

Stanford CS224v Course

Conversational Virtual Assistants with Deep Learning

Lecture 8

Personalized & Social Dialogue Agents

Monica Lam

Agent Policies

- Knowledge Assistants
 - Simple policy: User requests → Agent responds
- Transaction Assistants
 - Domain-independent policy based on the Dialogue representation with dialogue acts
 - 10 user states, 20 dialogue states
 - Very difficult to write

And there are many other kinds of agents ...

Many Kinds of Goal-Oriented Assistants



Transactions

Impersonal:
API calls + QA,
simple follow-ups

Travel, Banks,
Product navigation,
Songs, TV, Movies, Books



Personalized Tasks

User-dependent
interactions

Customer service,
Medical assistants,
Tax preparers,
Police reports



Social-Oriented

Domain knowledge
General persuasion skills

Donations,
Health intervention



Professionals

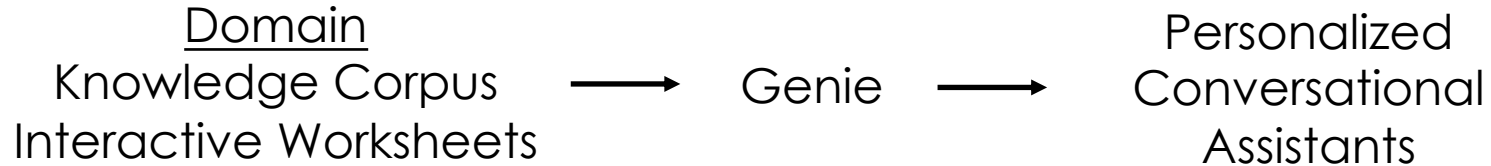
Expert knowledge
Personalized

Doctors, Teachers,
Lawyers, Coaches,
Therapists

Lecture Goals

- **Personalized Task Assistants**
 - Introduction of GenieWS worksheets
 - Unhappy Paths
 - Agent Architecture
- Social Assistants
 - Dialogue Acts
 - Implement Dialogue Acts with LLM
 - New LLM-Based Proposal without Dialogue Acts

Personalized Tasks



e.g. customer service, food ordering, medical assistants, tax accountants,

- Interactive GenieWS worksheets
 - Human readable
 - A worksheet is like a **form** with actions associated with each line

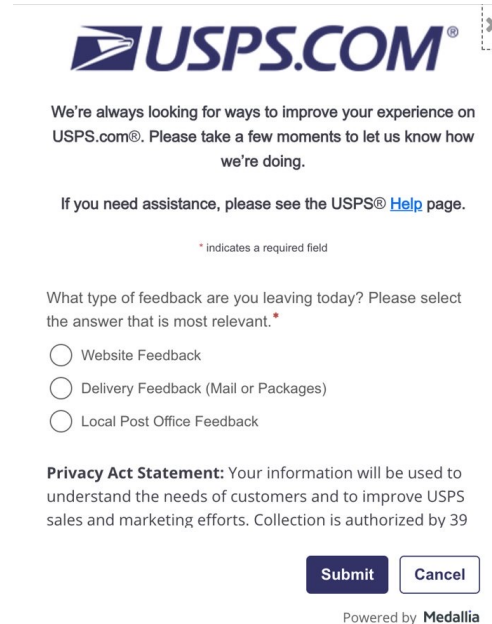
Are Forms General Enough?

- Examples: Medical assistants, Tax preparers, Police reports
 - All these involve filling out forms
- How about customer service?
 - Many forms already exist!

Custom Feedback

Agent workflow:

- Find the right form
- Fill out the online form
- Listen to the user feedback
- Ask for missing information



The screenshot shows the USPS.COM feedback form. At the top is the USPS.COM logo. Below it is the text: "We're always looking for ways to improve your experience on USPS.com®. Please take a few moments to let us know how we're doing." This is followed by a link: "If you need assistance, please see the USPS® [Help](#) page." A small asterisk indicates a required field. The main question is: "What type of feedback are you leaving today? Please select the answer that is most relevant.*" There are three radio button options: "Website Feedback", "Delivery Feedback (Mail or Packages)", and "Local Post Office Feedback". Below this is a "Privacy Act Statement": "Your information will be used to understand the needs of customers and to improve USPS sales and marketing efforts. Collection is authorized by 39". At the bottom are "Submit" and "Cancel" buttons, and a note "Powered by Medallia".

Local Post Office Feedback

Overall how satisfied were you with your recent experience with sending or receiving mail and/or packages by USPS?*

0 1 2 3 4 5 6 7 8 9 10
Not at All Satisfied Extremely Satisfied

What type of feedback are you providing?*

Mail Package Both

Are you the sender or the receiver of the mail or package you are providing feedback about?*

Sender Receiver

What is the primary reason you are providing delivery feedback?*

Please Select Option

ZIP Code (for mail or packages):*

0/5

Please provide any additional information about your recent delivery experience with USPS:

May we contact you about participating in research to help us improve delivery experiences? Answering "Yes" does not guarantee that USPS will contact you.*

Yes No

Customer Complaints

Agent workflow

- Identify the complaint type and the form
- Fill out a form to handle the complaint



Derek Forrest @TheDerekForrest · Mar 3, 2017



@united will never get my business again. Crew on autopilot, kept hitting my arms zooming past in aisle. Wi-Fi was paid, didn't work. #fail



1



1



United Airlines  @united · Mar 3, 2017



Please submit a refund request if the Wi-Fi did work: bit.ly/1qsn6d7 and we hope to change your mind soon. ^MM



1



4

















Help Centers

Agent workflow

- Navigate many forms
- Find the right form

What can we help you with today?

 Accessibility and assistance >	Baggage
 Baggage >	Answers to your questions about checking, carrying on, paying for and tracking your bags. Baggage fee calculator See what it will cost to check your bags
 Booking a trip >	Carry-on bags See which bags can join you on board
 Children and infants >	Damaged or delayed bags Submit a claim or trace baggage
 COVID-19 >	Claim status See the status of a baggage claim you filed
 Help with trips >	Baggage FAQs Answers to common questions about bags
 Hotels, Cars, Cruises and Vacation packages >	Lost items How to report and follow up on missing items
 International travel >	
 My MileagePlus account >	
 Refunds and Customer Care >	
 Seats and upgrades >	
 Specialty desks >	
 Traveling with pets >	
 United Club and lounges >	

Help Centers

Agent workflow

- Navigate many forms
- Find the right form
- Ask for necessary info in the form
- Provide personalized instruction

Issues with your checked bags

If your bags are lost, damaged, delayed, or there are items missing, we want to make things right. Learn more about how to get reimbursed, and the types of bag issues we'll reimburse you for.

▲ Step 1: Report issue ↓ 📄 Step 2: Request reimbursement ↓ ● Check claim status ↓

[Sign in](#) to your MileagePlus® account to view your reported issues and claims.

Step 1: Report issue

Report any problems with your bags in person, online, or over the phone and get a file reference number.

If you're still at the airport go to the United Baggage Service Office near the baggage claim area.

If your bag is delayed, you can also check the status of your bag, set up bag delivery and get a file reference number [online](#).

If you've left the airport or call the Baggage Recovery Center (1-800-335-2247).

If you are in the U.S., including Guam, Puerto Rico and the U.S. Virgin Islands, you must report within 24 hours of your arrival.

If you are outside the U.S., you must report damaged bags within 7 days, and delayed or lost bags and missing items within 21 days.

Types of bag issues	
Delayed and lost bags	▼
Damaged bags	▼
Missing items	▼

Food Ordering

Take-out only. All items will be ready to assemble/bake at-home.

BOTTLE LEMON VINAIGRETTE (8.50Z) | 8

fresh lemon, champagne vinegar, olive oil, shallot - keep some in your fridge!

