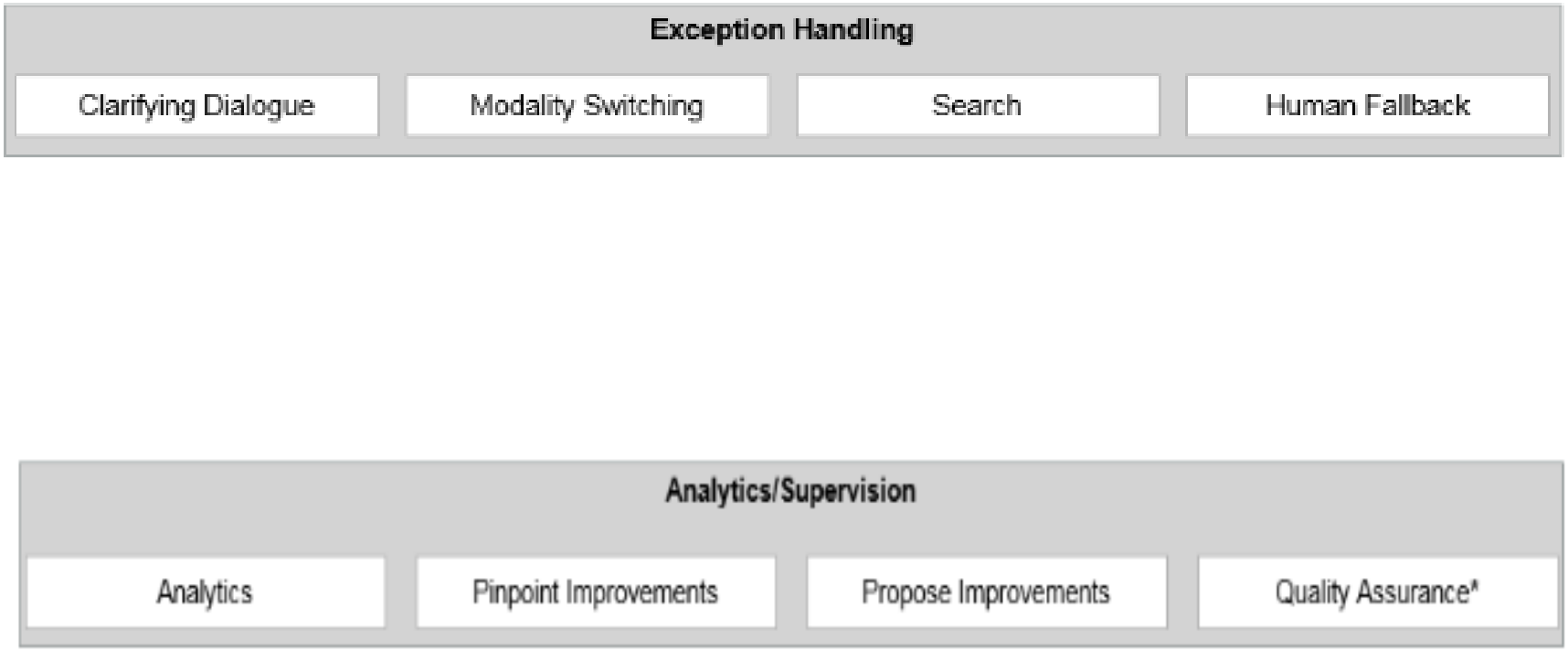
Response Generation Capabilities



Integration Capabilities



Exceptions and Analytics



Major Conversational Platforms

- **Many vendors in 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

CONVERSATION DESIGN

Exercise

- **Team-based**
- **Find a chatbot online, talk to it.**
- **Then state**
  - The objective of the chatbot
  - three things you like about the chatbot
  - three things you don't like
- **We'll share our experiences at:**

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
☐ Users already have human-to-human conversations about this task or topic. ☐ The interaction is brief, with minimal back-and-forth dialog.	**Conversation is intuitive.** It lets users say what they want to get what they want.
☐ Users would have to tap multiple times to complete the task with a screen. ☐ Users might have to navigate multiple apps or widgets to complete the task with a screen. ☐ The feature is difficult or cumbersome to find.	**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.
☐ Users can do this task while multitasking. ☐ Users can do this task when their hands or eyes are busy.	**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.
☐ 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

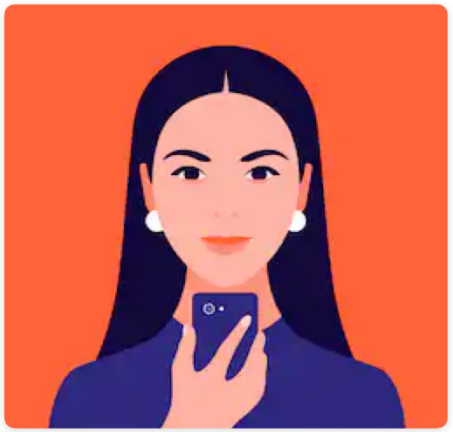
With persona design

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.
- Canceling appointments.
- Sending appointment reminders via push notifications.
- Adding humor to a mundane task.
- Giving the salon a fun, memorable voice.

