

Next-Generation Intelligent Assistants for AR/VR Devices

Xin Luna Dong @ Meta

4/2022

This talk does not represent the company's point of view

Everyone Deserves An Assistant



What is A Virtual Intelligent Assistant?

Respond to commands



“Hey Siri, set a timer to 7pm”

“Ok, added to today’s reminders”



What is A Virtual Intelligent Assistant?

Control devices



“Hey Alexa, turn off bedroom lights”



What is A Virtual Intelligent Assistant?

Provide information



“Hey, Google, when is Easter?”

“Easter will be on Sunday, April 17th.”



Meta's Assistant

Empowering connection to people and experiences in your life

Facebook Portal



“Hey Portal”

- Hop on a call hands-free
- Get help with music, timers, alarms, weather, show photos from your Facebook profile, and more.

Meta Quest 2



“Hey Facebook” (double press the button on your controller)

“Who’s online?”--meet up with friends

“Open Beat Saber”--jump straight in the game, and more.

Ray-Ban Stories



“Hey Facebook, take a picture” -- capture moments hands-free

“Hey Facebook”--call friends on Messenger, manage device settings, and more.

What is An Ideal Assistant?



Follow instructions



Know your needs



Capable



Knowledgeable

What is An Ideal Virtual Intelligent Assistant?

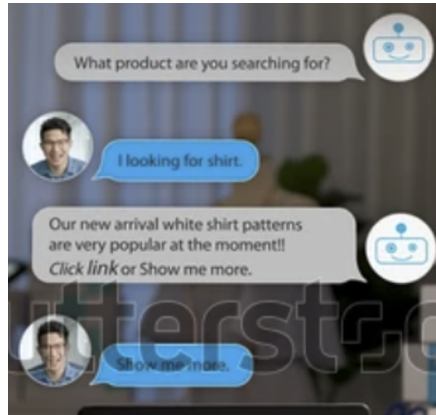
An *intelligent assistant* should be an agent that **knows you and the world**, can **receive your requests** or **predict your needs**, and provide you the **right services at the right time** with your permission.



Three Generations of Intelligent Assistant

V0.1 Chatbot

Text input



V1. Voice Asst

Voice input



V2. AR/VR Asst

Voice + Visual + Context



Structure of the Talk

Outline

- What is an Intelligent Assistant?
- Techniques to support current intelligent assistants
- Challenges and initial solutions for the next generation of intelligent assistants

Goals

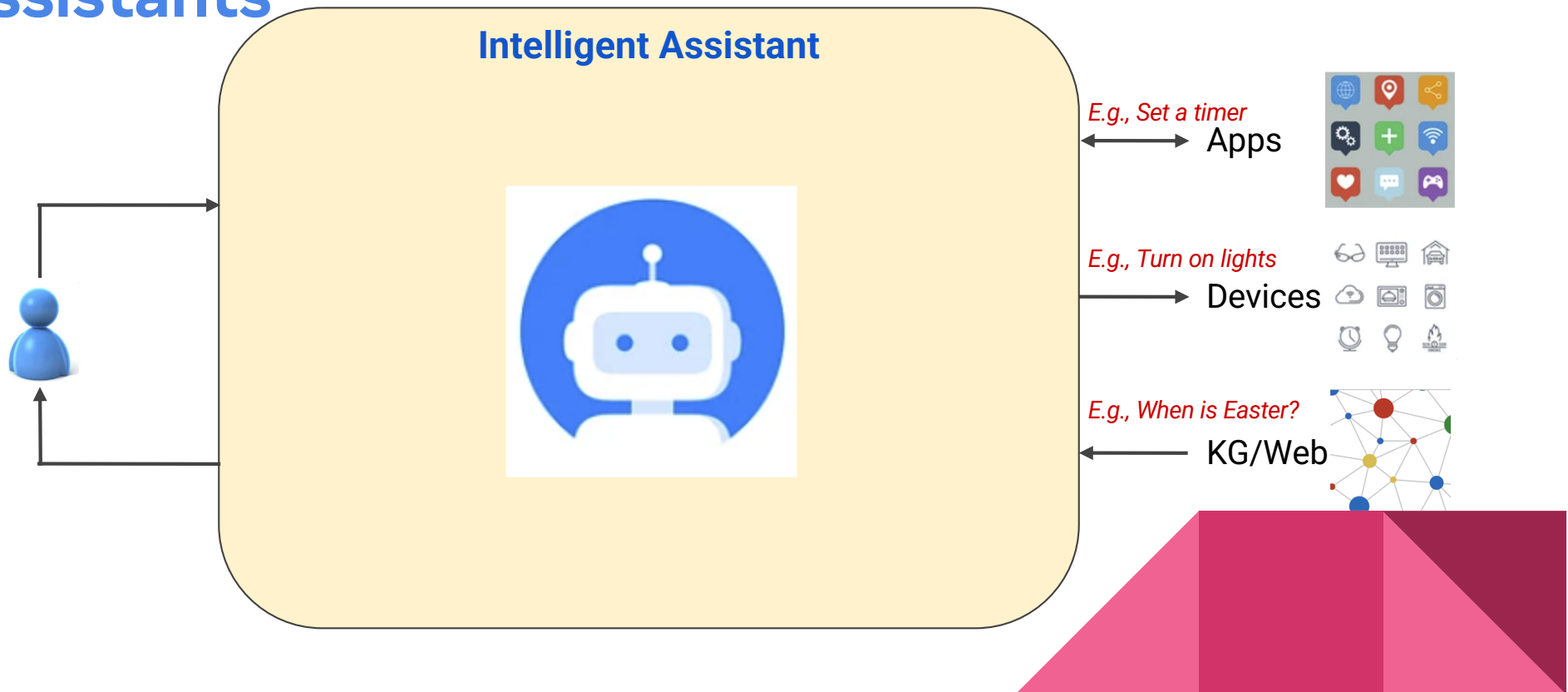
- Introduce you the interesting research problems for Intelligent Assistants
- Impress you with the many ML fields touched by Intelligent Assistants
- Invite you to open new doors to build next-generation Intelligent Assistants



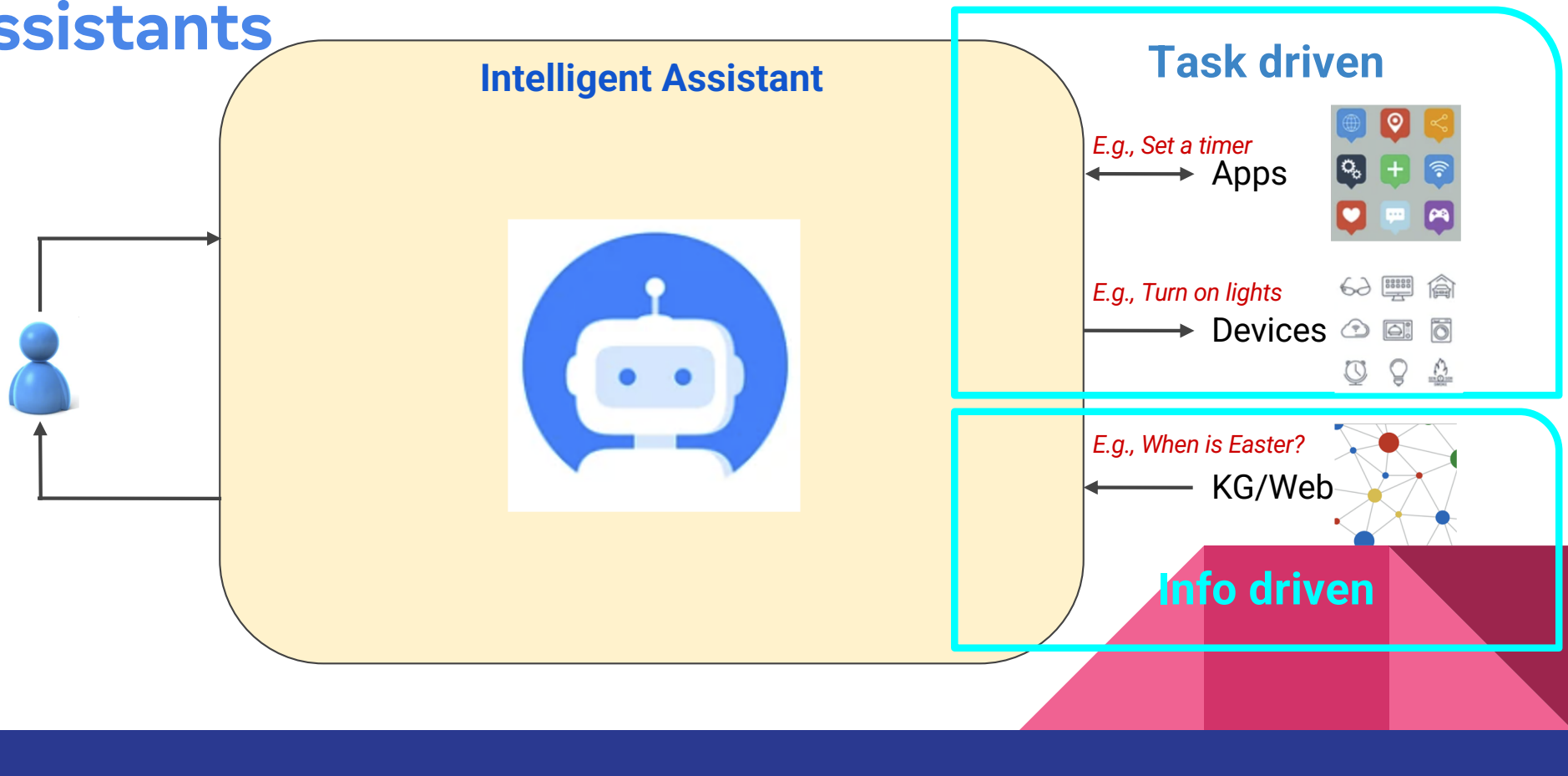


Technologies to Support Current Intelligent Assistants

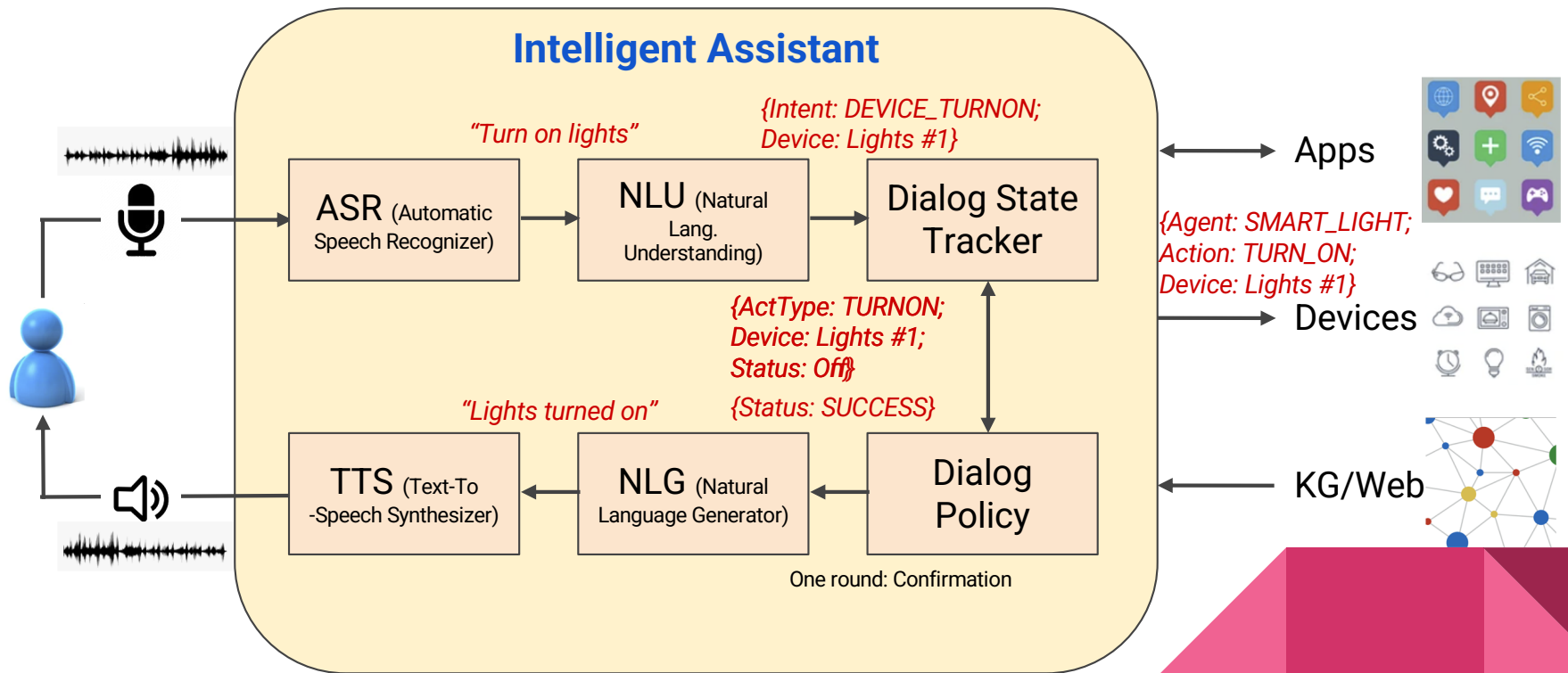
Two Types of Commands to Intelligent Assistants