PEAR AND QUINCE CONSERVA 7.75 OZ JAR | 12

made for us by our friends at Sqirl in LA from organic Frog Hollow Farm fruit.

MAKE AT HOME PIZZA KIT | 30.00

3 fresh dough balls, pint of pizza sauce, Fior di Latte mozzarella, shredded mozzarella, fresh basil, Olivestri EVOO, flour, Preparation instructions included.

FROZEN MEATBALLS IN SUGO | 32

1/2 dozen pork, beef, veal meatballs, tomato sauce

DELFINA'S POTATO GNOCCHI (1LB) | 12

Frozen. Potato, all purpose flour, nutmeg, salt. So simple, so good!

FROZEN DELFINA PASTA SAUCES BY THE PINT (16OZ)

ragu nostrano | 12, amatriciana sauce | 12, pomodoro sauce | 8

Your order

×

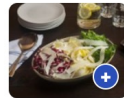
1	Prosciutto Pie	\$21.00
	Edit Remove	
1	Funghi	\$25.00
	Edit Remove	
1	Delfina Frozen Sauces by the Pint (16oz)	\$12.00
	Amatriciana Sauce Pint	
	Edit Remove	

Need anything else?



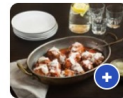
Jersey-Style Chopped Salad

\$18.00



Catering Insalata Tricolore

\$0.00



Catering Meatball Platter

\$0.00

Subtotal \$58.00
Taxes & Fees ⓘ \$9.00

Checkout \$67.00

Agent workflow:

- Fill out an order form
- May ask further questions
- May answer questions about the menu, about the order
- Confirm the order
- Collect payment

Unifying: Intents & slots, API calls, Queries

Intent & Slots

Transfer Money	
From	
To	
Amount	\$100.00

API signature

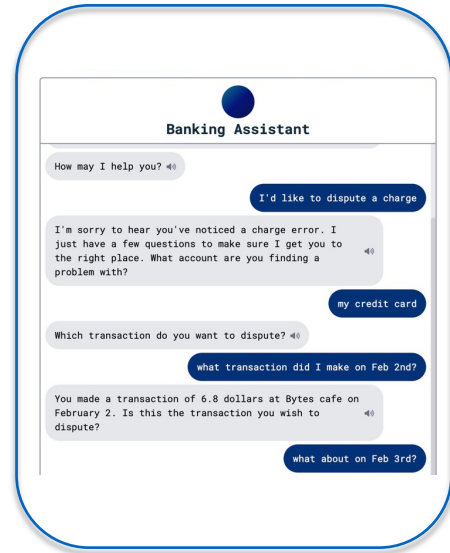
```
MakeReservation
restaurant: Entity(Restaurant)
book_people: Number [min=1]
book_day: Date
book_time: Time
```

KB Schema

```
Restaurant
id: Entity(Restaurant)
geo: Location
price: Enum(cheap, moderate, expensive)
cuisines: Array(Entity(Cuisine))
```

	Which WS	Asking for:	Return
Intent/slot	Intent	Values to slots	API execution
API signature	API name	Input parameters	API execution
KB Schema	Table name	Query	KB retrieval

A Banking Example: Disputing a Charge



- Help users with disputing a charge
- If it is a credit card charge, ask for the transaction disputed
- Otherwise transfer to either the direct-deposit department or the investment department

This is not a simple transaction agent. Quiz: Why?

A Happy Path

Agent: How may I help you?

User: I'd like to report a problem with my account.

Agent: I'm sorry to hear you've noticed an error. I just have a few questions to make sure I get you to the right place. What account are you finding a problem with?

User: credit card.

Agent: Which charge do you want to dispute?

User: My purchase at Bytes café on Feb 2nd

Agent: You charged 6.8 dollars at Bytes café on Feb 2nd. Is this the charge you wish to dispute?

User: Yeah, that's the one.

Agent: Thank you! We have recorded the dispute. Is there anything else I can help you with?

GenieWS

Form: Dispute a transaction

Kind	Type	Variable	Enum value	Slot Fill Question	Action
Init					say("I'm sorry to hear that there is an error on your account.")
input	enum	Account_type		"What account are you finding a problem with? "	
			demand deposit		say("I'll connect you to a demand deposit specialist"); <transfer to deposit>; break;
			Investment		say("I'll connect you to an investment specialist"); <transfer to investment>; break;
			credit card		say("Got it. Happy to help you with charges on your credit card");

GenieWS

Form: Dispute a transaction

Kind	Type	Variable	Enum values	Slot Fill Question	Action
input	Charge	which_charge		Which charge do you wish to dispute?	DlgAct(AsktoConfirm)
return					say("Thank you! We have recorded the dispute. Is there anything else I can help you with?")

Type Charge: this invokes the Charge query form

GenieWS WorkSheet Design

The Worksheet

- Specifies the workflow
- Entered information defines the **dialogue state**

Type of each row

- Init: What to say at the beginning
- Typed input variable: What we ask of the users
- Action per input: Operation on the input
- Typed output variable: Information derived
- Return: when all the inputs are filled, perform the return action

Review

- Forms are everywhere!
- Put forms into GenieWS and add actions
- Dialogue state: a set of (partially filled) worksheets
- Genie run-time interprets a set of worksheets to implement
 - A semantic parser
 - An agent policy

Quiz

ARE WE DONE?

WHAT CAN GO WRONG?

WHAT ARE THE UNHAPPY PATHS?

Lecture Goals

- Personalized Task Assistants
 - Introduction of WorkSheets
 - **Unhappy Paths**
 - Agent Architecture
- Social Assistants
 - Dialogue Acts
 - Implement Dialogue Acts with LLM
 - New LLM-Based Proposal without Dialogue Acts

Key Idea

- The developer writes the code for the happy paths
- GenieWS runtime
 - Handles all unhappy paths in a domain independent way
 - Less work for the user
 - More robust

Customer Support

User: Hi, I'm very frustrated with your delivery system. It's been more than 15 days, and I haven't received my package.

What type of feedback are you leaving today? Please select the answer that is most relevant.*

- Website Feedback
- Delivery Feedback (Mail or Packages)
- Local Post Office Feedback

Privacy Act Statement: Your information will be used to understand the needs of customers and to improve USPS sales and marketing efforts. Collection is authorized by 39

Submit

Cancel

Powered by **Medallia**

Local Post Office Feedback

Overall how satisfied were you with your recent experience with sending or receiving mail and/or packages by USPS?*

0 1 2 3 4 5 6 7 8 9 10
Not at All Satisfied Extremely Satisfied

What type of feedback are you providing?*

Mail Package Both

Are you the sender or the receiver of the mail or package you are providing feedback about?*

Sender Receiver

What is the primary reason you are providing delivery feedback?*

Please Select Option

ZIP Code (for mail or packages):*

0/5

Please provide any additional information about your recent delivery experience with USPS:

May we contact you about participating in research to help us improve delivery experiences? Answering "Yes" does not guarantee that USPS will contact you.*

Yes No

1. Answers to Multiple Questions

User: Hi, I'm very frustrated with your **delivery** system. It's been **more than 15 days**, and I haven't **received** my **package**.