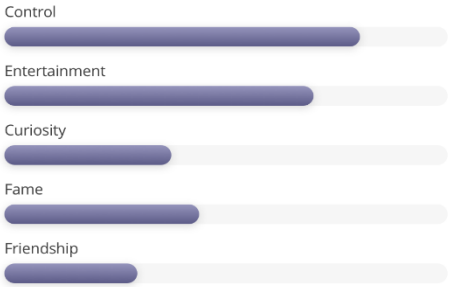
Pain Points

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

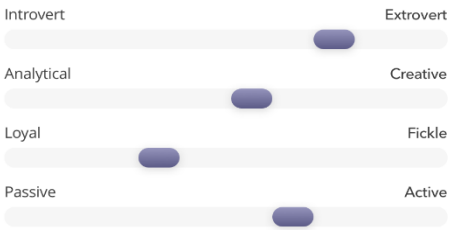
Quote

"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



Personality



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

Step 4. Important things about conversation

- **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

Examples

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 a size 10. Do you have a specific color in mind?

✓

SPOKEN PROMPT:

Got it. Patchwork sneakers. In what size?

✗

Examples

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:

Sorry, how many was that?

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.

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 sho

New Turn

USER INPUT:

The first pair

Turn-taking

- **It keeps the conversation in sync**
- **Relies on cues embedded in sentence structure, intonation, eye gaze, and body language**
- **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)

Examples

USER INPUT:

I want SportsTeam tickets

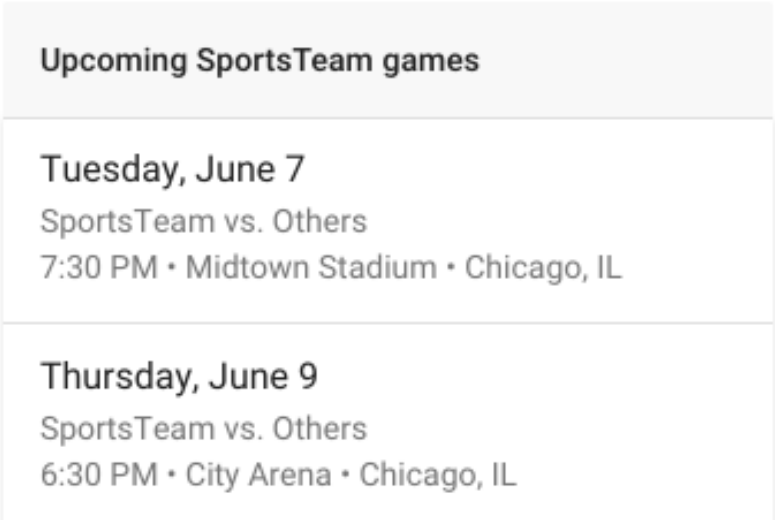
SPOKEN PROMPT:

Here are some upcoming SportsTeam games

DISPLAY PROMPT:

SportsTeam games

VISUAL:



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.