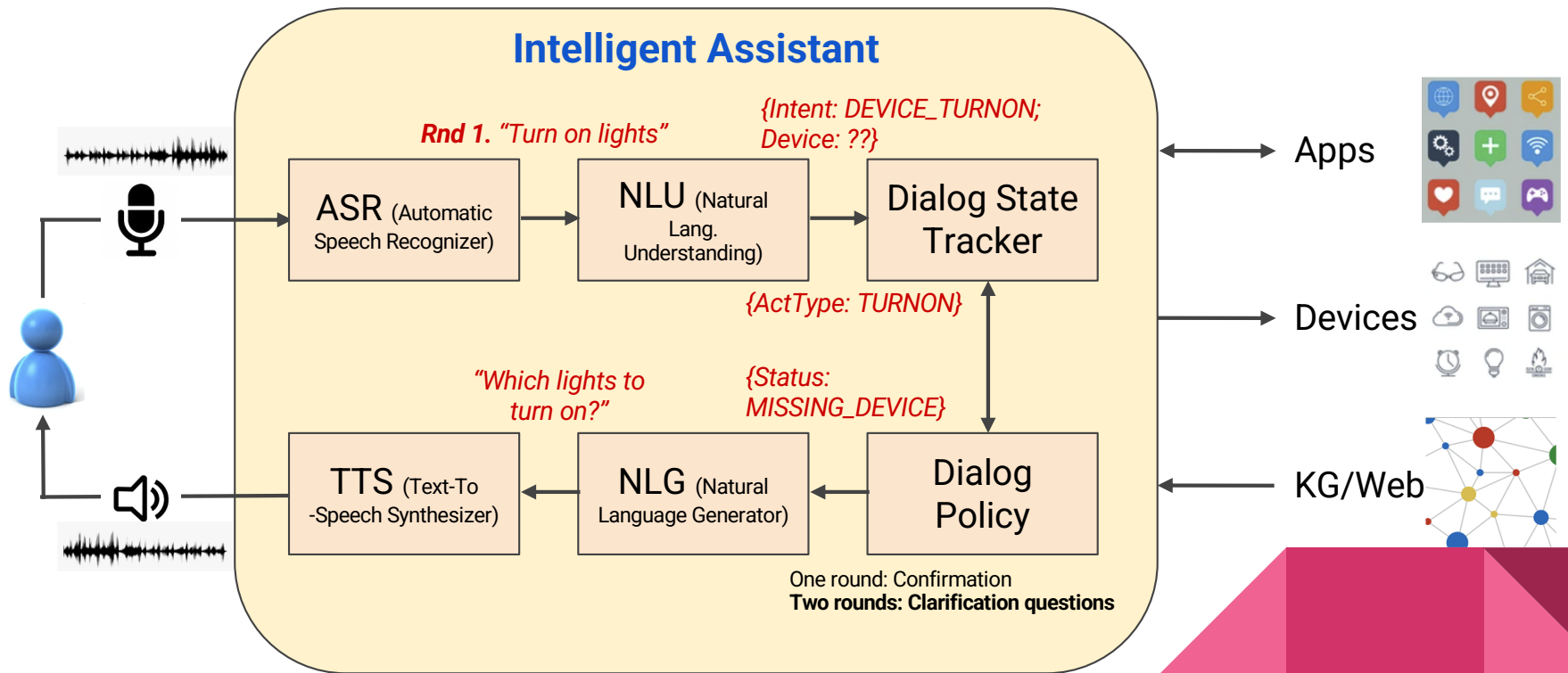
Two Types of Commands to Intelligent Assistants



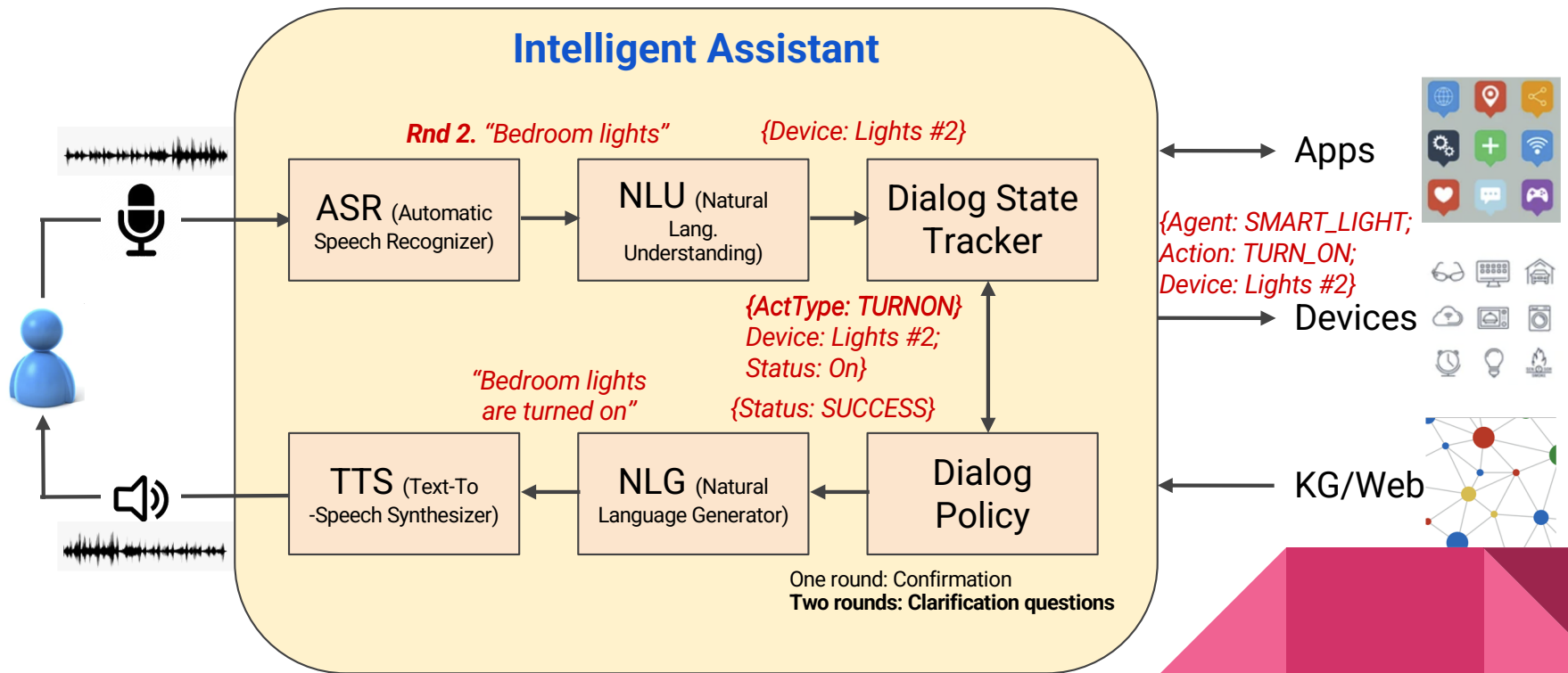
Int. Asst. Is Essentially a Conversation System



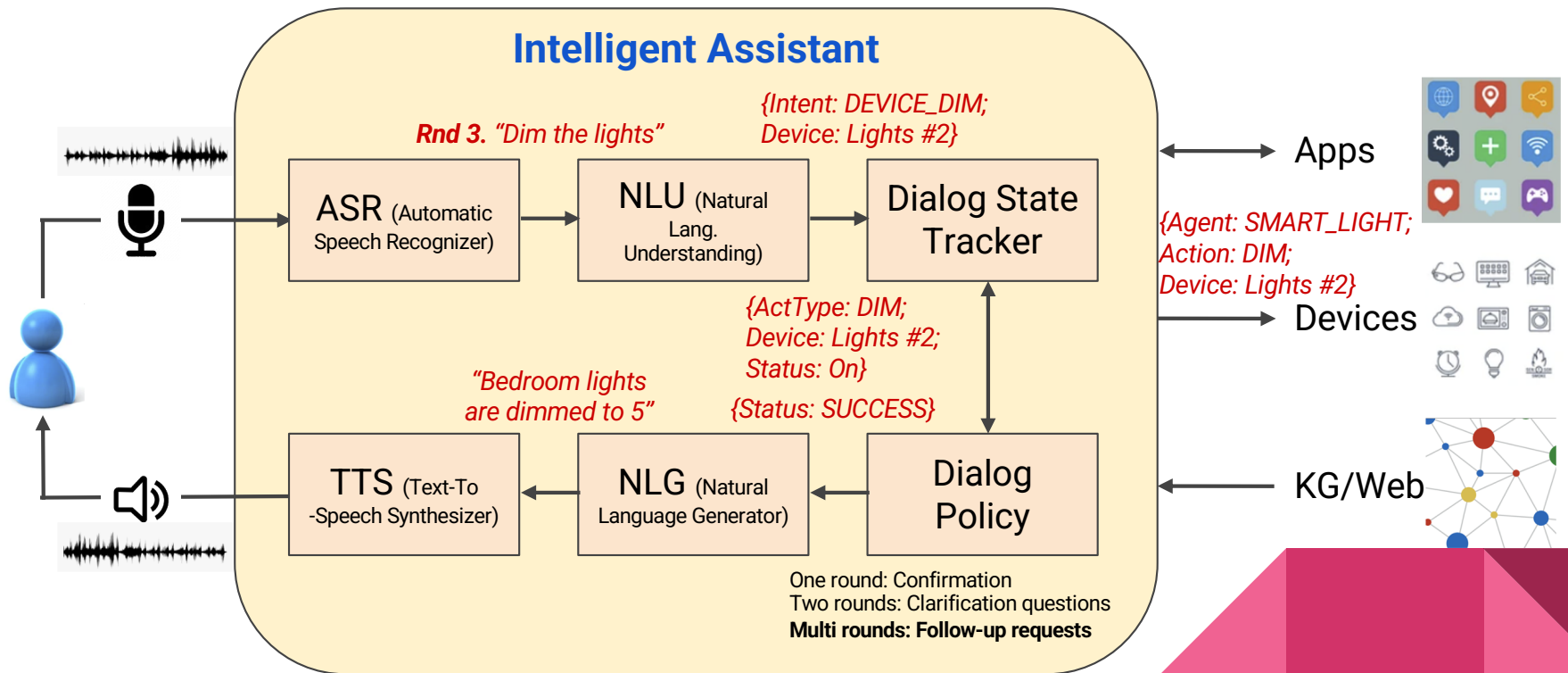
Int. Asst. Is Essentially a Conversation System



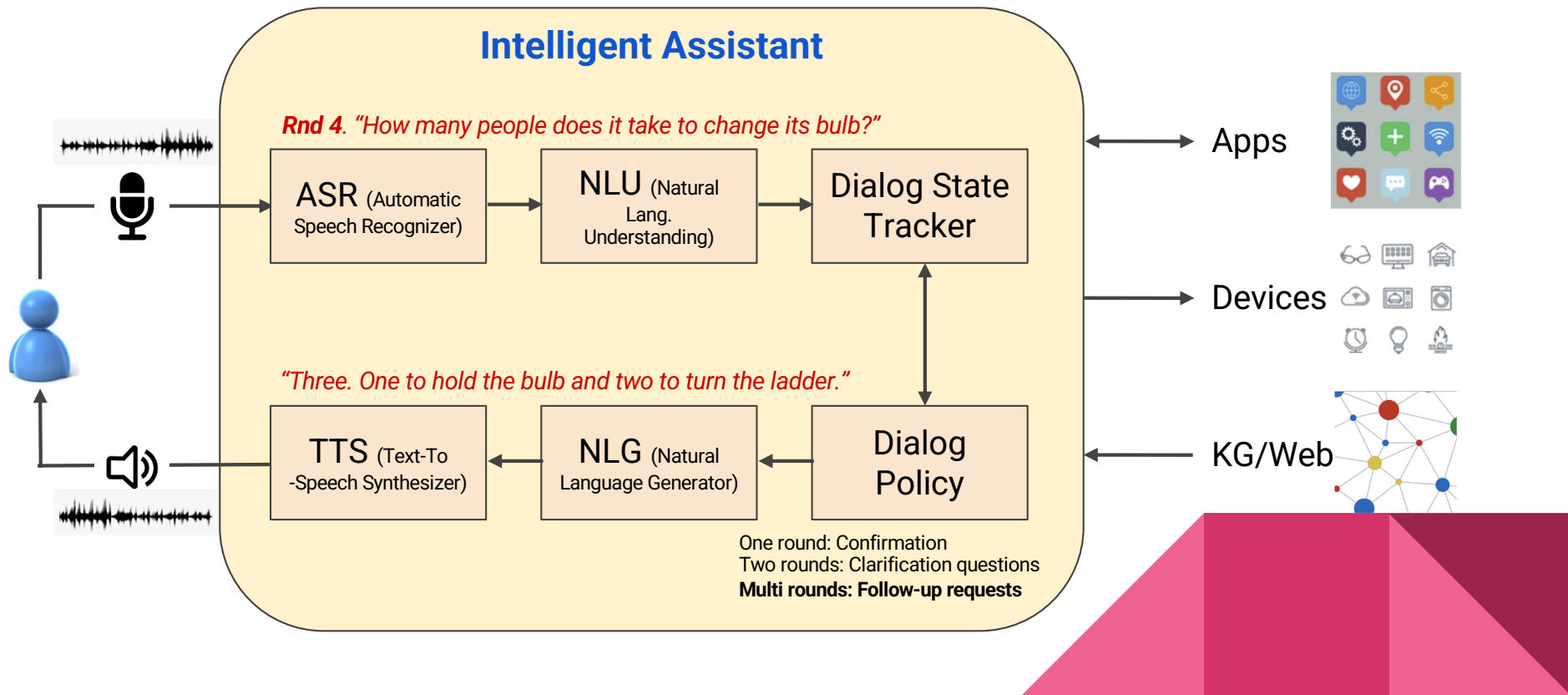
Int. Asst. Is Essentially a Conversation System