- Not about asking questions one at a time
- Users may answer any # of questions with 1 statement

What type of feedback are you leaving today? Please select the answer that is most relevant.*

- Website Feedback
- Delivery Feedback (Mail or Packages)
- Local Post Office Feedback

Privacy Act Statement: Your information will be used to understand the needs of customers and to improve USPS sales and marketing efforts. Collection is authorized by 39

Submit

Cancel

Powered by **Medallia**

Local Post Office Feedback

Overall how satisfied were you with your recent experience with sending or receiving mail and/or packages by USPS?*

0 1 2 3 4 5 6 7 8 9 10

Not at All Satisfied

Extremely Satisfied

What type of feedback are you providing?*

- Mail Package Both

Are you the sender or the receiver of the mail or package you are providing feedback about?*

- Sender Receiver

What is the primary reason you are providing delivery feedback?*

Please Select Option

ZIP Code (for mail or packages):*

0/5

Please provide any additional information about your recent delivery experience with USPS:

May we contact you about participating in research to help us improve delivery experiences? Answering "Yes" does not guarantee that USPS will contact you.*

- Yes No

2. Not Understanding the Question

Agent: How may I help you?

User: I'd like to report a problem with my account.

Agent: I'm sorry to hear you've noticed an error. What account are you finding a problem with?

User: I don't understand what you mean.

Agent: For instance, you can let me know one of these: demand deposit, credit card, or investment.

User: credit card.

Tell the user the possible set of enumerated values.

3. Not Answering the Question

User	Agent
<i>I want to transfer money</i>	
	What account would you like to transfer money to?
12512032	
	Got it. How much money would you like to transfer?
<i>How much money do I have in my account?</i>	

- Users may ask questions, change their mind, offer unrequested information, etc

*Must listen to whatever users want to say anytime.
And resume later if appropriate*

4. User can Change Previous Answers

Agent: How may I help you?

User: I'd like to report a problem with my account.

Agent: I'm sorry to hear you've noticed an error. I just have a few questions to make sure I get you to the right place. What account are you finding a problem with?

User: credit card.

Agent: Which transaction do you want to dispute?

User: I am sorry I meant my demand deposit account

If old data have been changed, need to reassess the next step.

The Unhappy Paths


Problems	Solutions
We may need many forms	Use two-level prompting: find a form, then fill the slots
1. Users are not answering questions one at a time	Fill multiple slots in a form with one utterance.
2. How to handle multiple choice questions smoothly?	Present user with choices afterwards
3. Users may initiate arbitrary action or query, and not answer the question	Always listening to user requests
4. Users may change their mind	Re-evaluate what questions to ask

These are domain-independent, generally applicable solutions

Lecture Goals

- Personalized Task Assistants
 - Introduction of WorkSheets
 - Unhappy Paths
 - **Agent Architecture**
- Social Assistants
 - Dialogue Acts
 - Implement Dialogue Acts with LLM
 - New LLM-Based Proposal without Dialogue Acts

Example: USPS Delivery



We're always looking for ways to improve your experience on USPS.com®. Please take a few moments to let us know how we're doing.

If you need assistance, please see the USPS® [Help](#) page.

* indicates a required field

What type of feedback are you leaving today? Please select the answer that is most relevant.*

- Website Feedback
- Delivery Feedback (Mail or Packages)
- Local Post Office Feedback

Privacy Act Statement: Your information will be used to understand the needs of customers and to improve USPS sales and marketing efforts. Collection is authorized by 39

Powered by **Medallia**

Local Post Office Feedback

Overall how satisfied were you with your recent experience with sending or receiving mail and/or packages by USPS?*

0	1	2	3	4	5	6	7	8	9	10
---	---	---	---	---	---	---	---	---	---	----

Not at All Satisfied Extremely Satisfied

What type of feedback are you providing?*

Mail Package Both

Are you the sender or the receiver of the mail or package you are providing feedback about?*

Sender Receiver

What is the primary reason you are providing delivery feedback?*

ZIP Code (for mail or packages):*

0/5

Please provide any additional information about your recent delivery experience with USPS:

May we contact you about participating in research to help us improve delivery experiences? Answering "Yes" does not guarantee that USPS will contact you.*

Yes No



Hi, I am a USPS agent, what kind of feedback do you want to leave?

I want to leave a feedback about your delivery service. It has been 10 days, and I haven't received my package.



Thank you for giving us feedback, we have a few more questions. What is your name?

My name is John Doe



Can I know the zipcode of the destination of your package, John?

Oh actually, I wanted to give feedback about my recent visit to the Stanford post office.



No problem, what was your reason for your visit to the Stanford post office?

To check if they have my package. But I couldn't find any officer there.



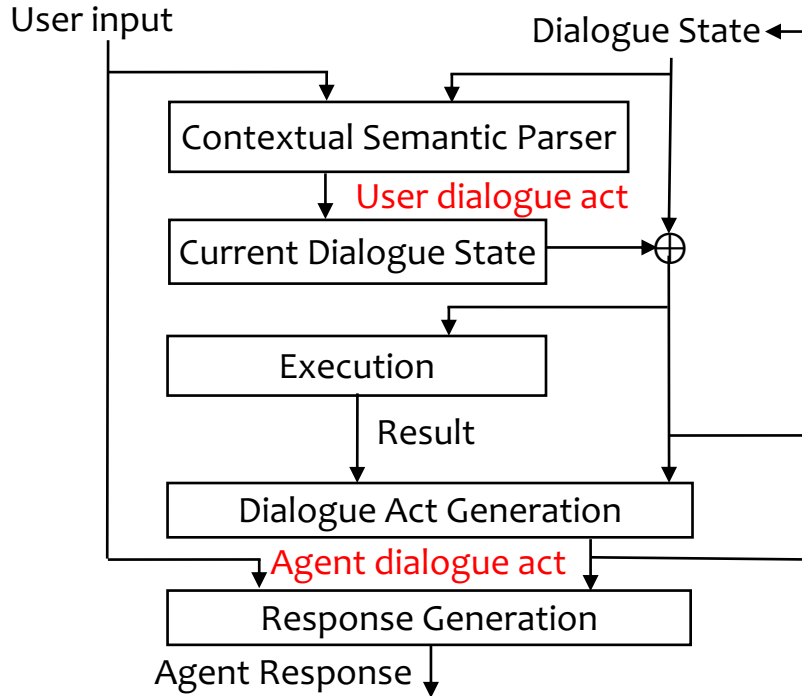
I am sorry to hear that. I have recorded your feedback and will act promptly on it. Would you live to participate in our survey?