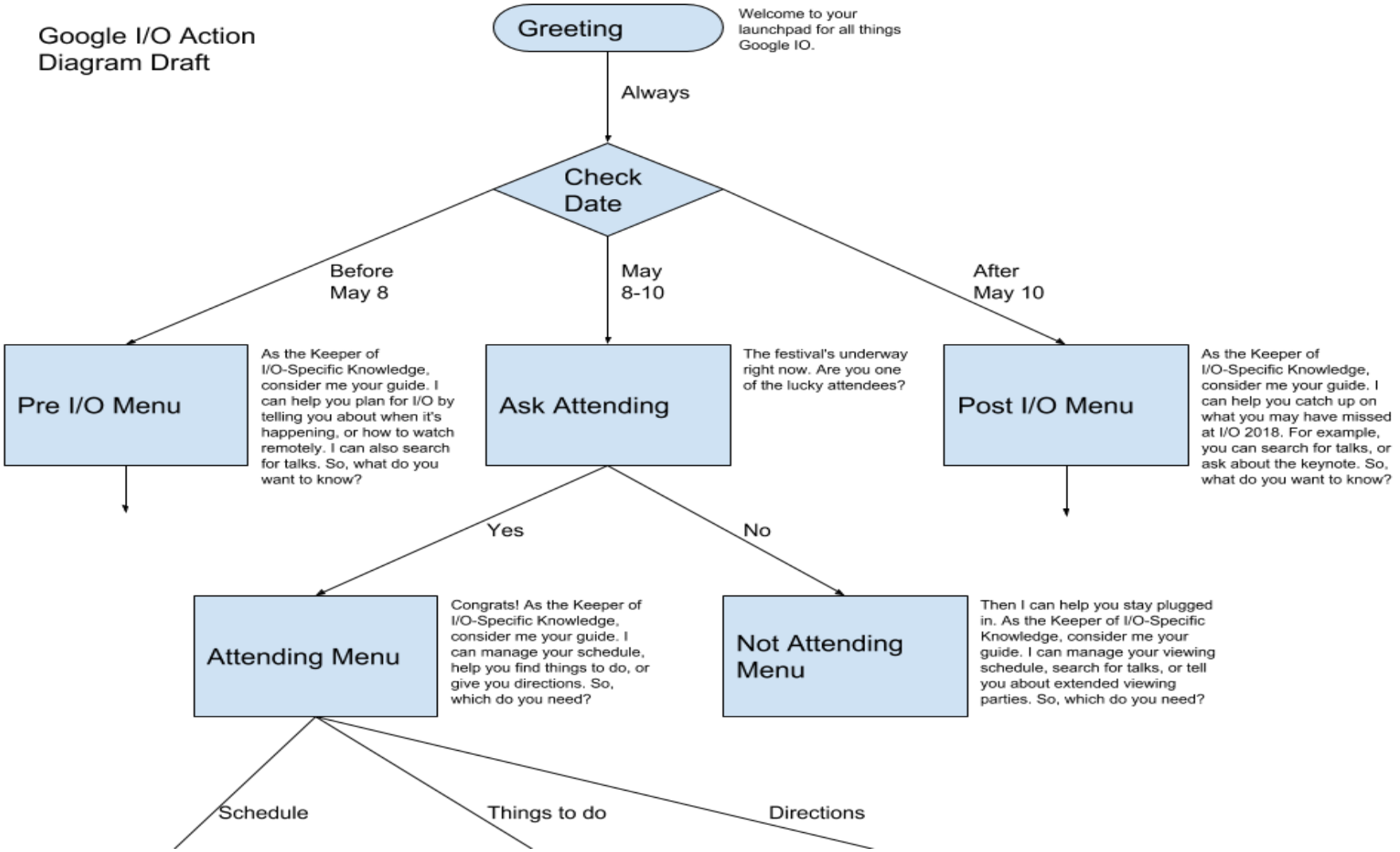
Step 5. Write sample dialogs

- **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?
  - Questions – What size do you want?
  - Acknowledgement – Okay
  - Confirmations – Got it. A pair of men's running shoes. In what size?
  - Information statements
  - Apologies – Sorry

