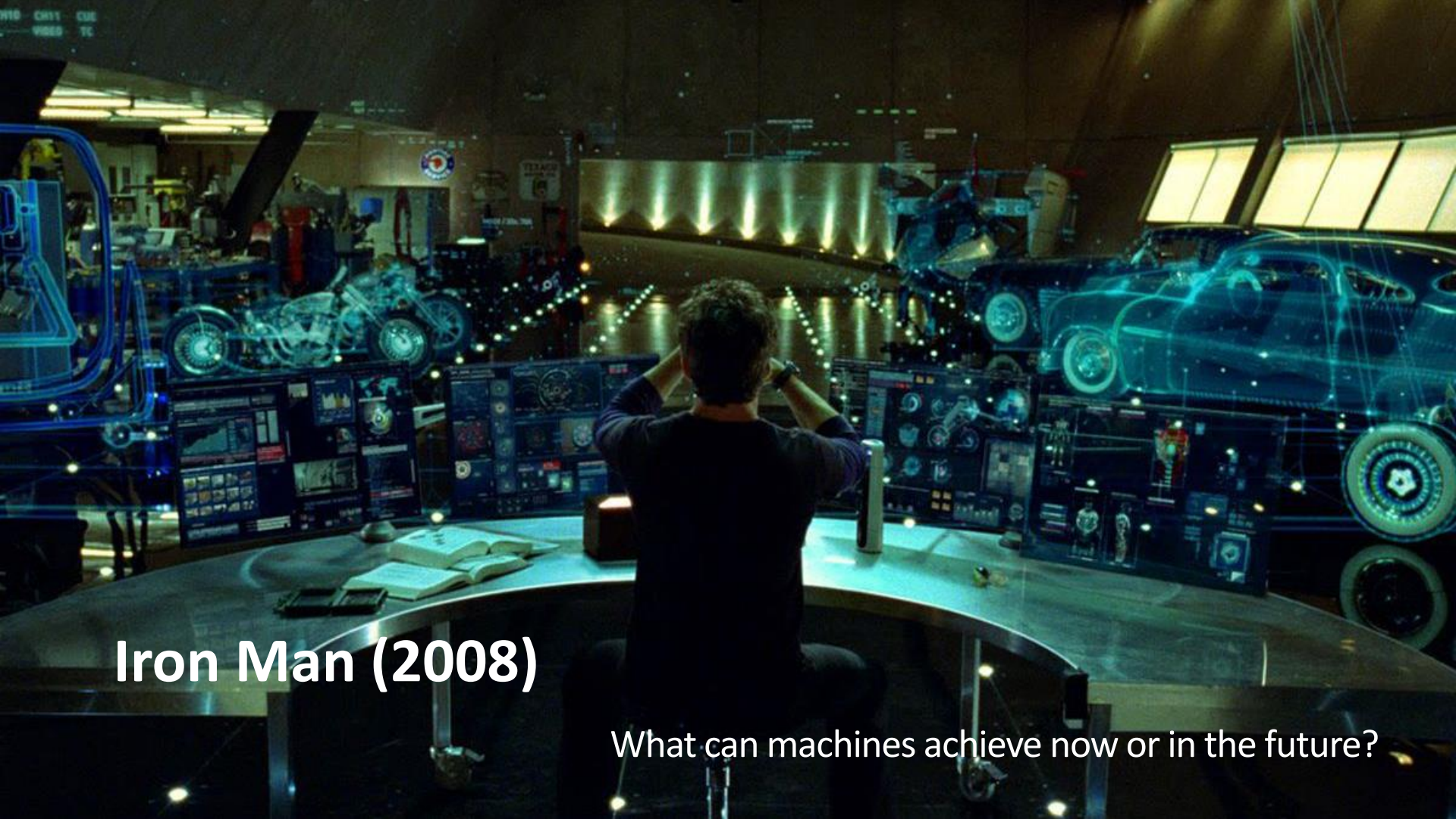
Towards Open-Domain Conversational AI

YUN-NUNG (VIVIAN) CHEN 陳縕儂
HTTP://VIVIANCHEN.IDV.TW

National Taiwan University

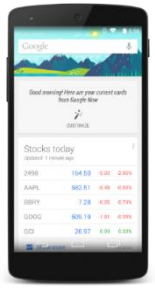


Iron Man (2008)

What can machines achieve now or in the future?