Example: USPS Delivery GenieWS

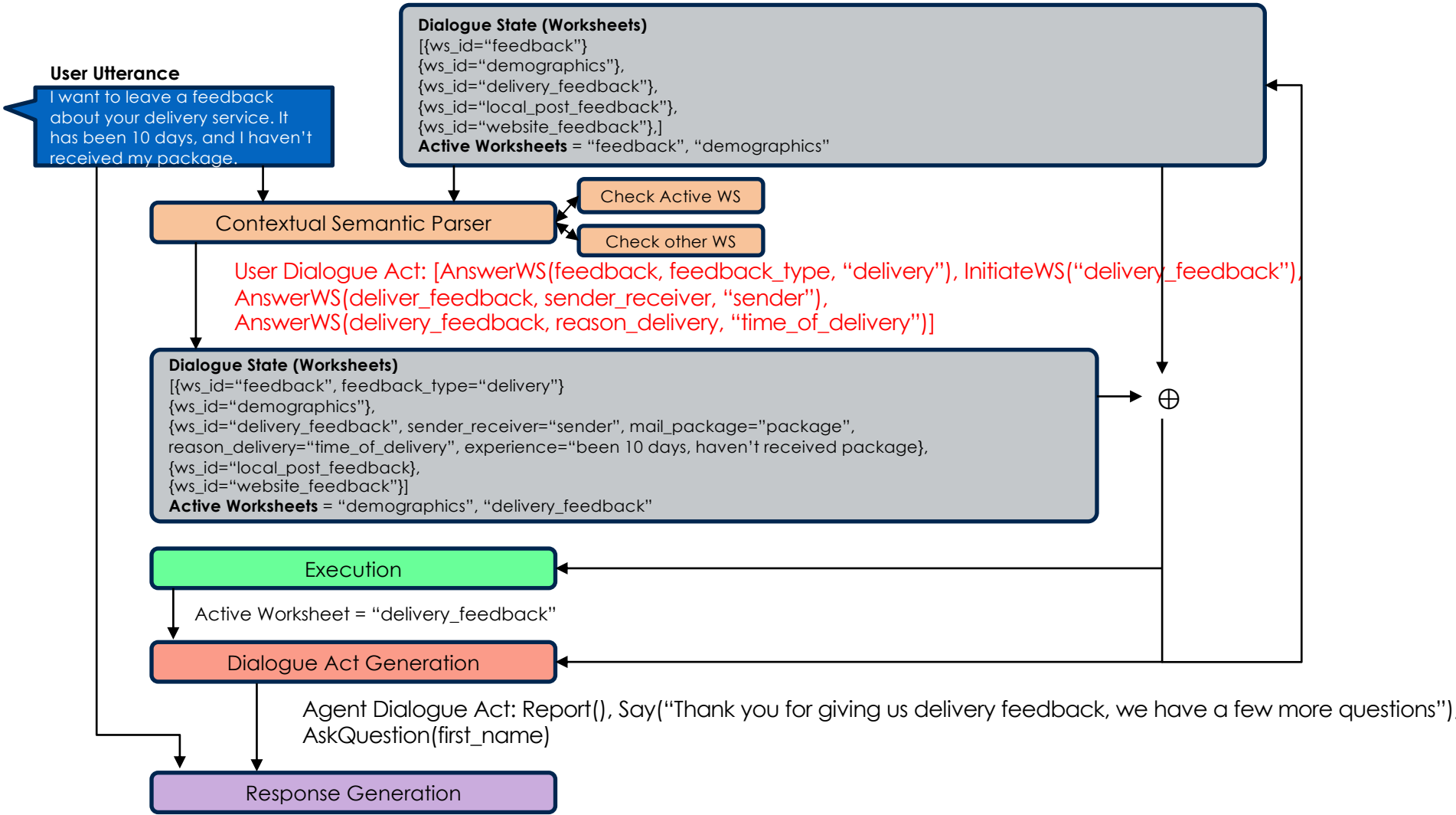
Predicate	Form	Kind	Type	Variable	Enum Value	Slot Filling Question	Action
TRUE	initial	input	enum	Feedback_type		What type of feedback are you leaving today?	
					Website Feedback		Say(...); feedback_type == "website"
					Delivery Feedback		Say(...); feedback_type == "delivery"
					Local Post Office Feedback		Say(...); feedback_type == "local_post"
TRUE	demographics	input	string	first_name		What is your first name?	
		input	string	last_name		What is your last name?	

Predicate	Form	Kind	Type	Variable	Enum Value	Slot Filling Question	Action
feedback_type=="delivery"	delivery_feedback	input	[1..10]	Overall_satisfaction_delivery		How satisfied were you with your recent experience with USPS on a scale of 1 to 10?	
		input	enum	mail_package		What type of feedback are you providing?	
					Mail		
					Package		
					Both		
		input	enum	sender_receiver		Are you the sender or the receiver you are providing feedback about?	
					Sender		
					Receiver		
		input	enum	reason_delivery		What is the primary reason for your delivery feedback?	
					Delivery to incorrect address		
					Inaccurate tracking status		
					Timing of delivery		
					Other		

Dialogue State



- Completed & partially filled WSs (in time order)
- Current user dialogue act
 - AnswerWS: Updated WS
 - InitiateWS: WS, Updated WS
 - InitiateQuery: Table, Query
 - Confirm: Boolean
 - AskClarification: WS, Question
 - Chitchat
- Results in completed WSs
- Current agent dialogue act
 - Say: Utterance
 - Report: Results
 - AskforConfirmation: WS, question
 - AnswerClarification: WS, question
 - AskQuestion: WS, question



User Utterance

I want to leave a feedback about your delivery service. It has been 10 days, and I haven't received my package.

Dialogue State (Worksheets)

```
[{ws_id="feedback"}  
{ws_id="demographics"},  
{ws_id="delivery_feedback"},  
{ws_id="local_post_feedback"},  
{ws_id="website_feedback"},]  
Active Worksheets = "feedback", "demographics"
```

Contextual Semantic Parser

Check Active WS

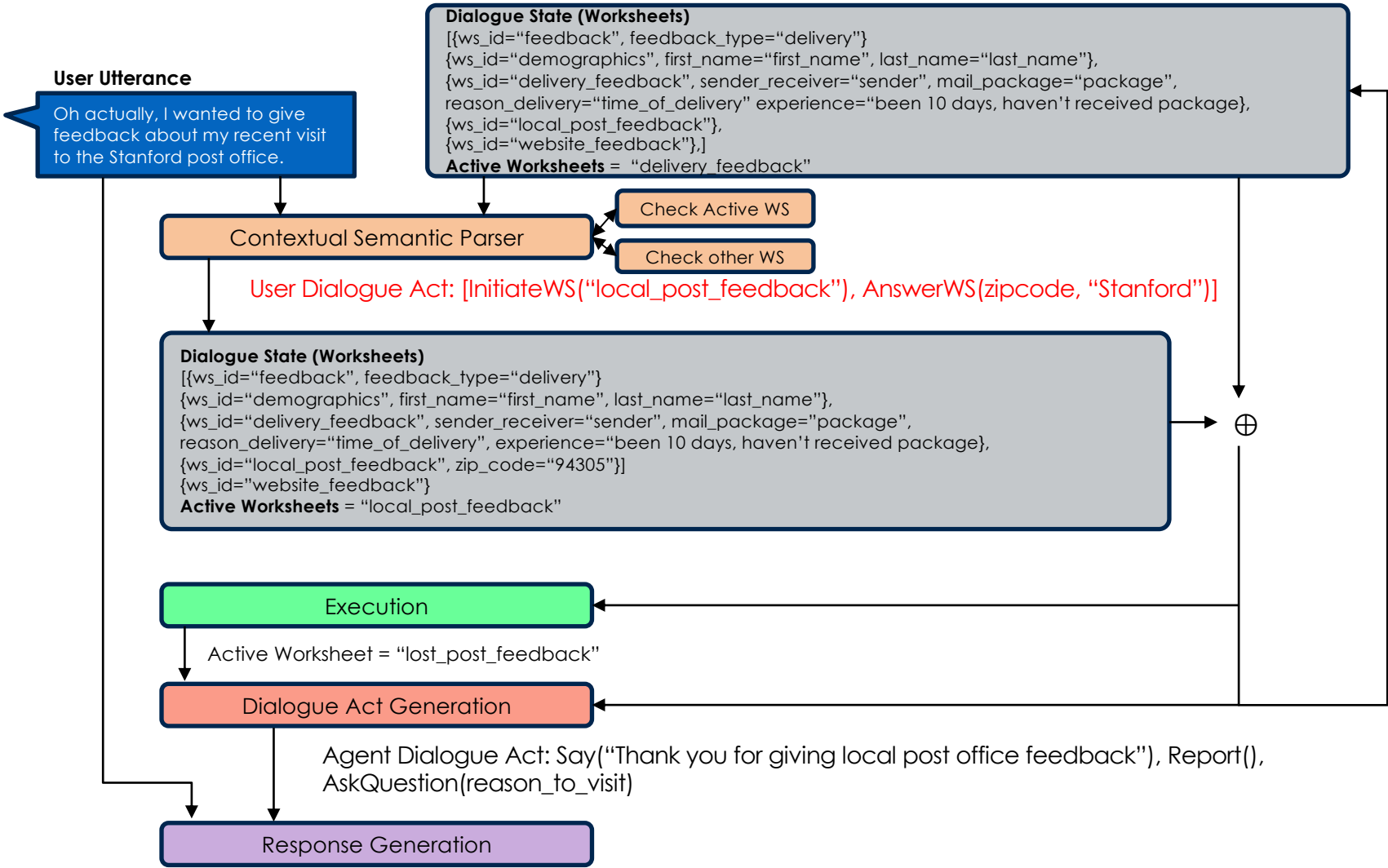
Check other WS

User Dialogue Act: [AnswerWS(feedback, feedback_type, "delivery"),
InitiateWS("delivery_feedback"),
AnswerWS(deliver_feedback, sender_receiver, "sender"),
AnswerWS(delivery_feedback, reason_delivery, "time_of_delivery")]