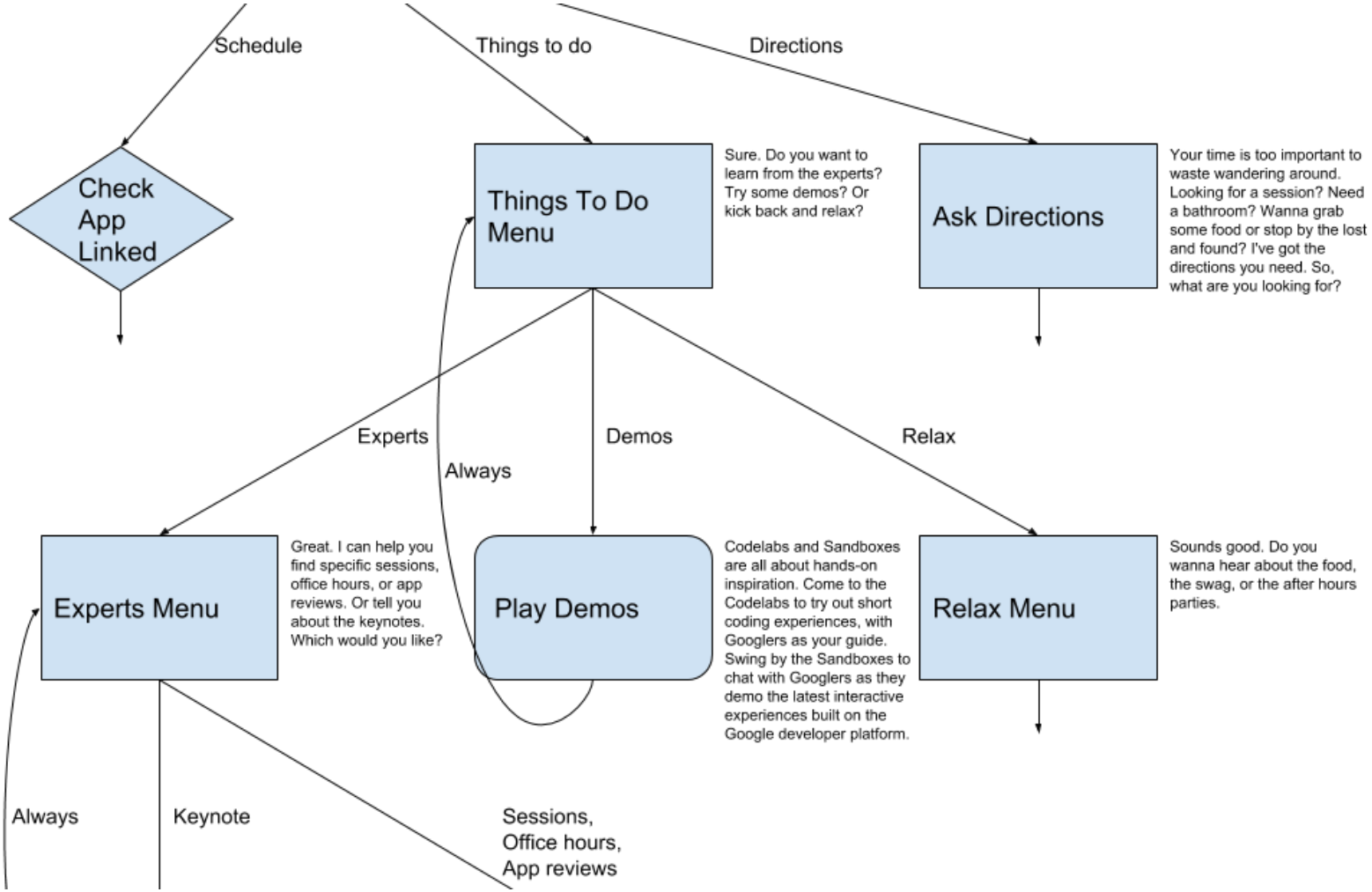
Prompts

- **Conversational components**
  - 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>