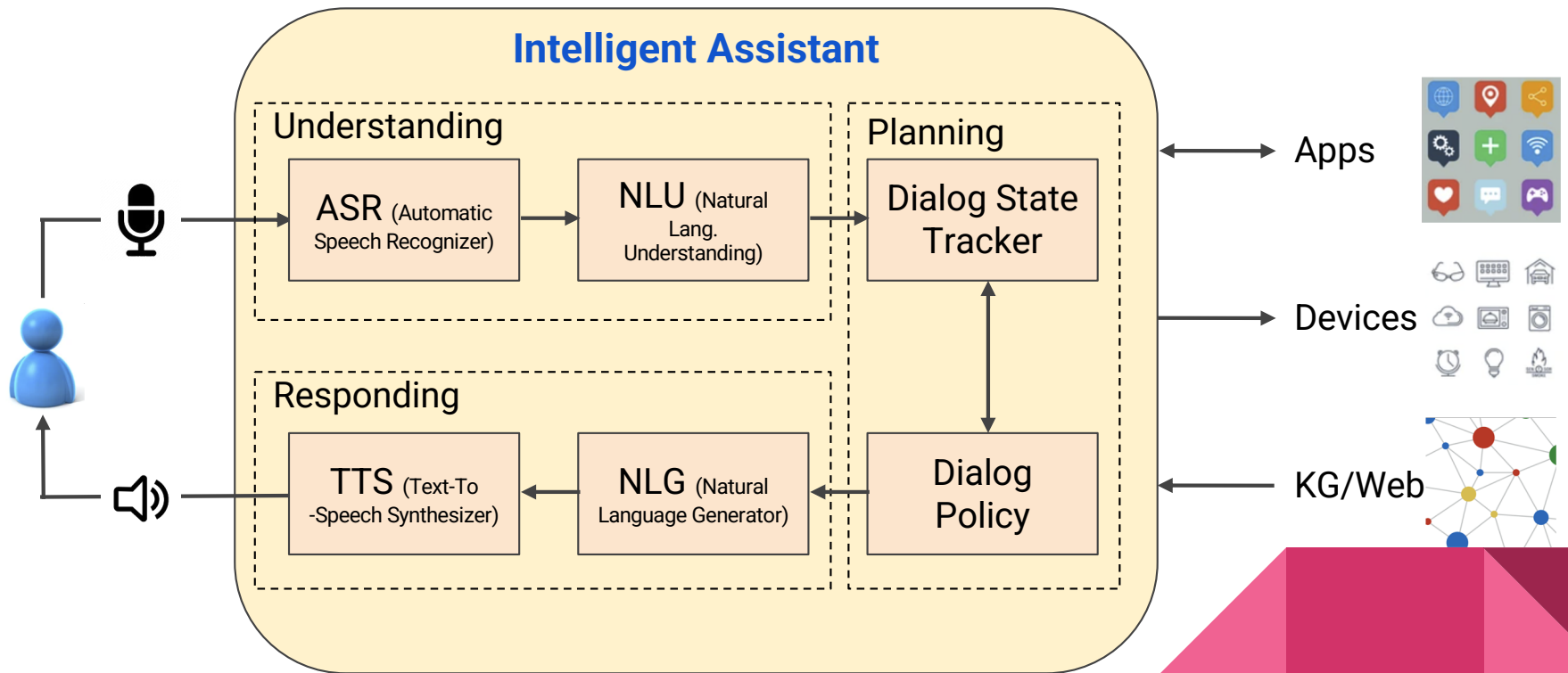
Int. Asst. Is Essentially a Conversation System



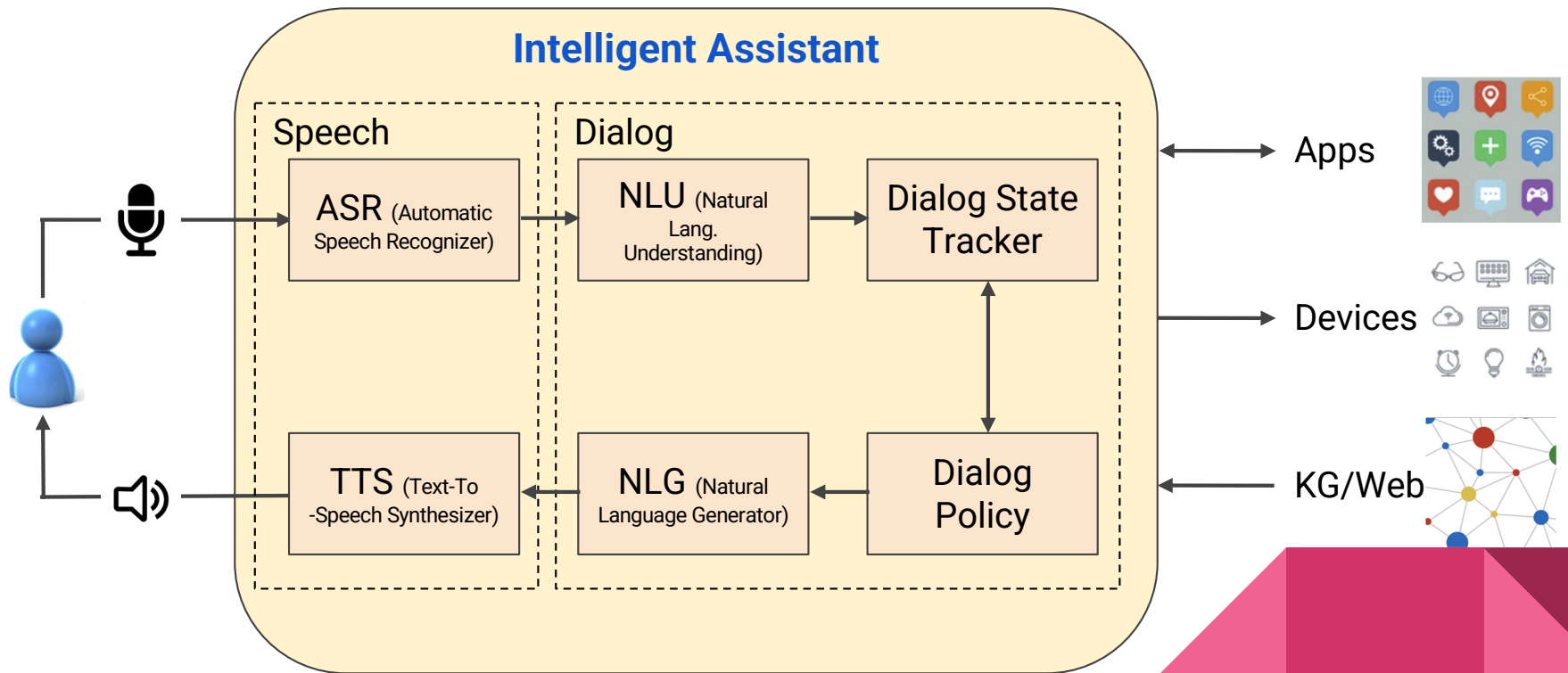
Int. Asst. Is Essentially a Conversation System



Int. Asst. Is Essentially a Conversation System

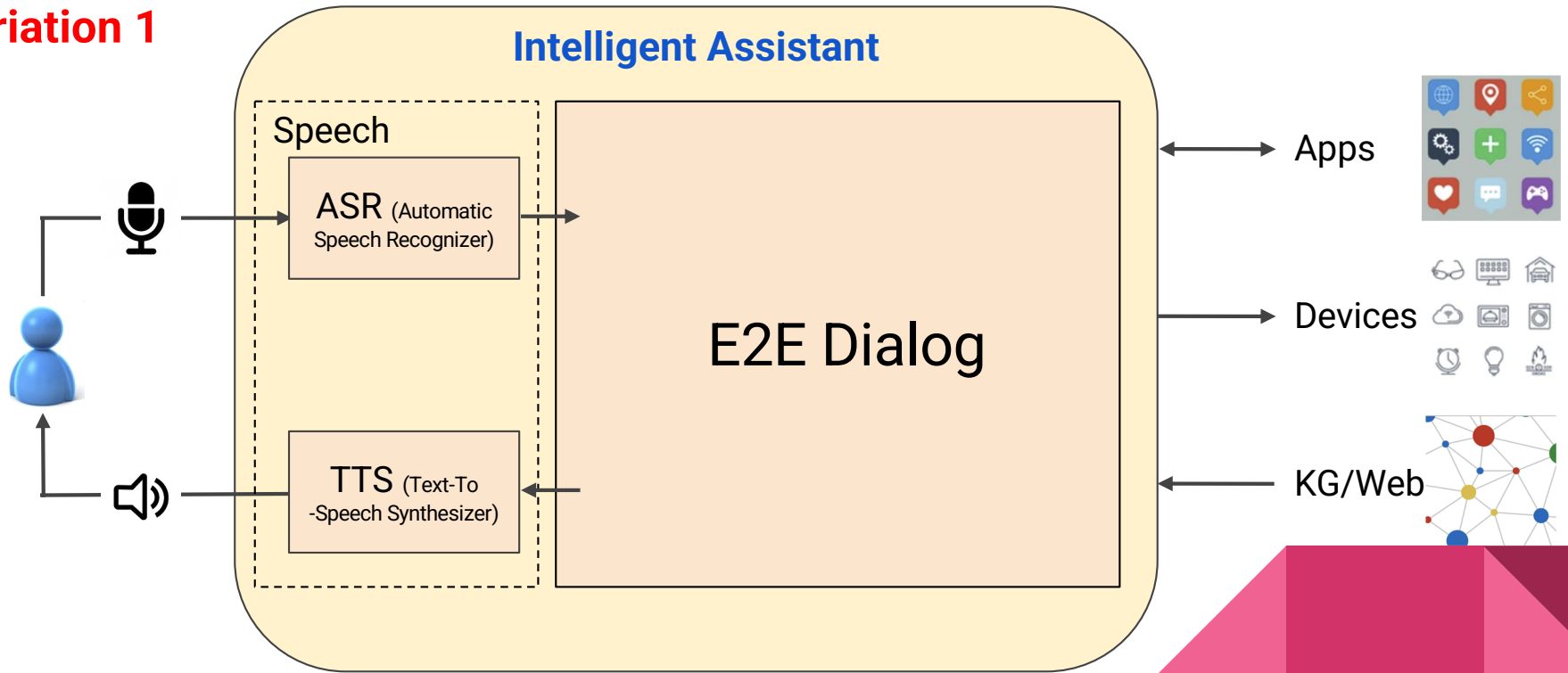


Int. Asst. Is Essentially a Conversation System



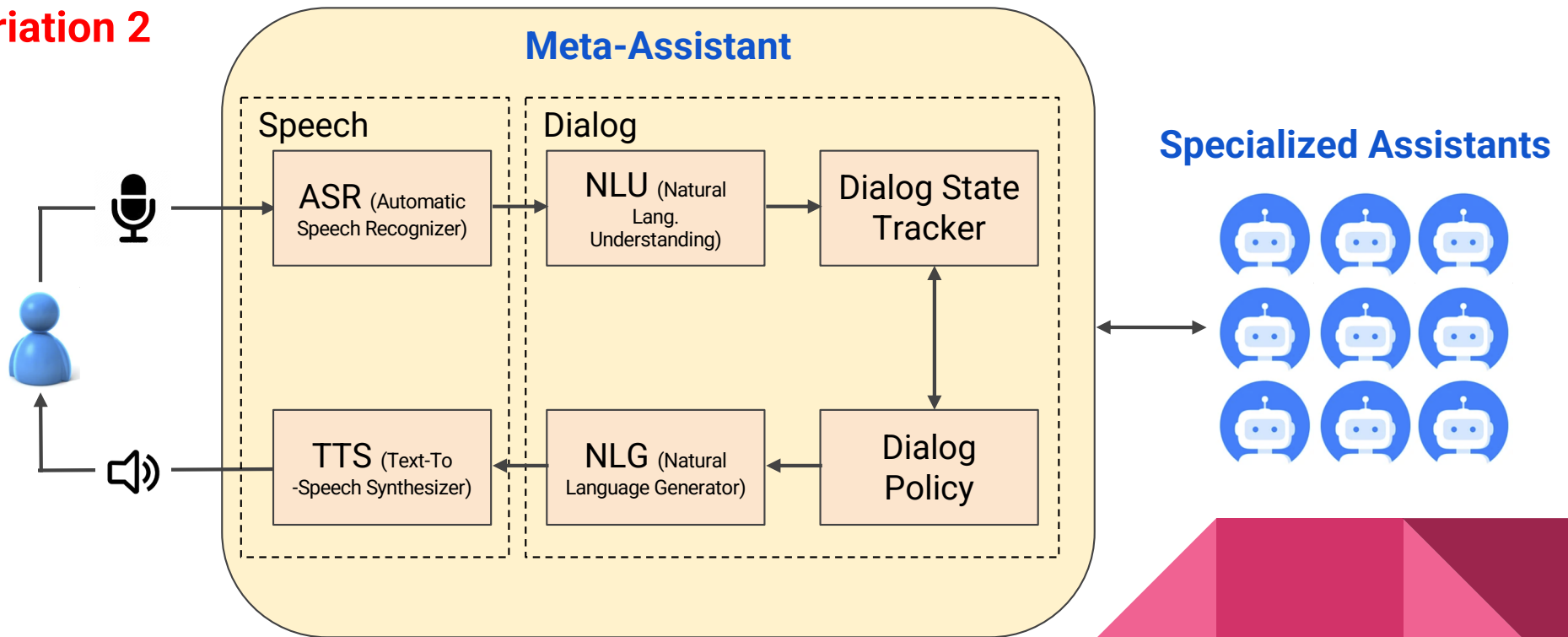
Int. Asst. Is Essentially a Conversation System