Language Empowering Intelligent Assistants

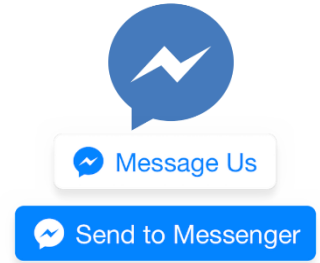
Apple Siri (2011)



Google Now (2012)
Google Assistant (2016)

Microsoft Cortana (2014)

Amazon Alexa/Echo (2014)



Facebook M & Bot (2015)

Google Home (2016)

Apple HomePod (2017)

Why and When We Need?

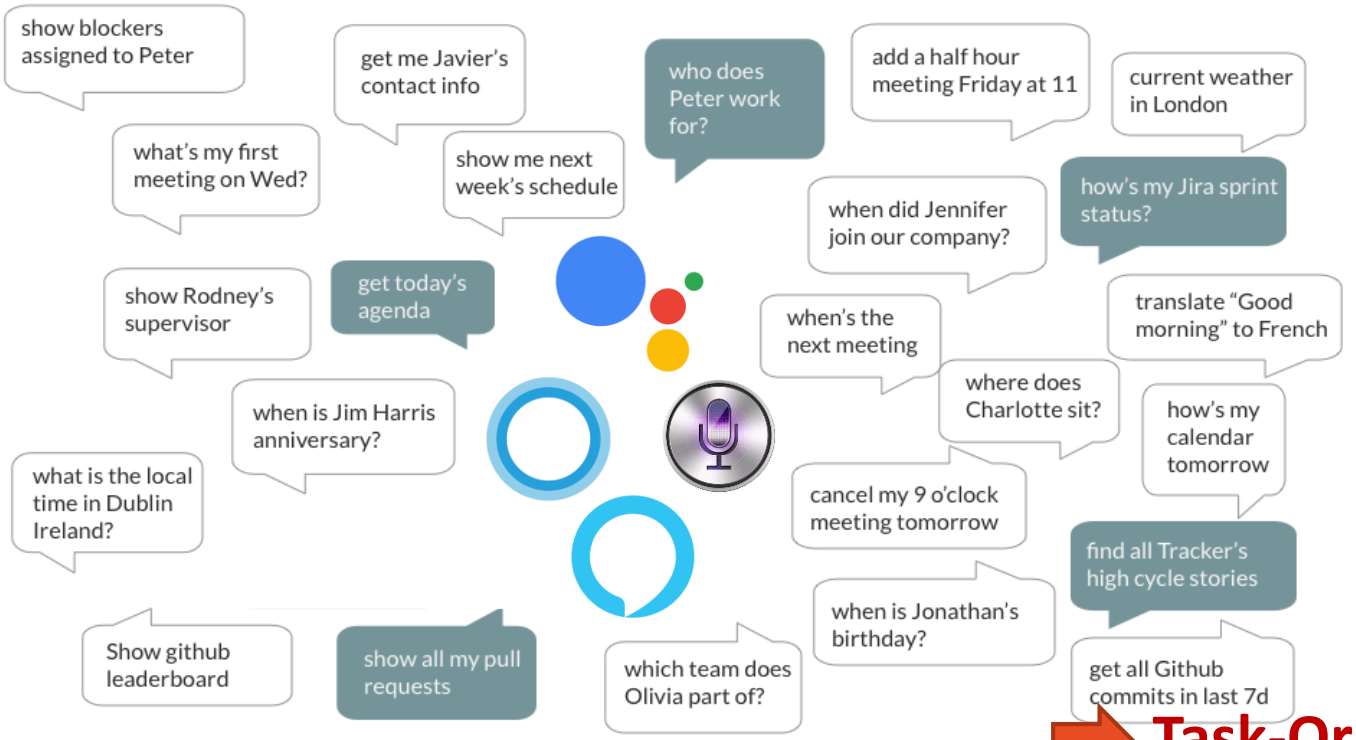
“I want to chat”	Turing Test (talk like a human)	Social Chit-Chat
“I have a question”	Information consumption	Task-Oriented Dialogues
“I need to get this done”	Task completion	
“What should I do?”	Decision support	