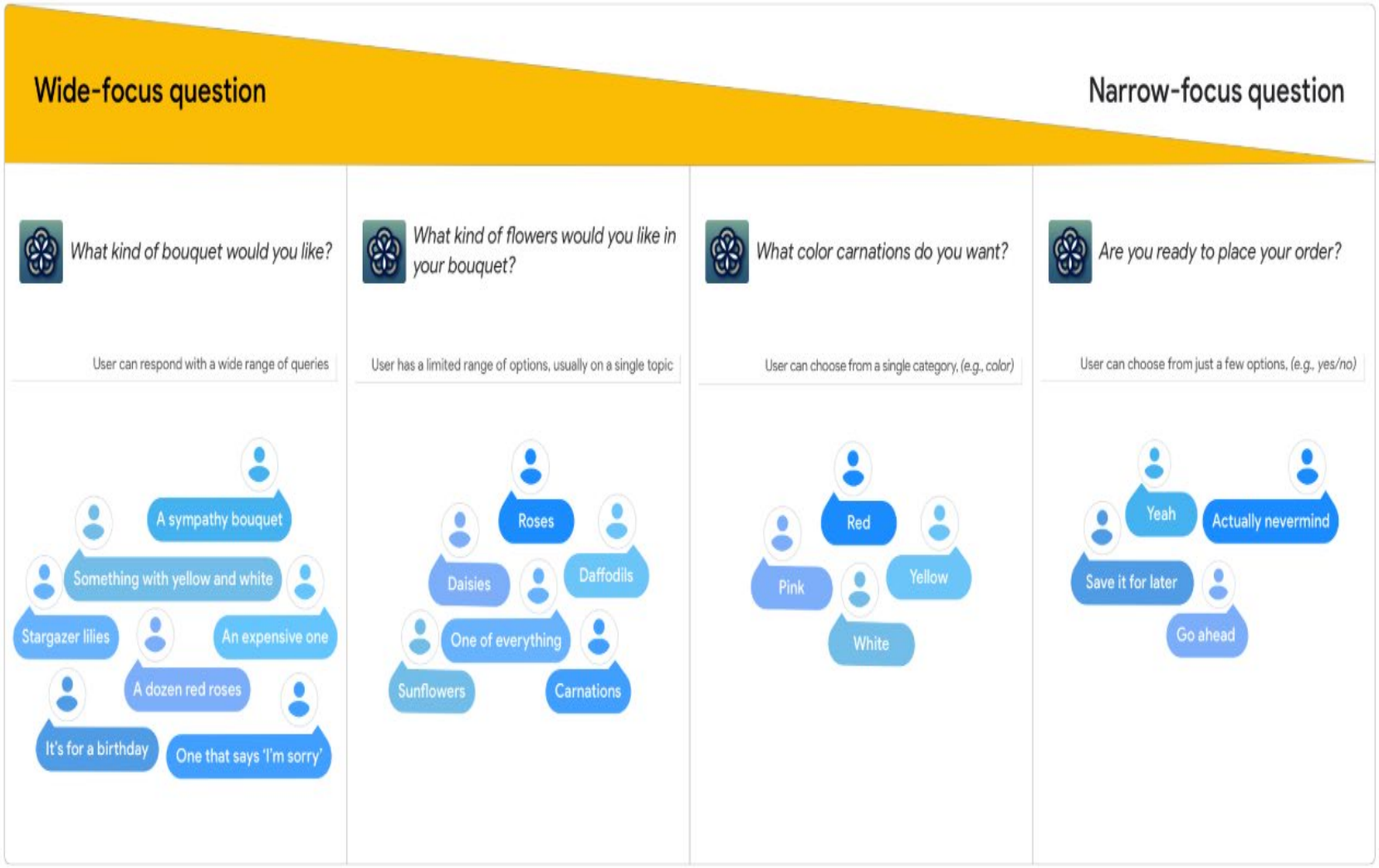
High level flows



High level flows



Questions



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

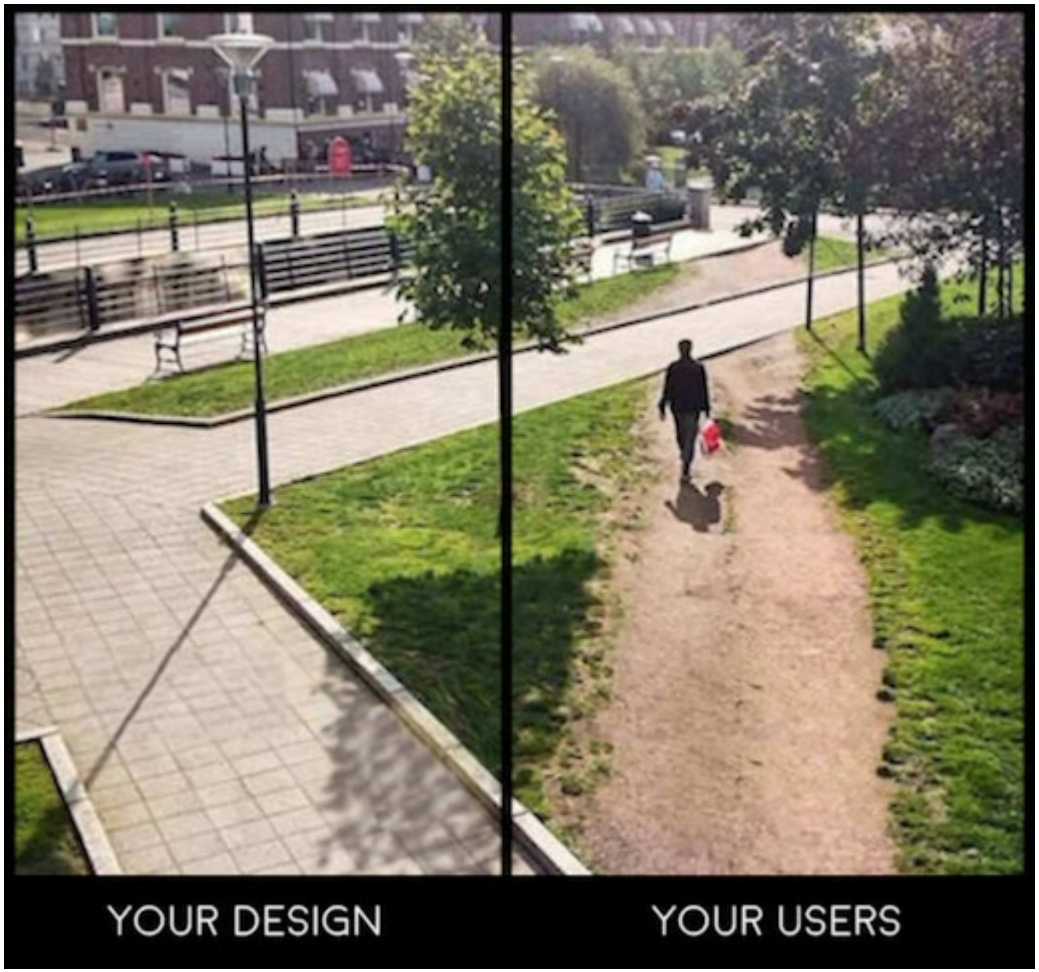
Error handling – no input

- **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

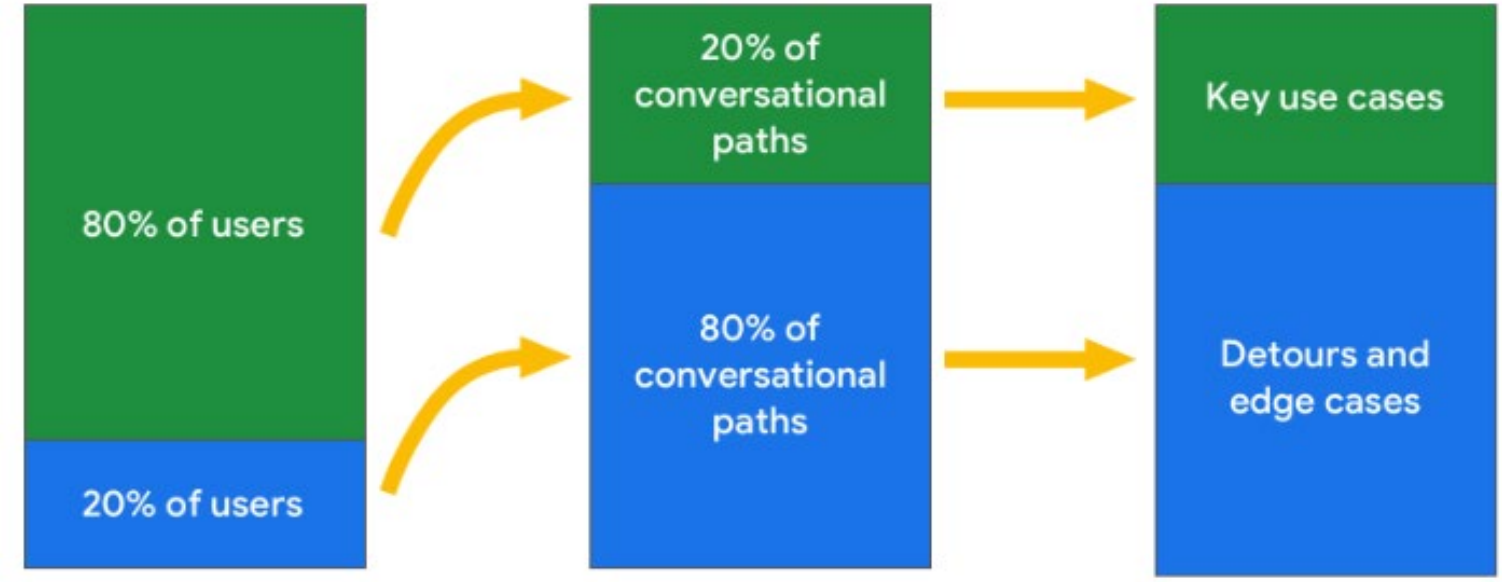


Step 6. Test and iterate

- **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 7. 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 8. 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.

EVALUATING CONVERSATIONAL UI

Testing conversational UI

- **Performance Evaluation**
  - To test the performance of the system and its components
- **Diagnostic evaluation**
  - To detect design and implementation errors
- **Adequacy evaluation**
  - To test how well the system meets its objectives and the users' expectations