Variation 1

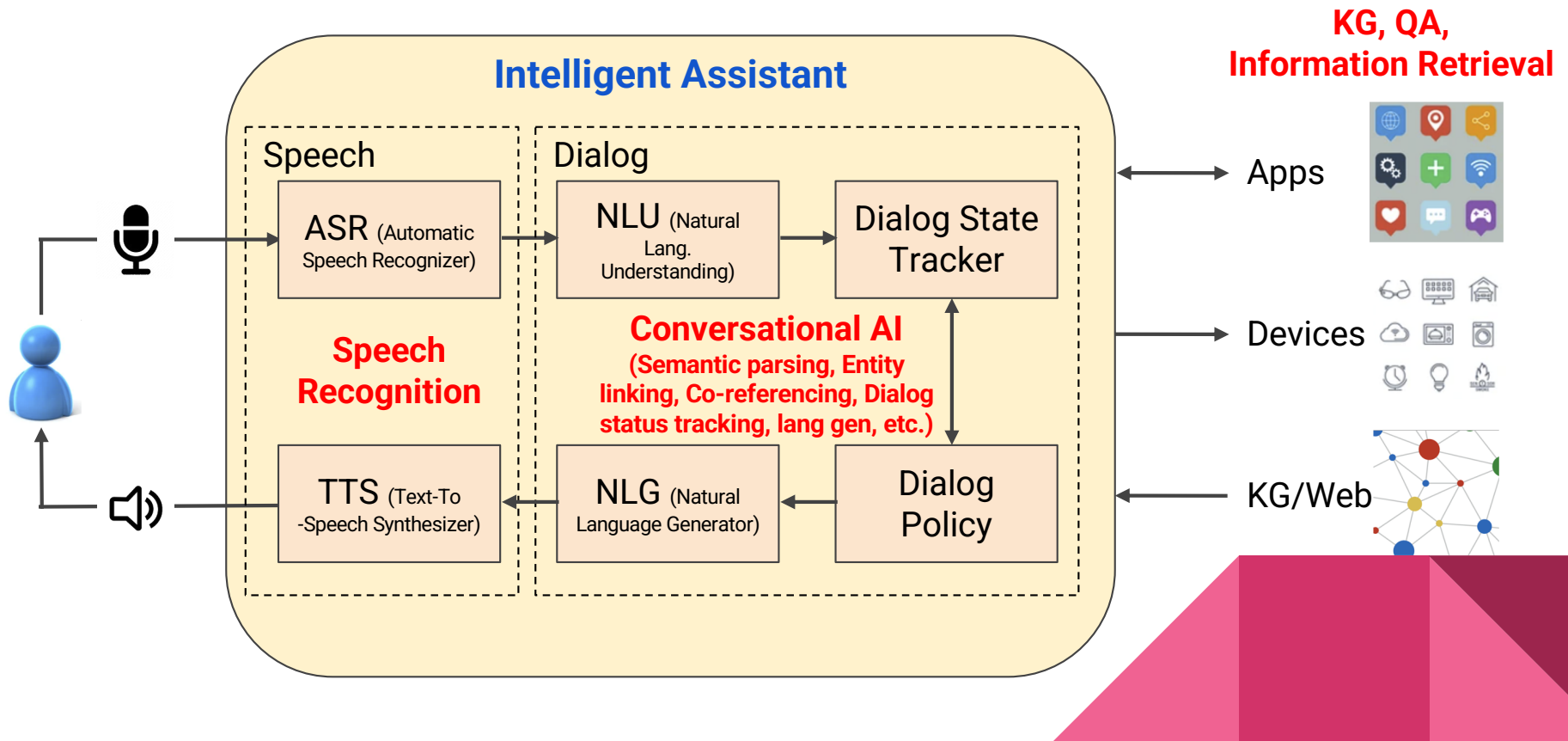


Int. Asst. Is Essentially a Conversation System

Variation 2



Related Research Areas



What Can Be Improved?



- How to increase accuracy?
- How to allow easy scale-up to new tasks, new domains, and new languages?
- How to make the assistants a know-it-all?
- How to make the multi-turn conversations smoother?
- etc.



Ideal Assistant Revisited–Missing Pieces

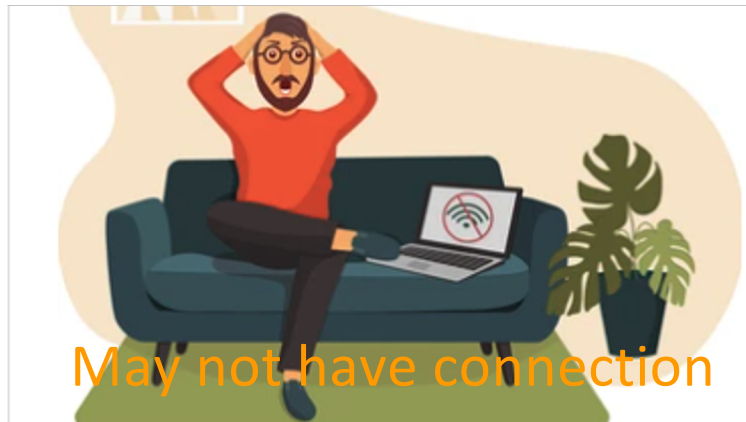
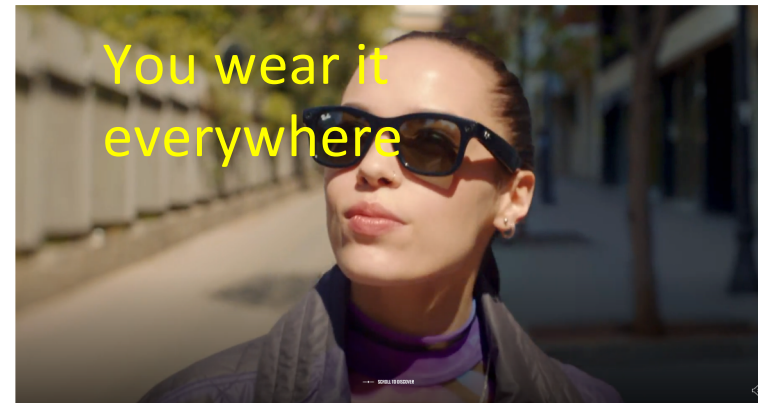
An *intelligent assistant* should be an agent that **knows you and the world**, can **receive your requests** or **predict your needs**, and provide you the **right services at the right time** with your permission.





Challenges and Initial Solutions to Next-Generation AR/VR Assistants

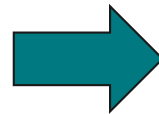
What Is Different for An AR/VR Assistant?



From Voice-Only to Multi-Modal



“How tall is Empire State Building?”

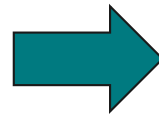


“What’s the name of this building and how tall is it?”

From Context-Agnostic to Context-Aware



“Show my shopping list”



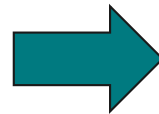
“Remember to buy apples and bananas at the grocery store around the corner”



From Reactive to Proactive



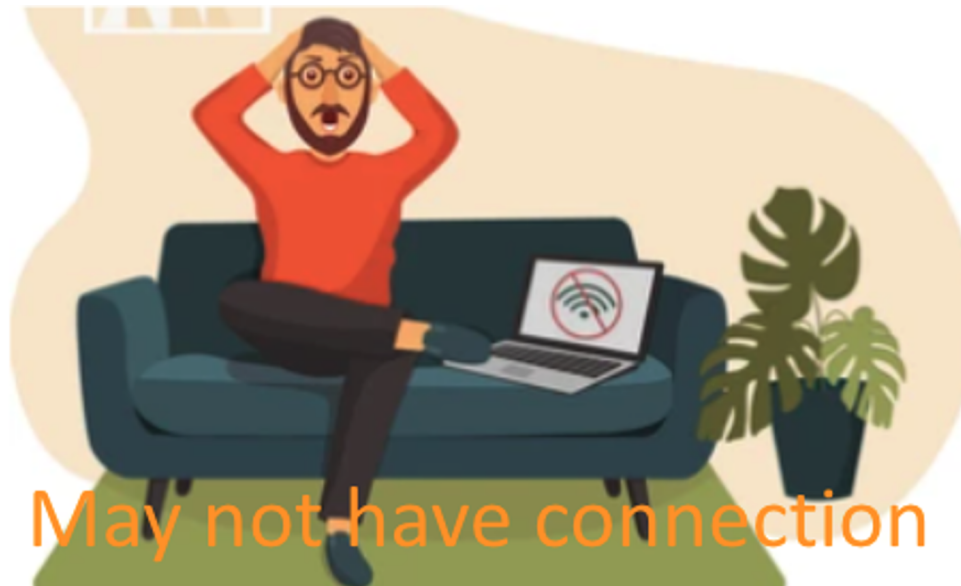
“What’s the weather today?”



“Today is sunny, 70 degree. Would you like to play your favorite morning music?”



From Server-Side to On-Device

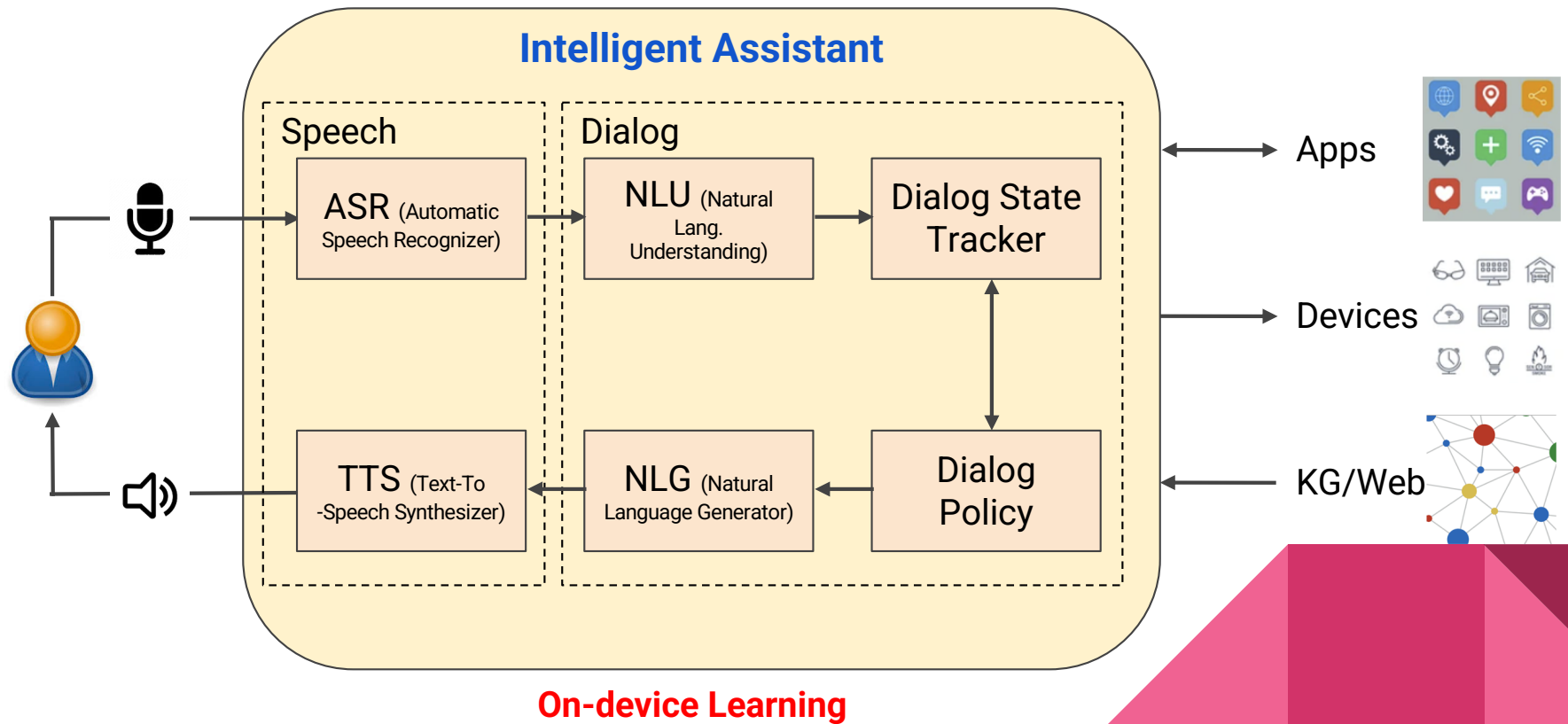


May not have connection

+ Privacy!!!