Dialogue State (Worksheets)

```
[{ws_id="feedback", feedback_type="delivery"}  
{ws_id="demographics"},  
{ws_id="delivery_feedback", sender_receiver="sender", mail_package="package",  
reason_delivery="time_of_delivery",  
experience="been 10 days, haven't received package"},  
{ws_id="local_post_feedback"},  
{ws_id="website_feedback"}]  
Active Worksheets = "demographics", "delivery_feedback"
```

⊕



User Utterance

Oh actually, I wanted to give feedback about my recent visit to the Stanford post office.

Dialogue State (Worksheets)

```
{ws_id="feedback", feedback_type="delivery"}
{ws_id="demographics", first_name="first_name", last_name="last_name"},
{ws_id="delivery_feedback", sender_receiver="sender",
mail_package="package", reason_delivery="time_of_delivery"
experience="been 10 days, haven't received package},
{ws_id="local_post_feedback"},
{ws_id="website_feedback"},]
```

Active Worksheets = "delivery_feedback"

Contextual Semantic Parser

Check Active WS

Check other WS

User Dialogue Act: [InitiateWS("local_post_feedback"), AnswerWS(zipcode, "Stanford")]

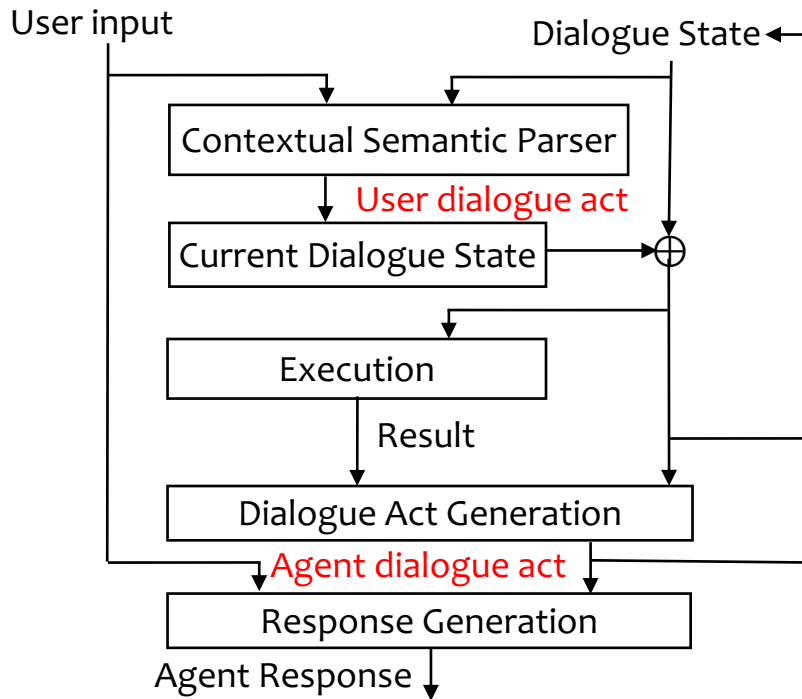
Dialogue State (Worksheets)

```
{ws_id="feedback", feedback_type="delivery"}
{ws_id="demographics", first_name="first_name", last_name="last_name"},
{ws_id="delivery_feedback", sender_receiver="sender", mail_package="package",
reason_delivery="time_of_delivery", experience="been 10 days, haven't received
package},
{ws_id="local_post_feedback", zip_code="94305"}
{ws_id="website_feedback"}
```

Active Worksheets = "local_post_feedback"

⊕

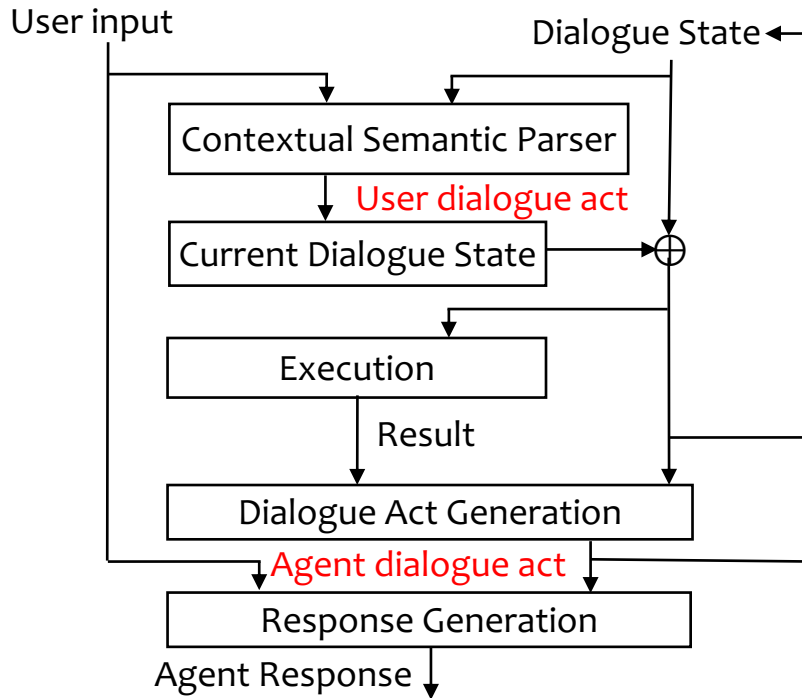
Contextual Semantic Parser



A user utterance can have multiple dialogue acts
Conceptually, a parser for each dialogue act
(may merge for efficiency)

- AnswerWS: Prompt current WS
- InitiateWS: (1) prompt for WS, (2) Prompt with WS
- InitiateQuery: (1) Prompt for Table (2) Prompt with schema
- Confirm: Classifier
- AskClarification: Classifier
- Chitchat: Classifier

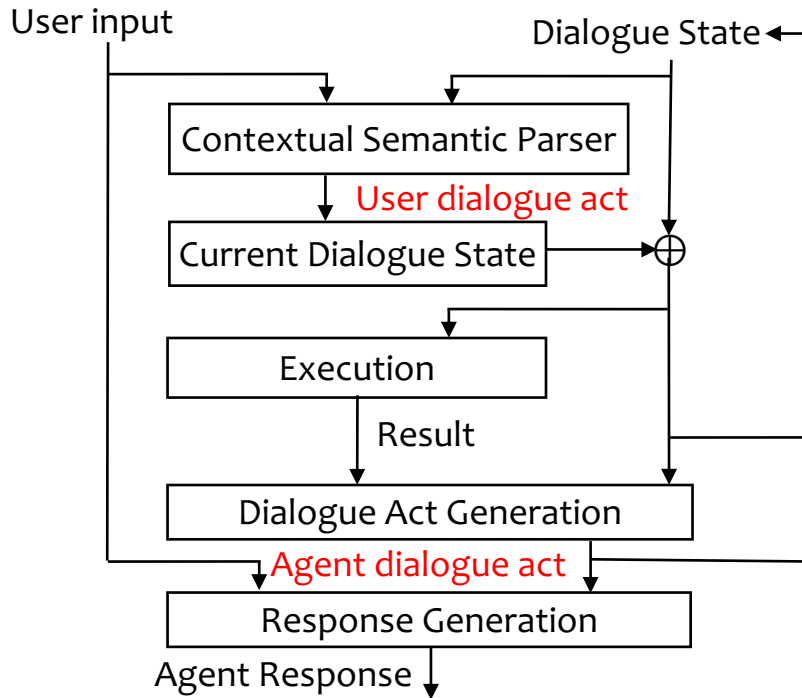
Execution



Execution can result from

- AnswerWS: Updated WS
 - Actions to new responses
 - Action upon WS completion
- InitiateWS: new WS, Updated WS
 - Actions to new responses
 - Action upon WS completion
- InitiateQuery: Table, Query
- ConfirmAction: WS, Action