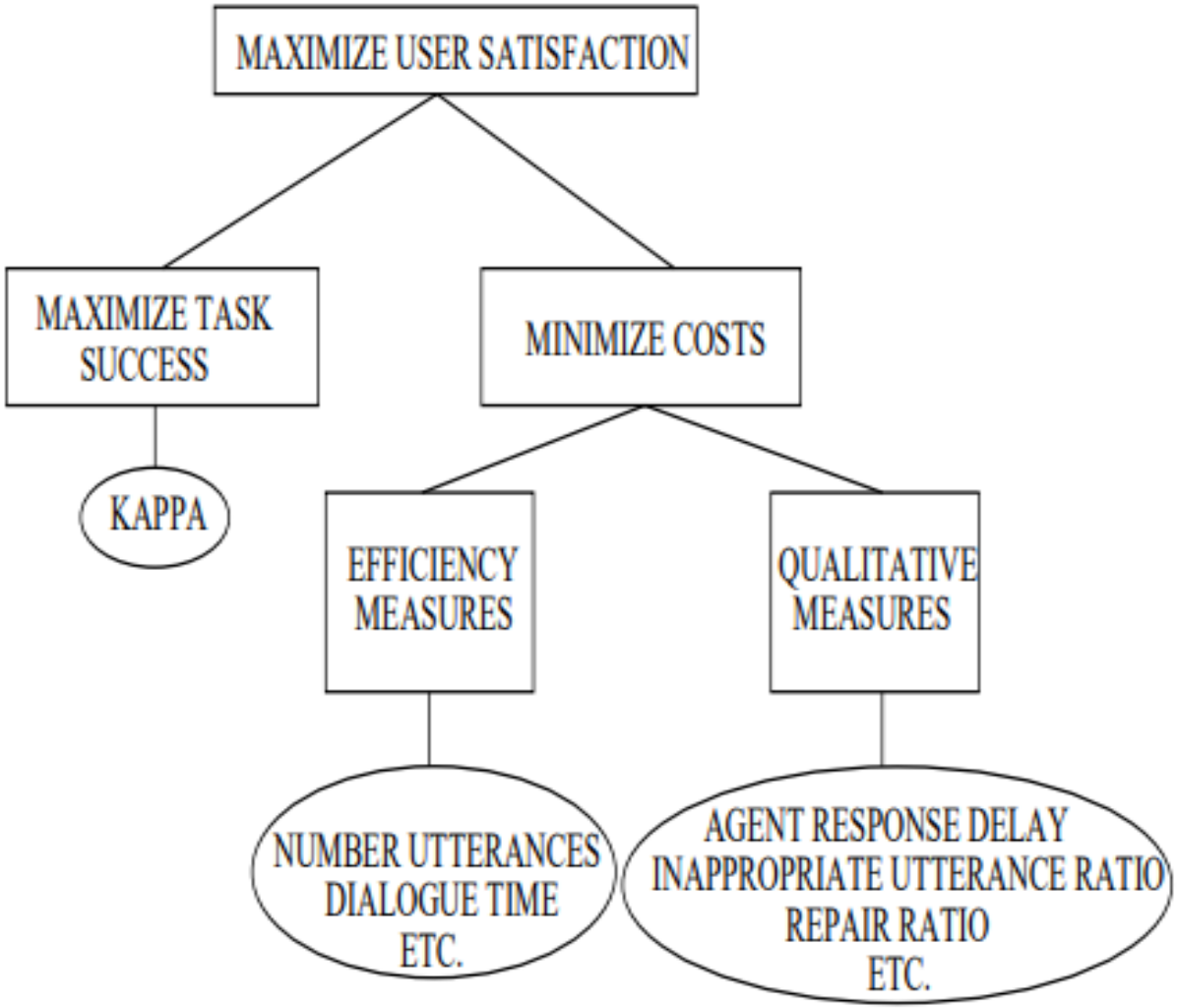
Common Metrics

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

PARADISE

- **PARAdigm for DIalogue System Evaluation**
  - A general framework for evaluating spoken dialogue agents
  - Performance is modeled as a weighted function of a task-based success measure and dialogue-based cost measures
  - Assumption: performance correlates with a meaningful external criterion such as usability.
  - Requires dialog corpora extracted from controlled experiments – users evaluate satisfaction after interacting with the system

Objectives for spoken dialogue performance



Component Evaluation

Evaluating the following components independently:

- Speech recognition
- Spoken language understanding
- Dialog management
- Natural language generation
- Text-to-speech synthesis

E.g. Evaluating 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 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

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

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

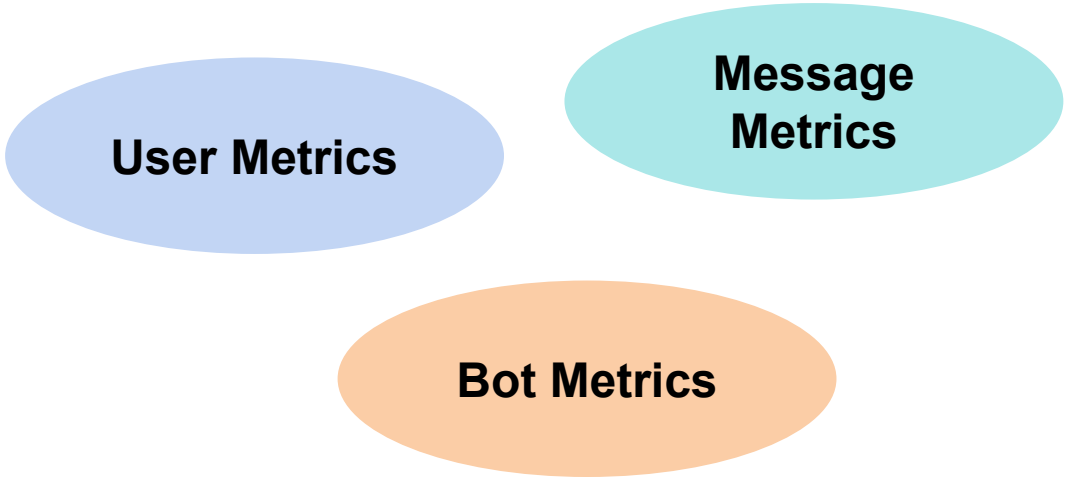
The system is useful
The system is pleasant
The system is friendly
I was able to recover easily from errors
I enjoyed using the system
It is clear how to speak to the system
It is easy to learn to use the system
I would use this system
I felt in control of the interaction with the system