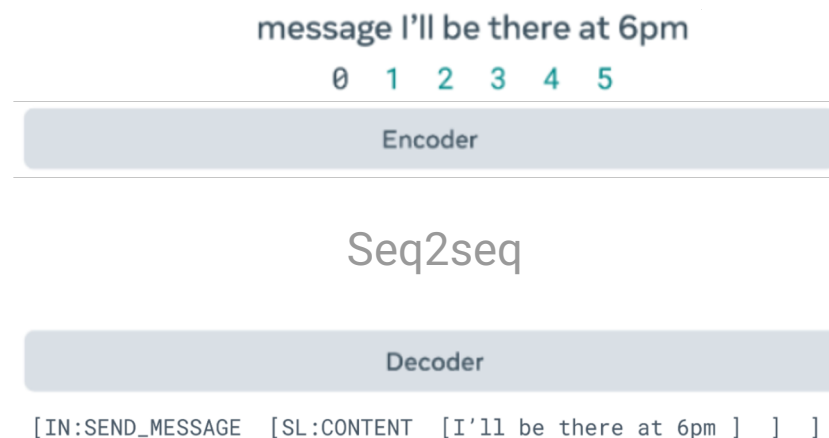
Direction 1. On-Device Machine Learning



Direction 1. On-Device Machine Learning

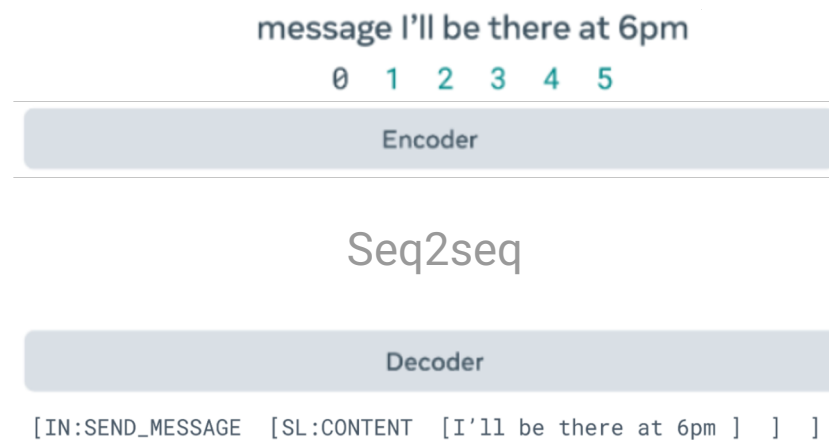
Traditional Autoregressive Semantic Parsing

- Pros
 - High accuracy
- Cons: Prohibitively expensive
 - ⇒ Server-side modeling
 - Flaky user experiences w. spotty internet connectivity
 - High latency
 - Compromised user data privacy



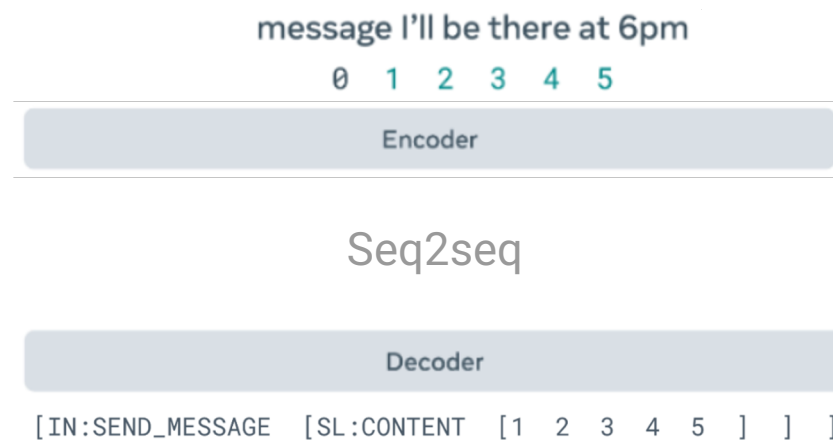
Direction 1. On-Device Machine Learning

Non-Autoregressive Semantic Parsing: Parallel prediction



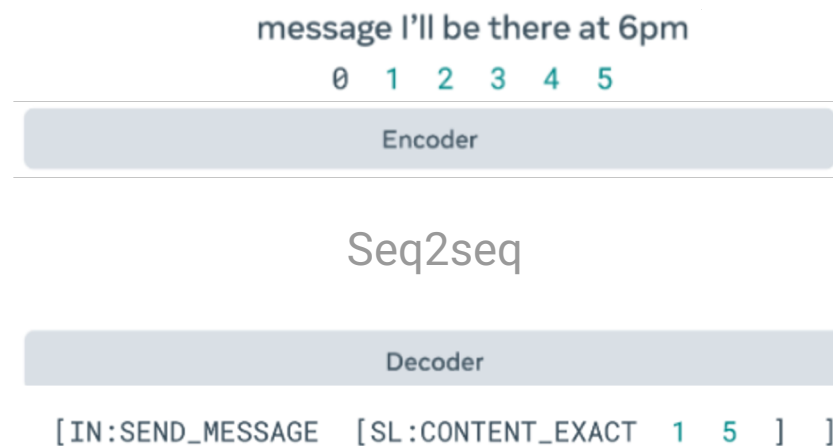
Direction 1. On-Device Machine Learning

Non-Autoregressive Semantic Parsing: Parallel prediction

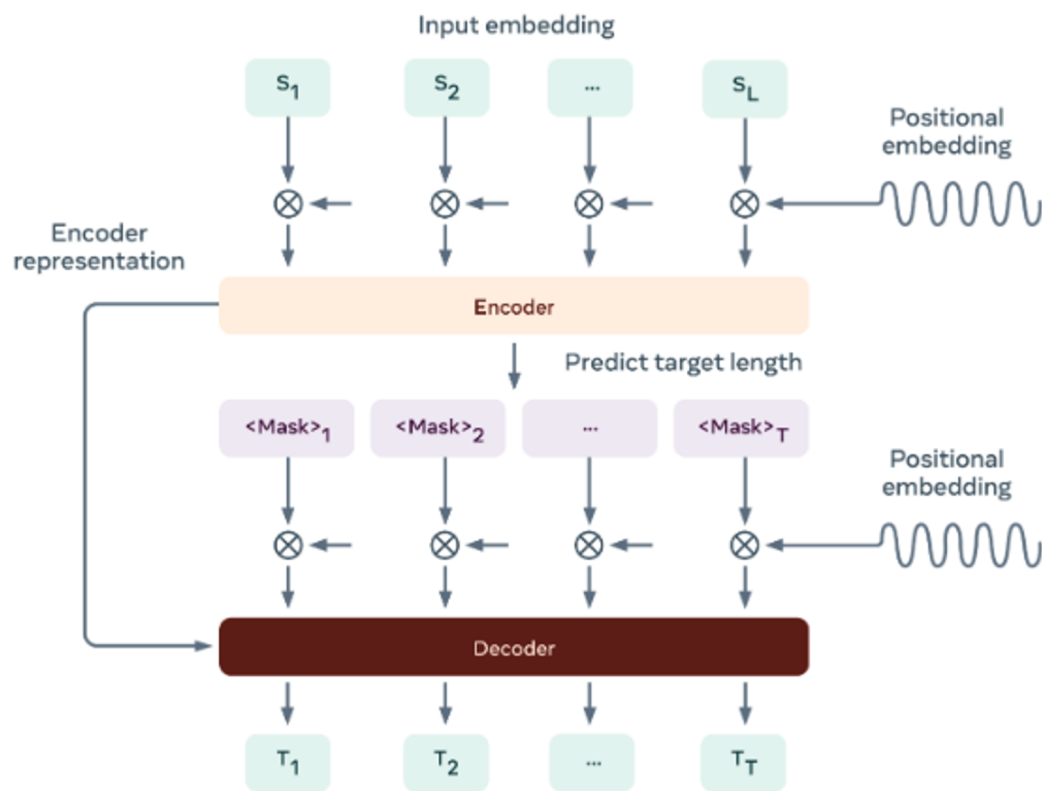


Direction 1. On-Device Machine Learning

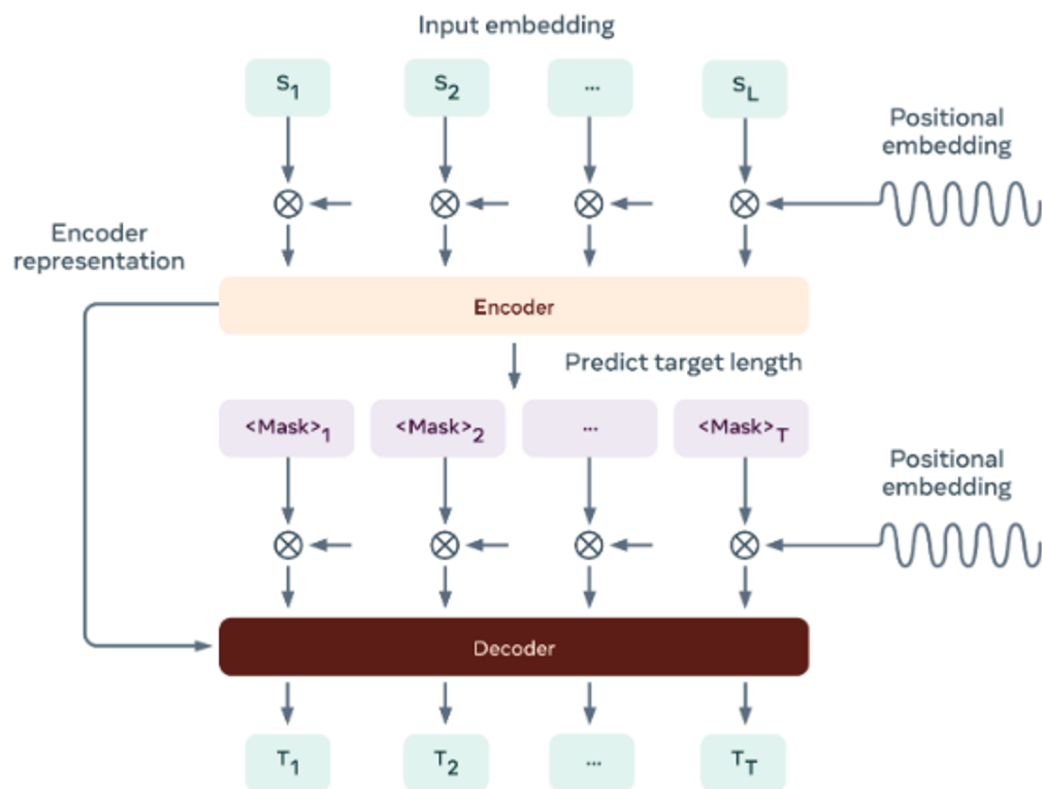
Non-Autoregressive Semantic Parsing: Parallel prediction