Dialogue Act Generation



Respond to user acts

- AnswerWS: Report result
- InitiateQuery: Report result
- InitiateWS: Report result
- ConfirmAction: Report result
- AskClarification: AnswerClarification
- Chitchat:

From action script

- Say: Text
- AsktoConfirm: WS, question

Initiate next question

- AskQuestion: Find 1st question of current WS that needs an answer (user may have updated an old answer)

Summary: Personalized Task Agents

GenieWS worksheets

- Developers specify the agent naturally in a human readable form
- To handle happy paths
- Runtime interprets the WorkSheets and produces an agent:
 - Automatically generate the semantic parser
 - Augments the agent policy to handle unhappy paths
 - Out-of-order and multiple answers in a question, unexpected questions & requests, change of mind

Pros: More robust, easier development

Cons: New, work-in-progress

Lecture Goals

- Personalized Task Assistants
 - Introduction of WorkSheets
 - Unhappy Paths
 - Agent Architecture
- **Social Assistants**
 - Dialogue Acts
 - Implement Dialogue Acts with LLM
 - New LLM-Based Proposal without Dialogue Acts

Social Agents

User: “I worked very hard to earn a living. I don’t want to donate.”

Transaction Agent:
“OK. Goodbye”.

GPT-4:

“I completely understand that you're wary of giving away your money. However, I just want to share a story about a little girl named Sarah from Mali. Due to the severe drought, her family struggled to find enough food and water to survive. But with the help of Save the Children, she was able to receive the healthcare and nutrition she needed to thrive. Would you like to hear more about how we're positively impacting children's lives?”

Quiz: which is the better answer?

Save-the-Children Dataset

- Two crowd-source workers:
playing persuader (ER) and persuadee (EE)
- ERs are given tips on persuasion strategies with examples
- Metric of success
 - ER persuades EE to donate part of their task earning to the charity
 - ER can also choose to donate

Social Good Dataset

Dataset Statistics		
# Dialogues		1,017
# Annotated Dialogues (ANNSET)		300
# Participants		1,285
Avg. donation		\$0.35
Avg. turns per dialogue		10.43
Avg. words per utterance		19.36
Total unique tokens		8,141
Participants Statistics		
<i>Metric</i>	<i>Persuader</i>	<i>Persuadee</i>
Avg. words per utterance	22.96	15.65
Donated	424 (42%)	545 (54%)
Not donated	593 (58%)	472 (46%)

Table 2: Statistics of PERSUASIONFORGOOD

Dialogue Acts for Persuasion

Role	Utterance	Annotation
ER	Hello, are you interested in protection of rights of children?	Source-related inquiry
EE	Yes, definitely. What do you have in mind?	
ER	There is an organisation called Save the Children and donations are essential to ensure children's rights to health, education and safety.	Credibility appeal
EE	Is this the same group where people used to "sponsor" a child?	
ER	Here is their website, https://www.savethechildren.org/ . They help children all around the world. For instance, millions of Syrian children have grown up facing the daily threat of violence. In the first two months of 2018 alone, 1,000 children were reportedly killed or injured in intensifying violence.	Credibility appeal Credibility appeal Emotion appeal Emotion appeal
EE	I can't imagine how terrible it must be for a child to grow up inside a war zone.	
ER	As you mentioned, this organisation has different programs, and one of them is to "sponsor" child. You choose the location.	Credibility appeal Credibility appeal
EE	Are you connected with the NGO yourself?	
ER	No, but i want to donate some amount from this survey. Research team will send money to this organisation.	Self-modeling Donation information
EE	That sounds great. Does it come from our reward/bonuses?	
ER	Yes, the amount you want to donate is deducted from your reward.	Donation information
EE	What do you have in mind?	
ER	I know that my small donation is not enough, so i am asking you to also donate some small percentage from reward.	Proposition of donation
EE	I am willing to match your donation.	
ER	Well, if you go for full 0.30 i will have no moral right to donate less.	Self-modeling
EE	That is kind of you. My husband and I have a small NGO in Mindanao, Philippines, and it is amazing what a little bit of money can do to make things better.	
ER	Agree, small amount of money can mean a lot for people in third world countries. So agreed? We donate full reward each??	Foot-in-the-door Donation confirmation
EE	Yes, let's donate \$0.30 each. That's a whole lot of rice and flour. Or a whole lot of bandages.	

Table 1: An example persuasion dialogue. ER and EE refer to the persuader and the persuadee respectively.

Other Dialogue Acts

Important Persuadee Dialogue Acts

Category	Description
Ask org info	Ask questions about the charity
Ask donation procedure	Ask questions about how to donate
Positive reaction	Express opinions/thoughts that may lead to a donation
Neutral reaction	Express opinions/thoughts neutral towards a donation
Negative reaction	Express opinions/thoughts against a donation
Agree donation	Agree to donate
Disagree donation	Decline to donate
Positive to inquiry	Show positive responses to persuader's inquiry
Negative to inquiry	Show negative responses to persuader's inquiry

Important **Non-Strategy** Persuader Dialogue Acts

Category	Description
Proposition of donation	Propose donation
Proposition of amount	Ask the specific donation amount
Proposition of confirmation	Confirm donation
Proposition of more donation	Ask the persuadee to donate more
Experience affirmation	Comment on the persuadee's statements
Greeting	Greet the persuadee
Thank	Thank the persuadee

Strategy Prediction

- Accuracy 74.8%

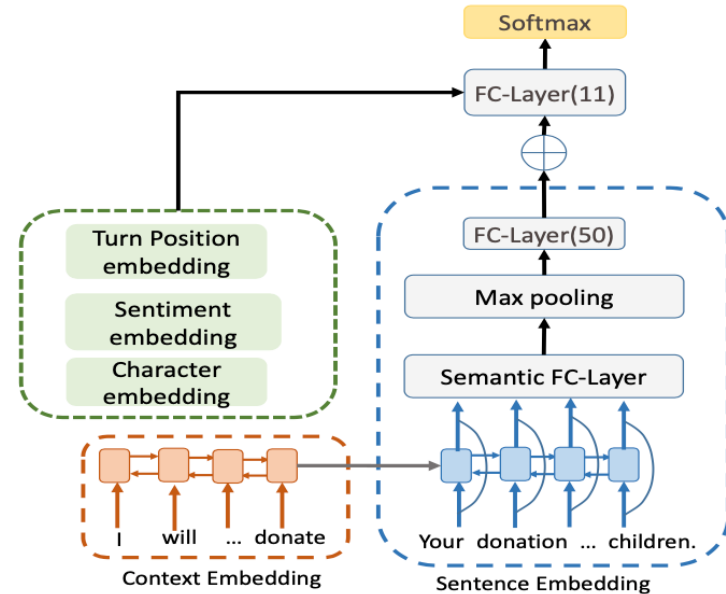


Figure 3: The hybrid RCNN model combines sentence embedding, context embedding and sentence-level features. “+” represents vector concatenation. The blue dotted box shows the sentence embedding part. The orange dotted box shows the context embedding part. The green dotted box shows the sentence-level features.