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

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

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



User Metrics

- **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

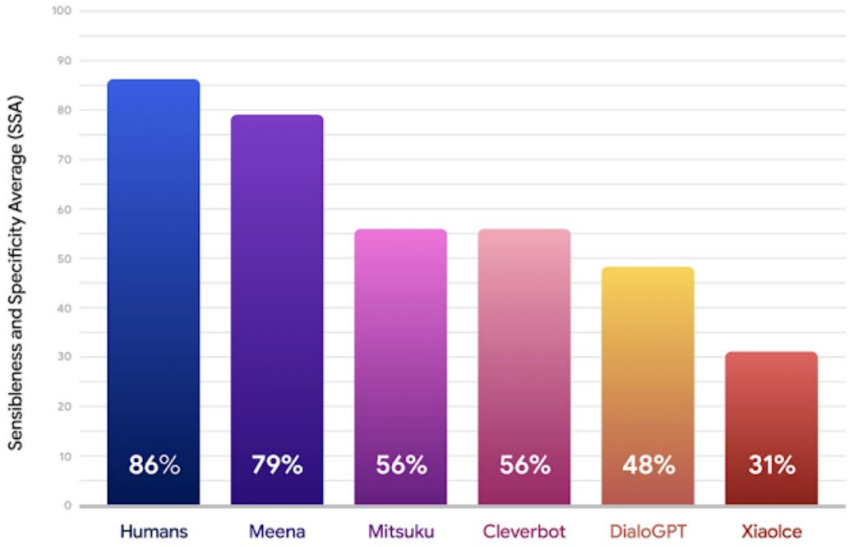
- **How individuals interact with the bot**
  - Conversation starter messages – number of messages from the bot initiating interaction
  - Bot messages – total number of messages sent by the bot in each interaction (conversation length)
  - In messages – number of messages sent by users in each interaction
  - 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**
  - 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

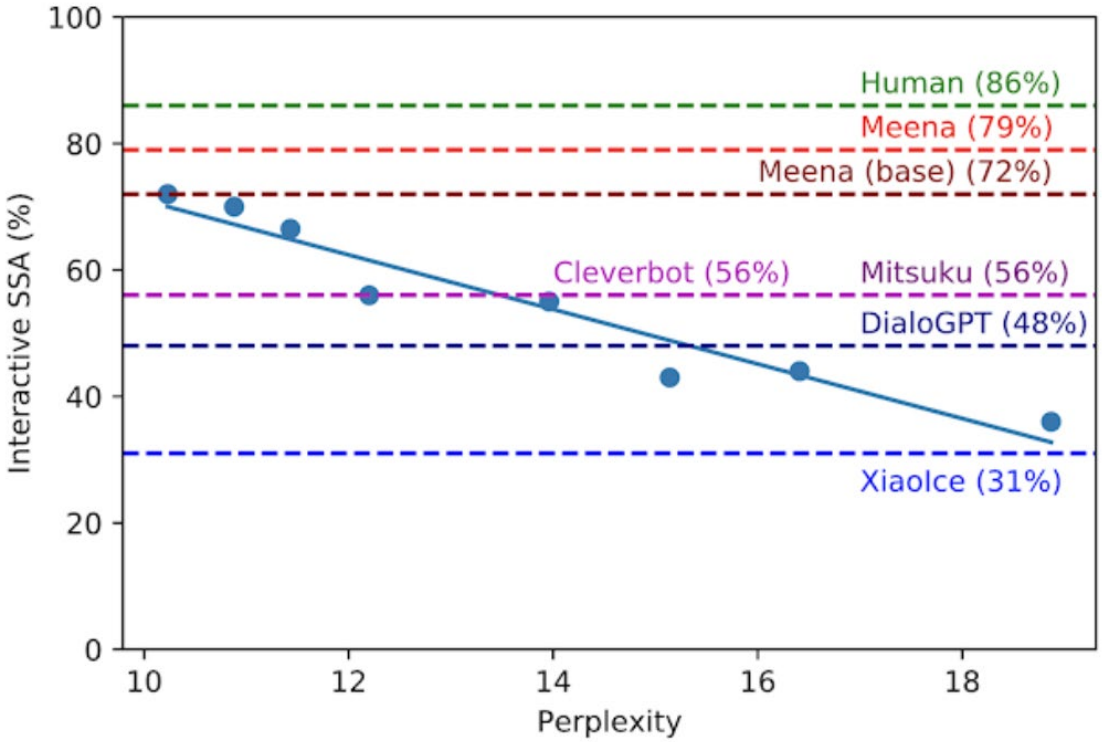
- **Naturalness of conversation**
  - for end-to-end neural conversational agents, which need to handle a wide variety of topics
- **Human evaluation metric: Sensibleness and Specificity Average (SSA)**
  - Proposed by Google, using common sense to judge if a response is reasonable and specific in the given context.
  - average of the sensibleness score (percentage of sensible responses) and specificity score (percentage of specific responses)
  - crowd-sourced evaluation



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

Exercise

- **Form discussion groups**
- **Each group picks a chatbot task and determine the objective of the chatbot**
- **Gather requirements**
  - Identify users
  - Identify key use cases
  - Choose a persona
- **Design the conversation**
  - Create the conversation flow
  - Choose one key use case, and create sample dialogue
  - Handle a few exceptions