Direction 1. On-Device Machine Learning

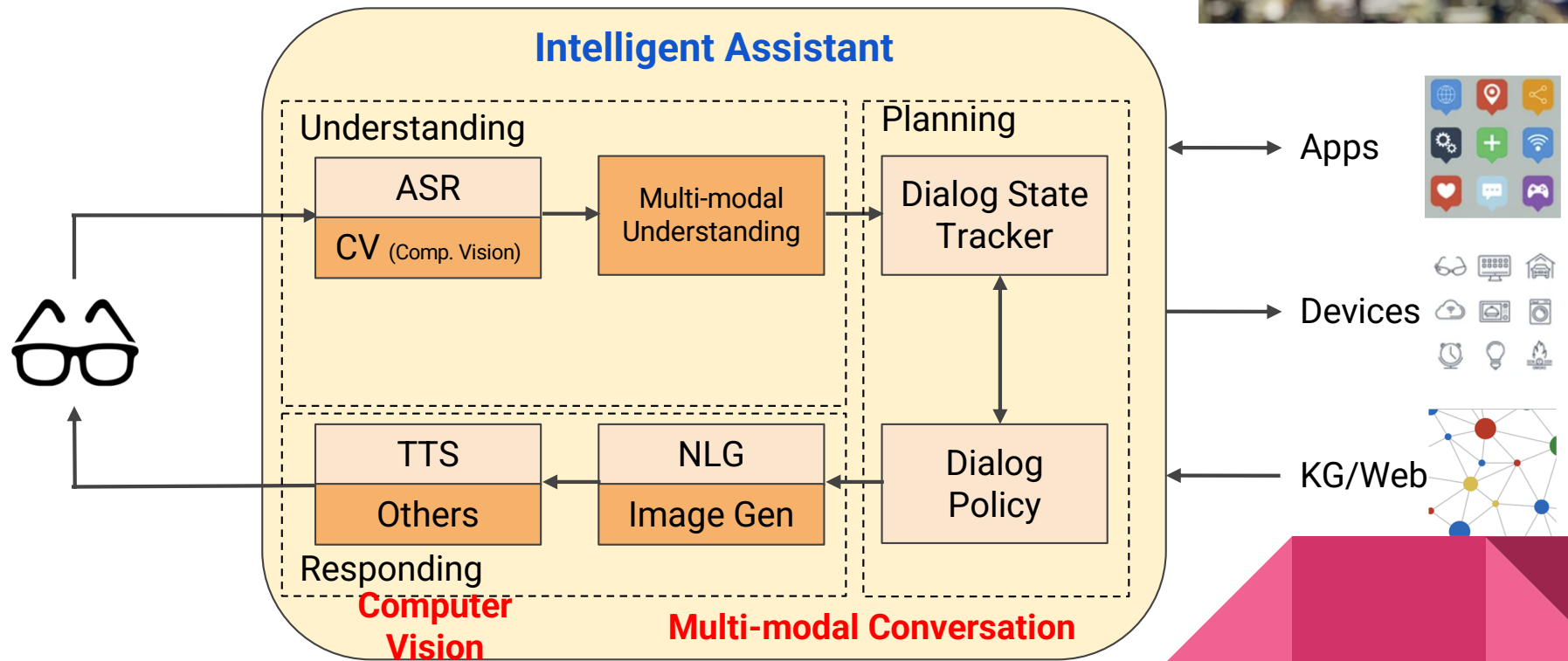


Direction 1. On-Device Machine Learning

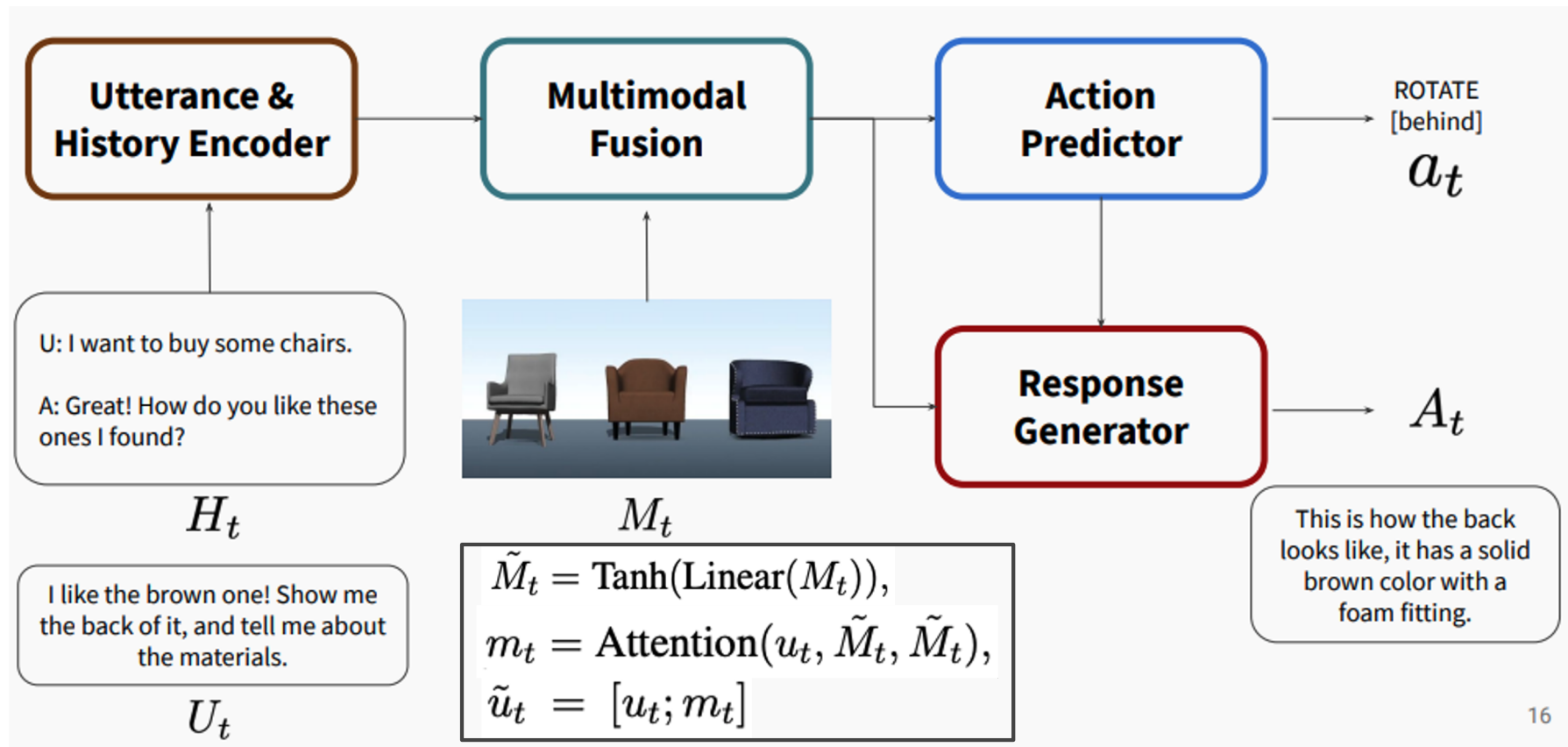


- Memory usage: -83%
- Latency: -70%
- Quality: +1.2% vs. non-AutoRegressive STOA
- Cross-lingual: +14% vs. AutoRegressive baseline

Direction 2. Multi-Modal Assistant



Direction 2. Multi-Modal Assistant



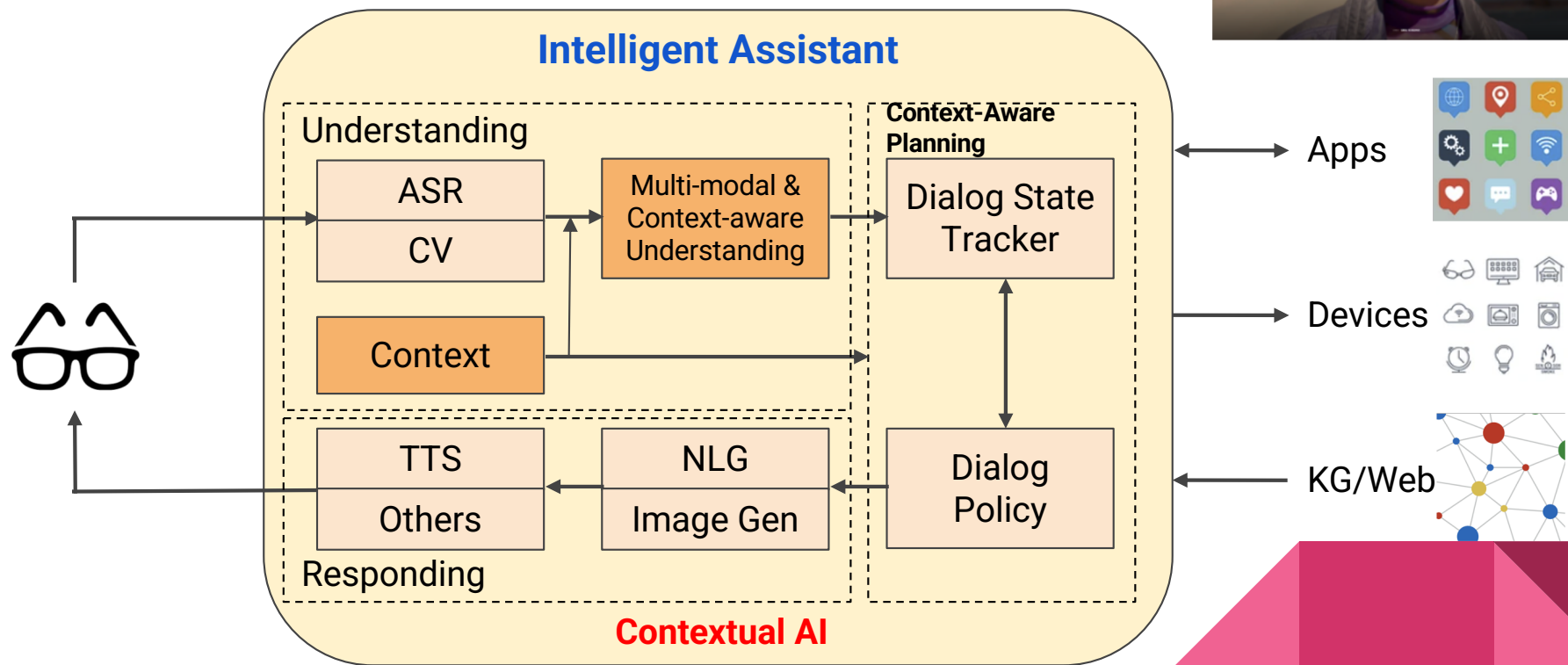
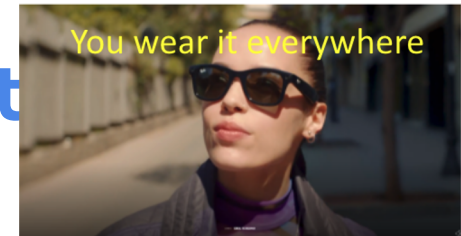
Direction 2. Multi-Modal Assistant

Model	T3. DST	
	In.F1↑	Sl.F1↑
SIMMC-Furniture		
TRADE	-	45.5
SimpleTOD	75.0	50.1
SimpleTOD+MM	74.1	60.2
SIMMC-Fashion		
TRADE	-	32.8
SimpleTOD	56.5	37.3
SimpleTOD+MM	59.1	43.5

The MultiModal model improves intent classification and slot filling.

Table 5: Results for: **(3) Dialog State Tracking (DST)**, measured with Intent and Slot prediction F1 metrics. ↑: higher is better, ↓: lower is better. Bold denotes the best for each metric.

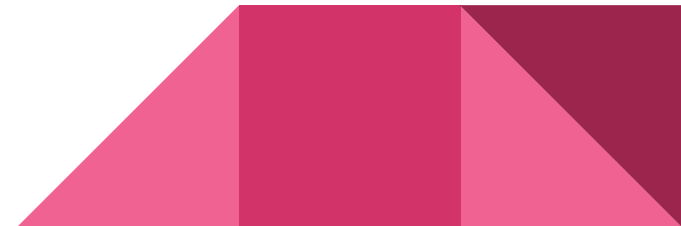
Direction 3. Context-Aware Assistant



Direction 3. Context-Aware Assistant



- Context-aware assistants
 - examine your surroundings, and
 - use this context to personalize a product experience.



Direction 3. Context-Aware Assistant



- Context-aware assistants
 - examine your surroundings, and
 - use this context to personalize services

Context

(Time, Location, Scene,
Activity, Event, etc.)

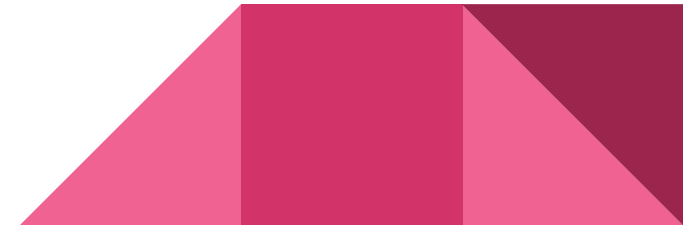
When is it?

Where are you?

What are you doing?

Whom are you together with?

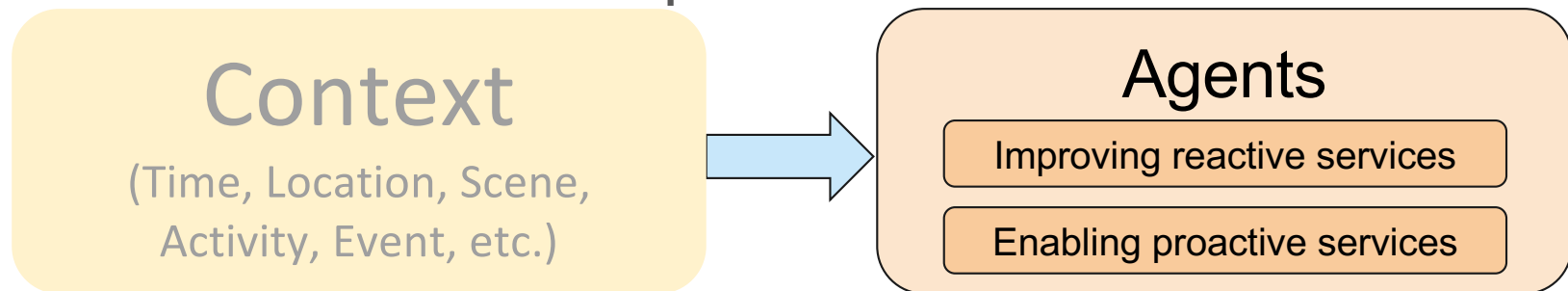
What's surrounding you?



Direction 3. Context-Aware Assistant



- Context-aware assistants
 - examine your surroundings, and
 - use this context to personalize services



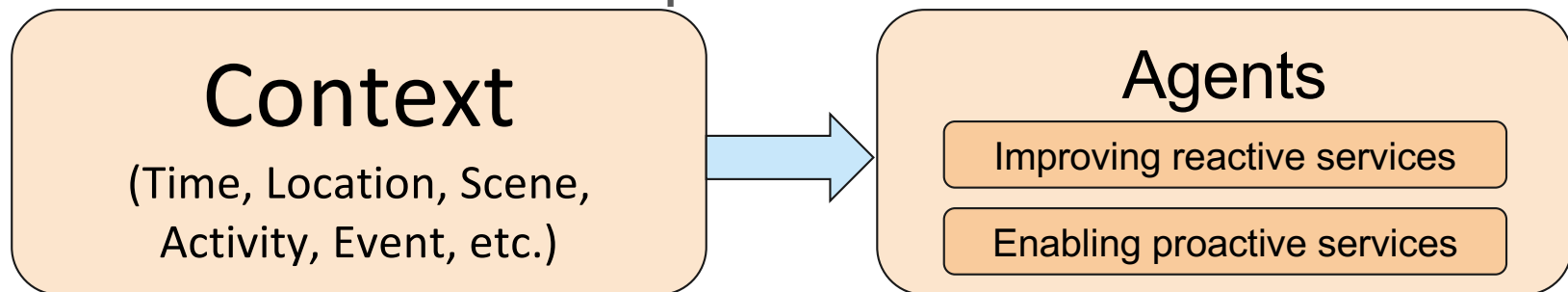
When is it?
Where are you?
What are you doing?
Whom are you together with?
What's surrounding you?

Context-aware ranking
Contextual recommendation
Contextual reminder, etc.