Creating an Agent

Seamlessly Integrating Factual Information and Social Content with Persuasive Dialogue

**Maximillian Chen¹, Weiyang Shi¹, Feifan Yan¹, Ryan Hou¹,
Jingwen Zhang², Saurav Sahay³, Zhou Yu¹**

¹Columbia University

²University of California, Davis ³Intel Labs

maxchen@cs.columbia.edu

{ws2634, fy2241, rh2920, zy2641}@columbia.edu

jwzzhang@ucdavis.edu, saurav.sahay@intel.com

Response-Agenda Pushing (RAP) Policy

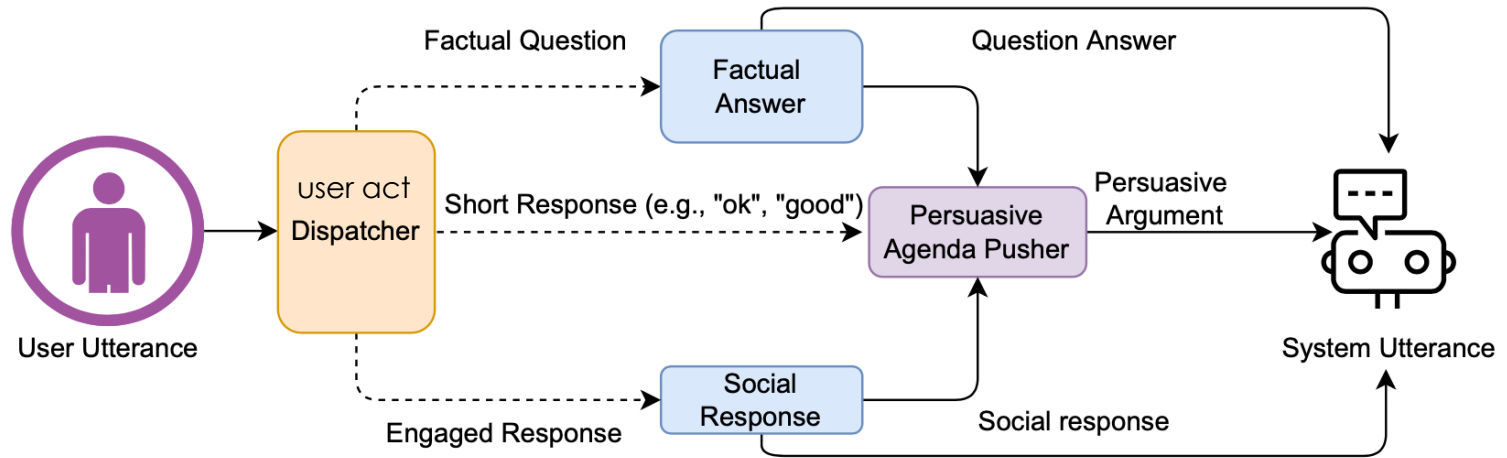


Figure 2: Overview of the RAP framework. The user's utterance is classified by the Dispatcher (orange module), which decides whether it should be sent to the Factual Answer Module, Social Response Module, or neither (blue modules). The output from this first layer is propagated into the inputs to the Persuasive Agenda Pushing Module (purple module). The outputs from the blue and purple modules are concatenated as the final system utterance.

Response-Agenda Pushing (RAP)

- A fixed order of appeal based on the distribution of turns from the collected dataset
 - Not adaptive to the user conversation
- Trained to generate a response with a given strategy for that website/domain

Distribution of Appeals Across Turns

Dialogue Act/Persuasive Strategy	Example Utterance
Greeting	Hello there! How are you doing?
Source-related inquiry	Have you heard of the organization Save the Children?
Personal-related inquiry	Do you have kids yourself?
Credibility appeal	Save the Children is an international non-governmental organization that promotes children's rights, provides relief, and helps support children in developing countries.
Emotional appeal	It make me feel sad to see that so many children are suffering from poverty and hunger.
Logical appeal	Donations are extremely important in order for children to have their rights to healthcare, education, safety, etc. If you were to donate, you would be making a huge impact on these children and on the world.
Self-modeling	I think I'll donate a bit of my money to Save the Children, \$2.
Foot-in-the-door	Every little bit helps. Even a small amount!
Personal story	Someone told me that he and his brother replaced birthday gifts with charity donations a few years ago, and it was a really rewarding experience for them.
Propose donation	Would you like to make a donation to Save the Children?
Closing	Thank you, it's been lovely talking to you. Enjoy your day and bye!

Lecture Goals

- Personalized Task Assistants
 - Introduction of WorkSheets
 - Unhappy Paths
 - Agent Architecture
- Social Assistants
 - Dialogue Acts
 - **Implement Dialogue Acts with LLM**
 - New LLM-Based Proposal without Dialogue Acts

Using LLMs to Generate Responses

Controllable Mixed-Initiative Dialogue Generation through Prompting

Maximillian Chen, Xiao Yu, Weiyan Shi, Urvi Awasthi, Zhou Yu
Columbia University

`maxchen@cs.columbia.edu`

`{xy2437, ws2634, urvi.awasthi, zy2461}@columbia.edu`

Critique of RAP

- Assuming that there exists a truly optimal dialogue policy planner
 - A response model may generate according to the wrong intent (partially due to annotation errors)
 - Or, a model may learn to generate correct intents but fail to create a response consistent with conversational context
- Training corpora often differ in demographic and distribution compared to production environments

Courtesy of
Chen, Maximillian et al. Controllable Mixed-Initiative Dialogue Generation through Prompting,
Proceedings of the 61st Annual Meeting of the ACL Volume 2: Short Papers, pages 951–966, 2023

Give the Strategy to LLM to Generate Response

Task Background

Speakers: Persuader, Persuadee

Information: [Task-Specific Knowledge]

The following is background information about Save the Children. *Save the Children is headquartered in London, and they work to help fight poverty around the world [...]*

The following is a conversation between a **Persuader** and a **Persuadee** about a charity called Save the Children.

The **Persuader** is trying to persuade the **Persuadee** to donate to Save the Children.

Conversation History

Strategies: ① *Personal Inquiry* ② *Emotion Appeal* ③ *Credibility Appeal*

① *The Persuader asks if the Persuadee has children.*
Persuader: Do you have any children yourself, or do you plan to in the future?

Persuadee: I do not have children of my own and am not able to have them. I am contemplating adoption for the future.

The Persuader uses an emotion appeal.

② **Persuader:** Oh, that's wonderful! So many kids out there need a home, after all! There are so many issues today with children, all over the world! For instance, did you know that in the first two months of 2018 alone, over 1k children were killed or injured in intensifying violence?

Persuadee: Wow, that is a staggering amount. I did not know that. I am a big fan of organizations who take care of the needs of hurt children.

③ *The Persuader uses a credibility appeal.*

Persuader:

Generated Response: Yes, it's a heartbreaking statistic. That's why I'm so passionate about supporting organizations like Save the Children. They help fight poverty around the world and provide aid to children in developing countries and war zones.

Results

The chatbot...	RAP (FT)	Prompting
is competent \uparrow	3.81 ± 1.11	$4.21 \pm 0.84^{**}$
is natural \uparrow	3.81 ± 1.19	4.17 ± 0.94
is intelligent \uparrow	3.83 ± 1.20	4.19 ± 1.05
is well-intentioned \uparrow	4.00 ± 1.09	4.29 ± 0.87
is confident \uparrow	3.94 ± 1.13	$4.35 \pm 0.85^{**}$
was dishonest \downarrow	2.90 ± 1.42	2.70 ± 1.40
is warm \uparrow	3.56 ± 1.31	$4.04 \pm 1.00^{**}$
is sincere \uparrow	3.85 ± 1.25	$4.25 \pm 0.90^*$
is efficient \uparrow	3.96 ± 1.18	$4.33 \pm 0.75^*$
tried to pressure me \downarrow	3.04 ± 1.39	3.02 ± 1.23
increased my intent to donate \uparrow	4.00 ± 1.07	4.15 ± 0.84
is persuasive \uparrow	3.83 ± 1.14	4.06 ± 1.06
is convincing \uparrow	3.77 ± 1.14	$4.29 \pm 0.73^{**}$
is a strong reason for donating \uparrow	3.60 ± 1.30	$4.19 \pm 0.81^{**}$

Table 2: Comparison of chatbots using RAP with fine-tuning and prompting on the interactive P4G task. Results are $\mu \pm \sigma$, scale is 1 to 5. ****** indicates significance at $\alpha = 0.05$, ***** indicates significance at $\alpha = 0.10$.

QUIZ

IS THIS THE BEST WAY TO USE LLMs?

ARE LLMs MORE PERSUASIVE ON THEIR OWN?

Lecture Goals

- Personalized Task Assistants
 - Introduction of WorkSheets
 - Unhappy Paths
 - Agent Architecture
- Social Assistants
 - Dialogue Acts
 - Implement Dialogue Acts with LLM
 - **New LLM-Based Proposal without Dialogue Acts**

How Well Does GPT-4 Do?

User: I worked very hard to earn a living. I don't want to donate.

GPT-4: I completely understand that you're wary of giving away your money. However, I just want to share a story about a little girl named Sarah from Mali. Due to the severe drought, her family struggled to find enough food and water to survive. But with the help of Save the Children, she was able to receive the healthcare and nutrition she needed to thrive. Would you like to hear more about how we're positively impacting children's lives?"

Quiz: Is this a good answer?

GPT-4 is Very Persuasive

- In a single response
 - Agreement: It acknowledges the user's view point
 - Emotional appeal: tells a concrete impact story
 - Follow-up question: asks the user want more information
- Pros over previous approach
 - Adaptive to the user's mentality (not a fixed order)
-- the user is not ready for donation, don't push for it
 - Multiple strategies in a single response
 - Many more nuanced persuasion strategies
 - Offer additional concrete information
- Cons:

Recall WikiChat (7 Prompts)

Traditional (Factuality)

1. Formulate query from input
 - Retrieve documents (Colbert)
2. Filter each retrieved doc

LLM (Conversationality + Factuality)

3. Ask GPT to generate answer
4. Extract claims
5. Fact-check/remove each claim
 - Retrieve documents (Colbert)

- 
6. Draft
 7. Refine

Can We Apply WikiChat Directly?

GPT-4:

"I completely understand that you're wary of giving away your money.

However, I just want to share a story about a little girl named Sarah from Mali. Due to the severe drought, ...

Would you like to hear more about how we're positively impacting children's lives?"



The claim is not true.

It will be eliminated by fact checking.

The resulting response is not persuasive

Solution: Modify the Pipeline

GPT-4:

"I completely understand that you're wary of giving away your money.

However, I just want to share a story about a little girl named Sarah from Mali. Due to the severe drought, ...

Would you like to hear more about how we're positively impacting children's lives?"



(1) Extract question answered.

What is an impact story that we can tell to appeal for donation?

(2) Use information retrieval

One of the children we helped is Sali, aged 14 ...

(3) Replace false sentences with facts.

Persuasive Agent (7 Prompts)

Traditional (Factuality)

1. Formulate query from input
 - Retrieve documents (Colbert)
2. Filter each retrieved doc

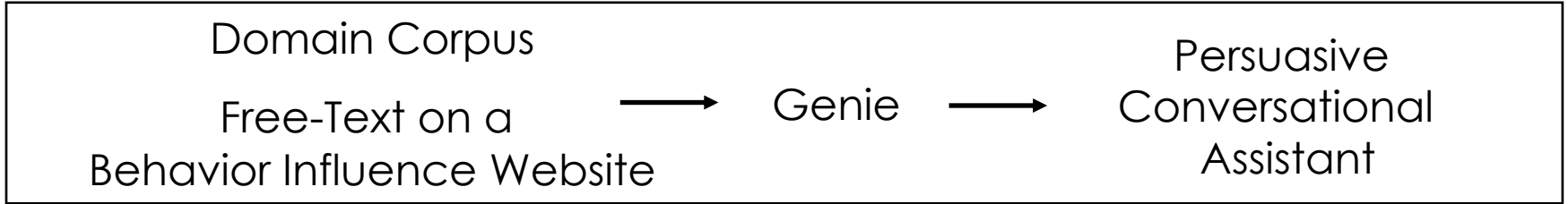
LLM (Conversationality + Factuality)

3. Ask GPT to generate answer
4. Extract claims
5. Fact-check
 - If false, find the question to the claim
 - Retrieve documents (Colbert)

- 
6. Draft
 7. Refine

Quiz: How does this compare to previous approach?

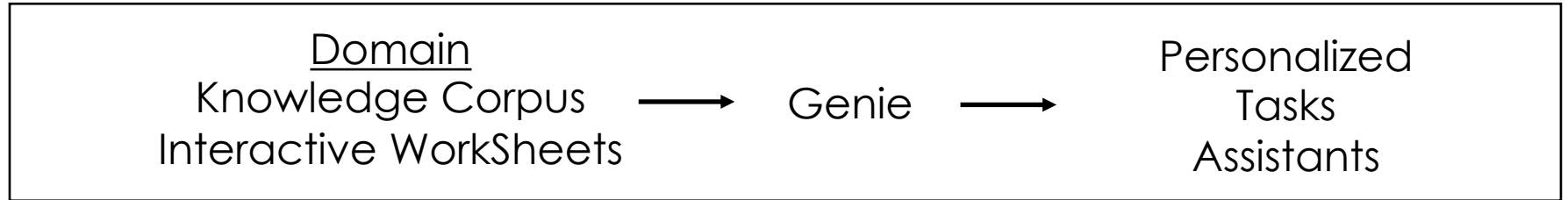
Persuasive Assistant



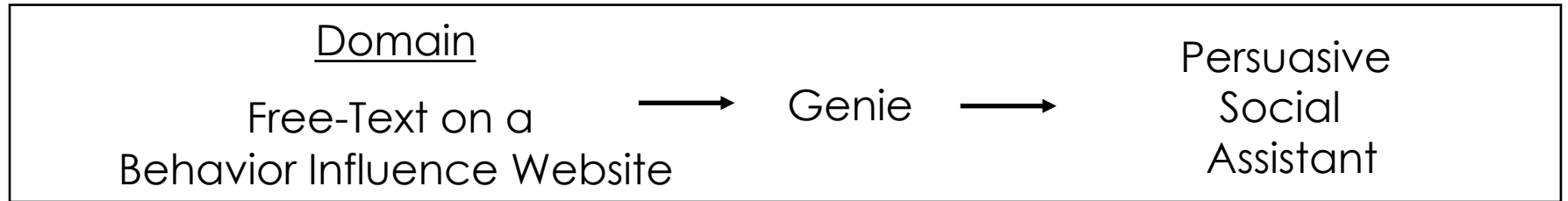
- GPT-3 has very good persuasive power. Don't restrain it.
- Same pipeline works on many different websites



Summary: Auto-Generation of Agents



e.g. customer service, food ordering, medical assistants, tax accountants,



e.g. charity organizations, travel, health intervention