Direction 3. Context-Aware Assistant



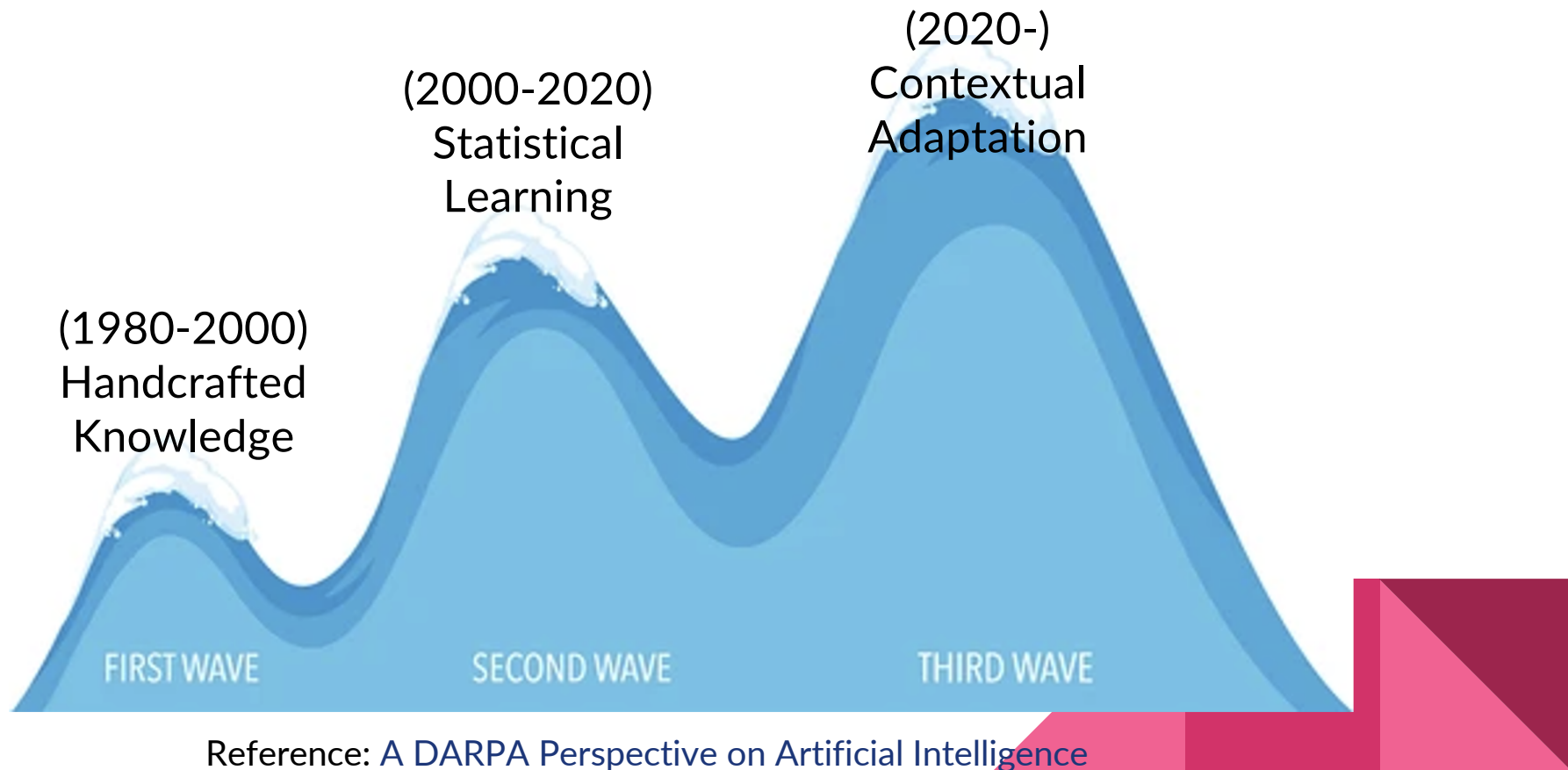
- Context-aware assistants
 - examine your surroundings, and
 - use this context to personalize services



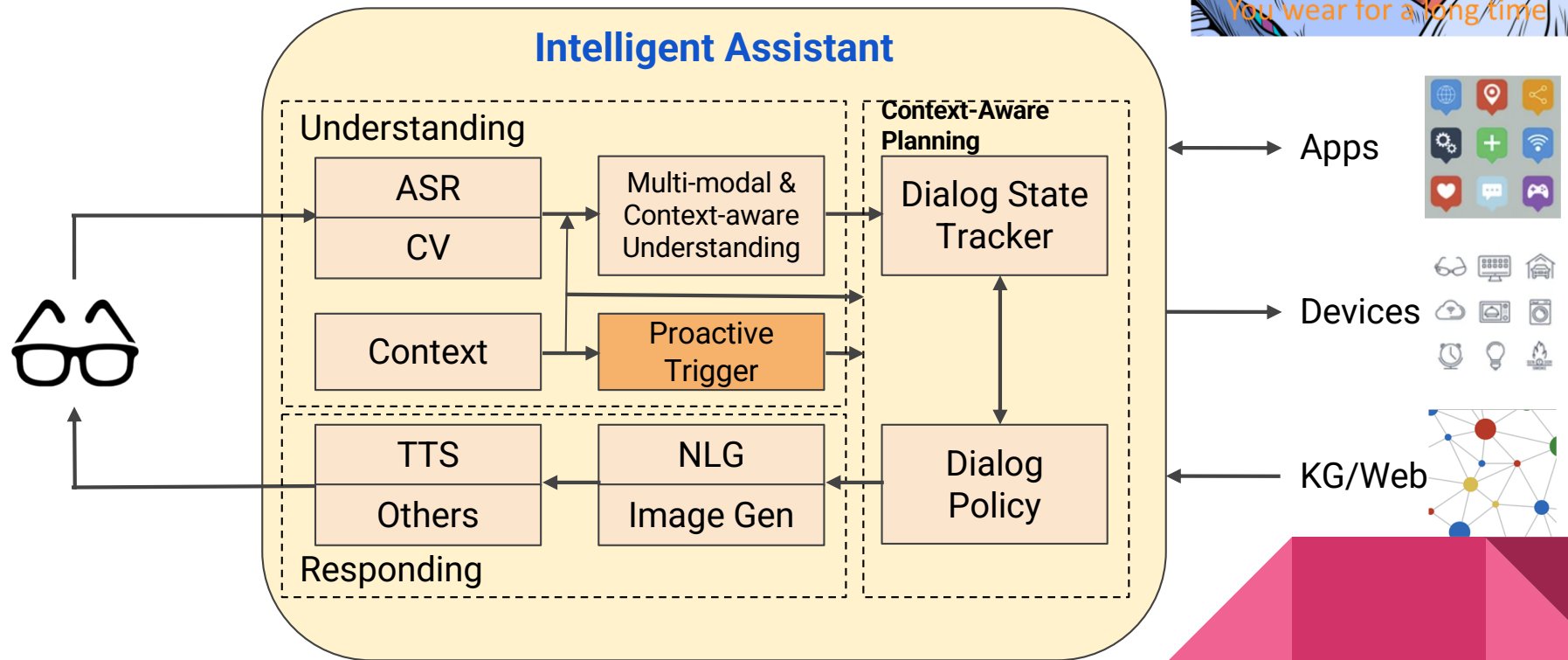
When is it?
Where are you?
What are you doing?
Whom are you together with?
What's surrounding you?

Context-aware ranking
Contextual recommendation
Contextual reminder, etc.

Direction 3. Context-Aware Assistant

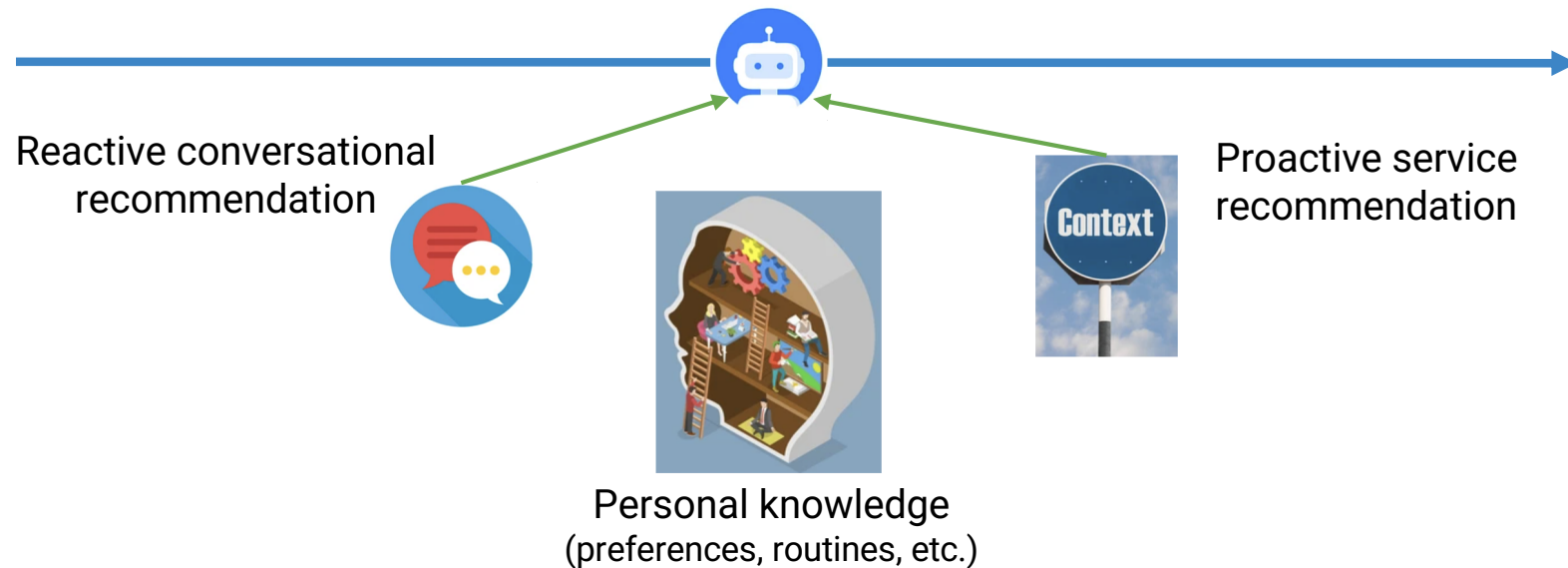


Direction 4. Assistant Recom.



Personalization, Federated Learning

Direction 4. Assistant Recommendation



Challenges: Leveraging personal action log, preferences, routines, etc., to improve context-aware recommendation w/o sacrificing privacy.

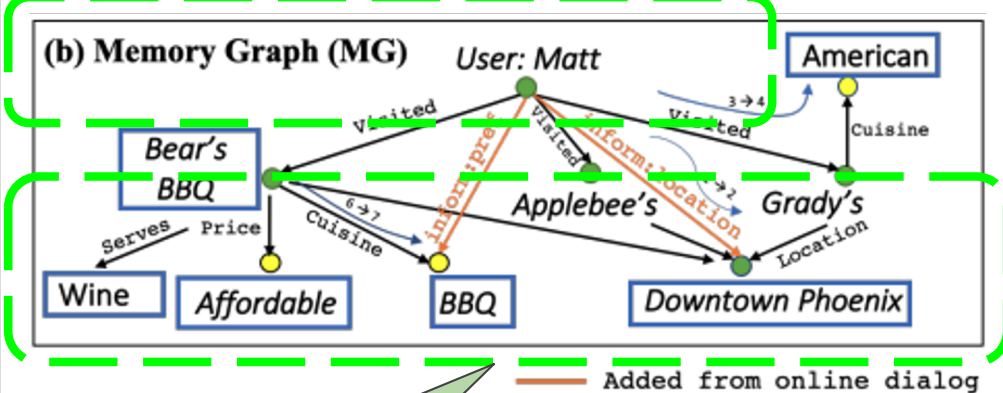
4-1. Conversational Recommendation

(a) Conversational Recommendations

- 1 Hello, I'm looking for a good place to eat.
- 2 Is Downtown Phoenix a good place to start your search as you've been there a few times?
- 3 Yes, I'd like something there please.
- 4 You ordered *American* food a few times, do you want something similar or feeling adventurous today?
- 6 No, I'm in the mood for a *BBQ* today.
- 7 *Bear's BBQ* got some *affordable* yet great *BBQ*. They also serve *wine*.

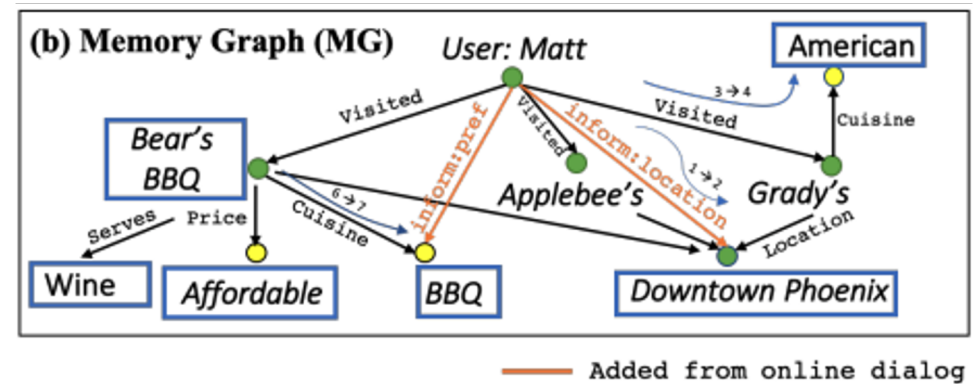
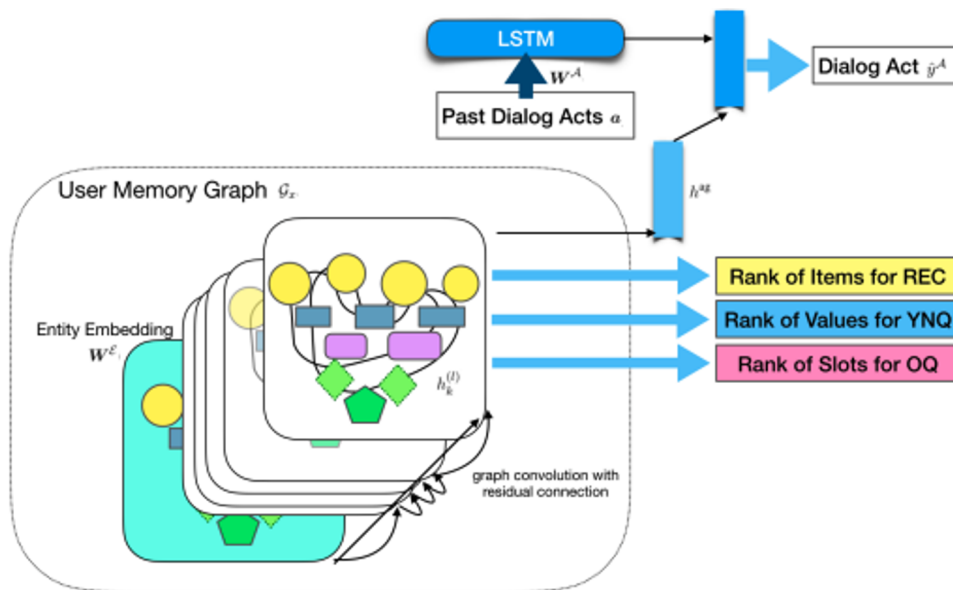
Personal KG

(b) Memory Graph (MG)



Public KG

4-1. Conversational Recommendation



4-1. Conversational Recommendation

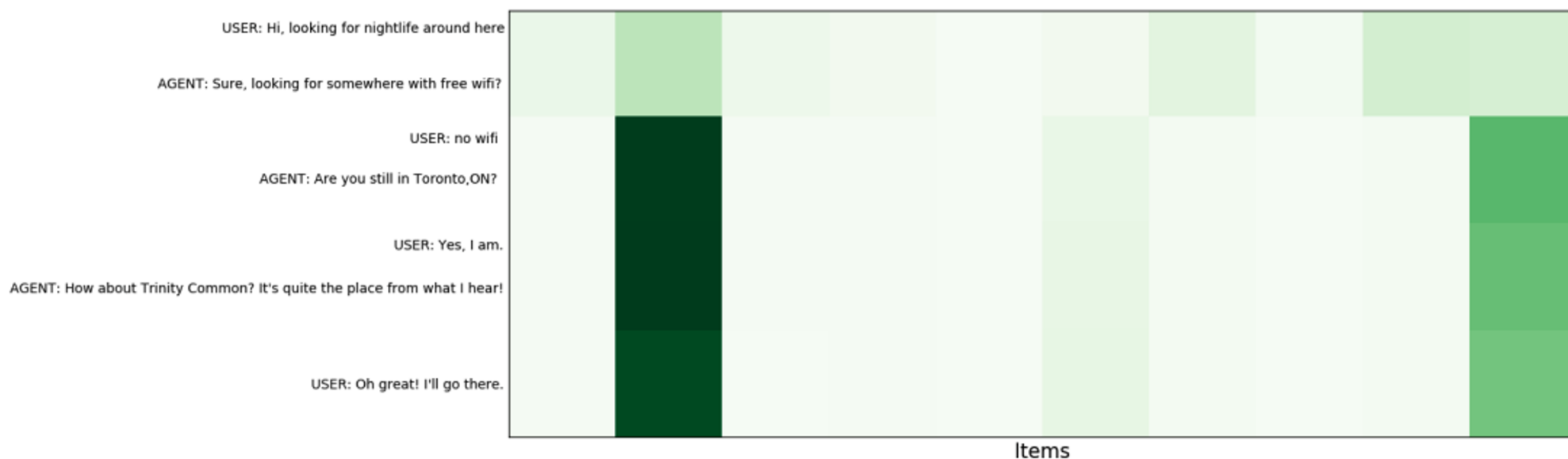


Figure 5: Visualization of item-level conversational reasoning, given an example dialog. Darker color indicates more salient items for recommendation at each given turn (row), predicted by our UMGR model.

4-2. Federated Learning

- ~~Push data to model~~ → Push models to data

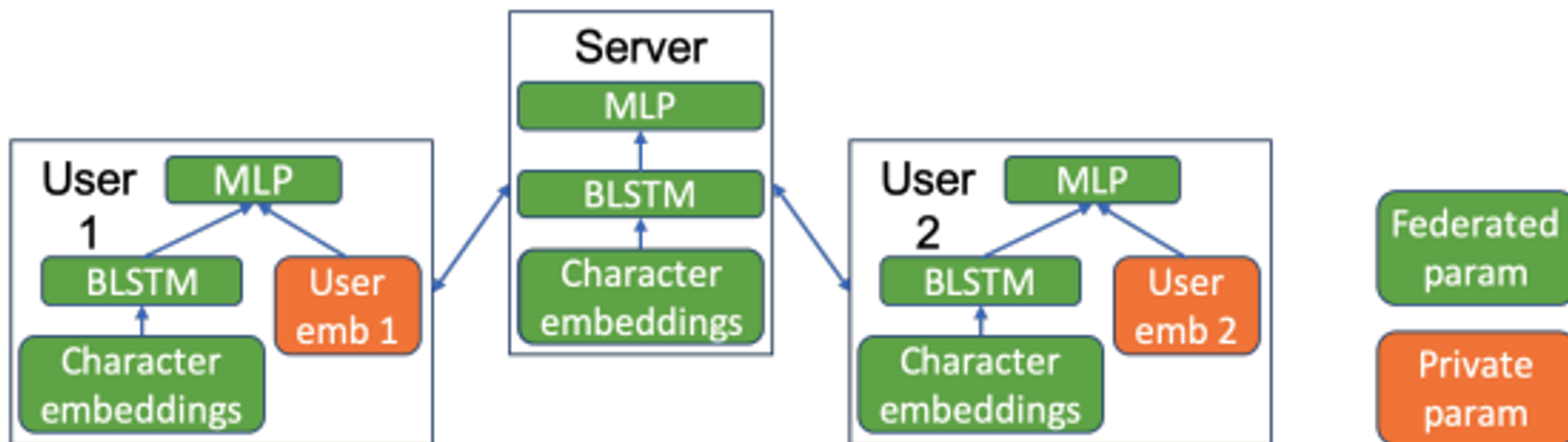
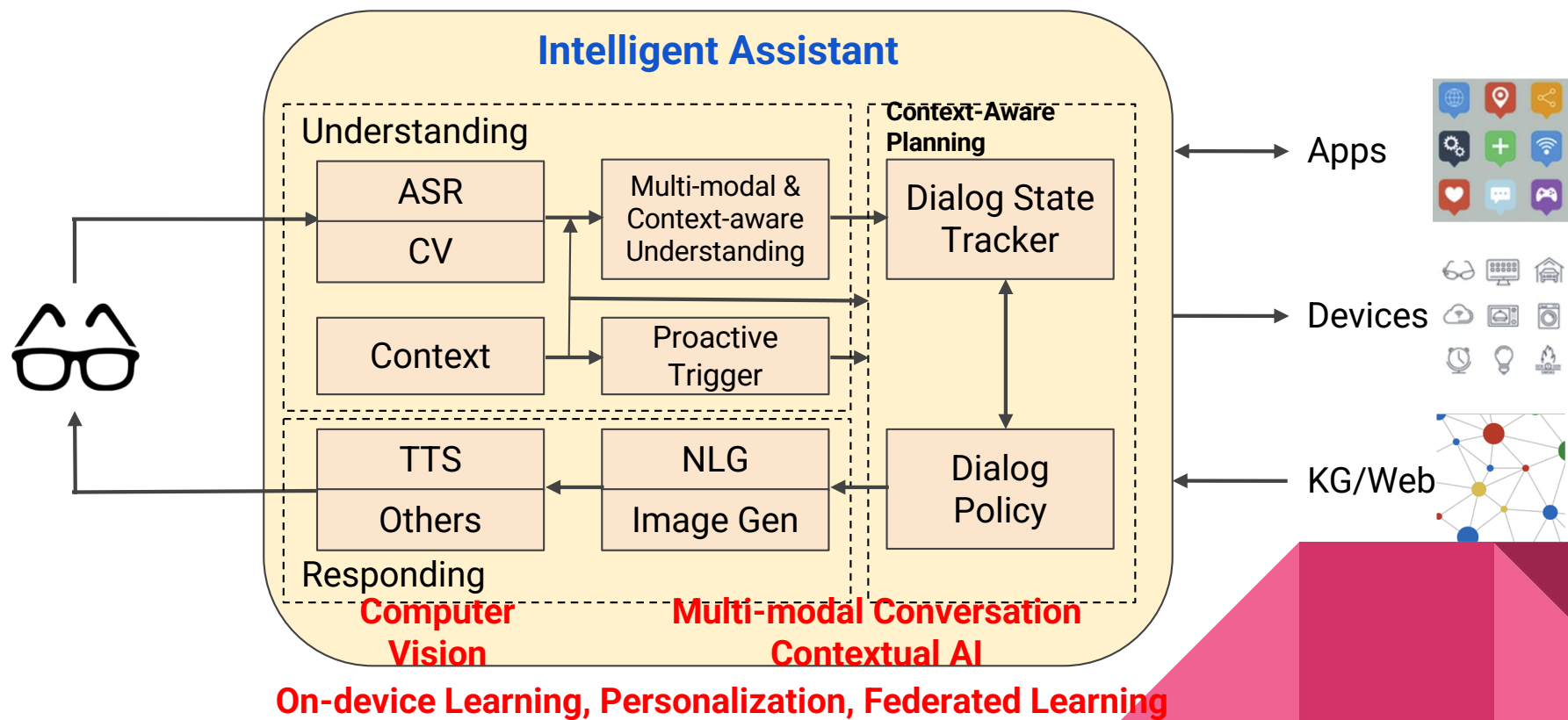
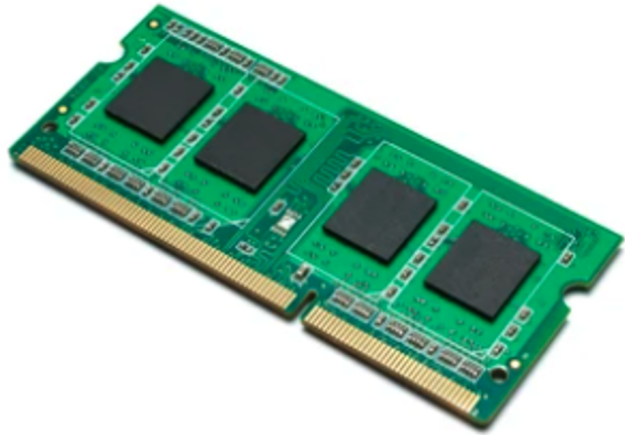


Figure 1: Personalized Document Model in FL.

Recap: New Architecture & Research Areas



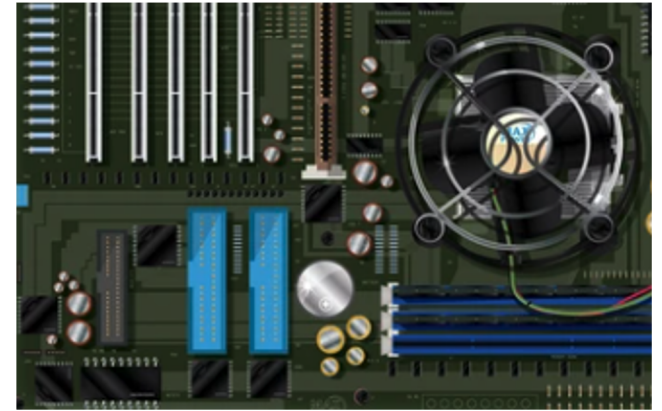
My Everyday Worries When Working on Devices



Memory



Battery



Thermal



I Had A Dream (2002-2021)



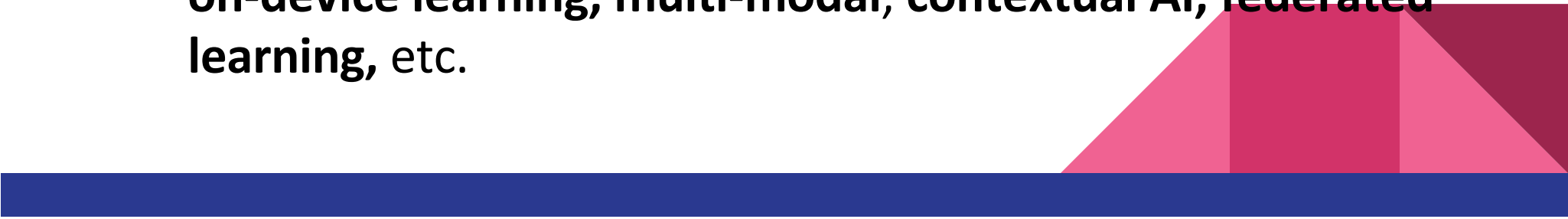
Trucks of *Data*, to enrich KGs

I Have A Dream (Now)



A Big Nose, to Wear A Computer

Take-Aways

- An intelligent assistant should be an agent that *knows you and the world*, can *receive your requests* or *predict your needs*, and provide you *the right services at the right time* with your permission
 - An intelligent assistant is essentially a **conversation system**, *task-driven or information-driven*
 - Next-generation AR/VR assistants require new research on **on-device learning, multi-modal, contextual AI, federated learning**, etc.
- 



Thank You

Q&A?