- *What is today’s agenda?*
- *What does SLT stand for?*

- *Book me the flight ticket from Taipei to Athens*
- *Reserve a table at Din Tai Fung for 5 people, 7PM tonight*

- *Is SLT conference good to attend?*

Intelligent Assistants



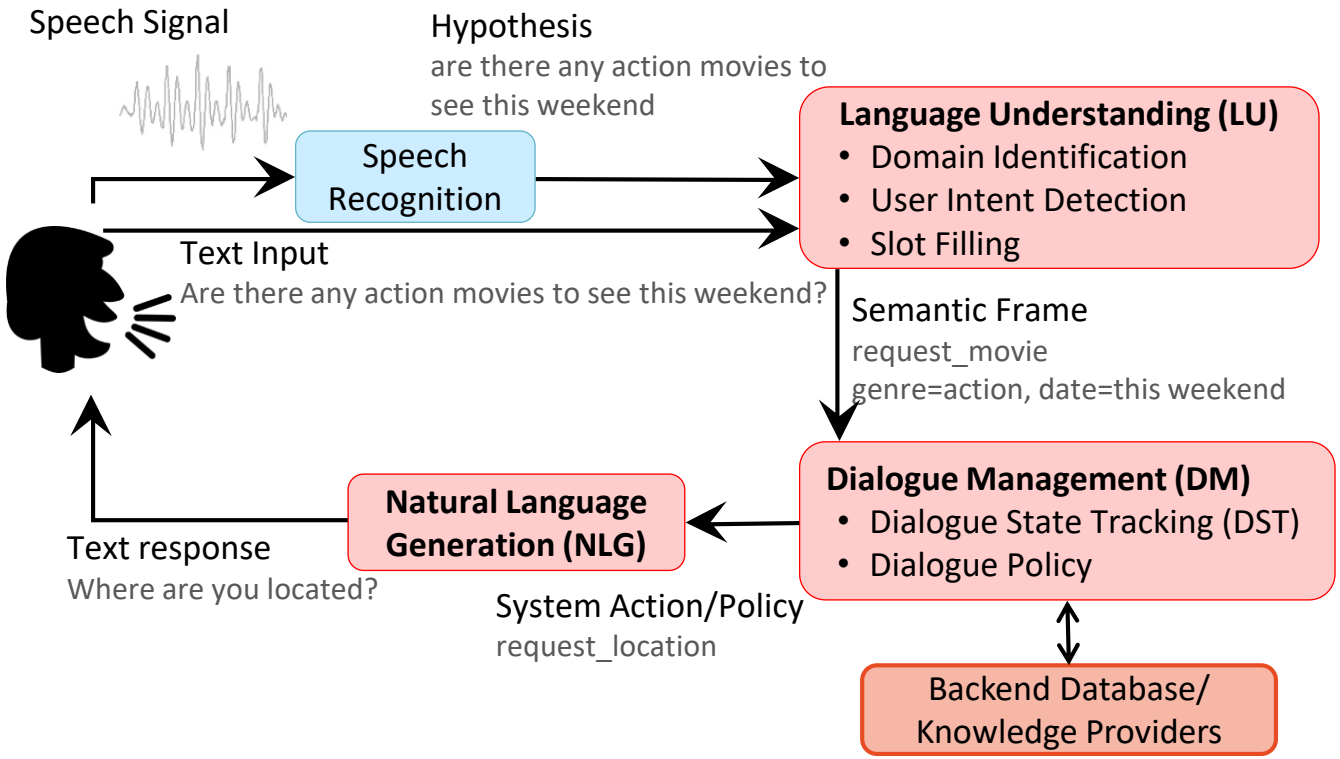
Task-Oriented

Task-Oriented Dialogue Systems

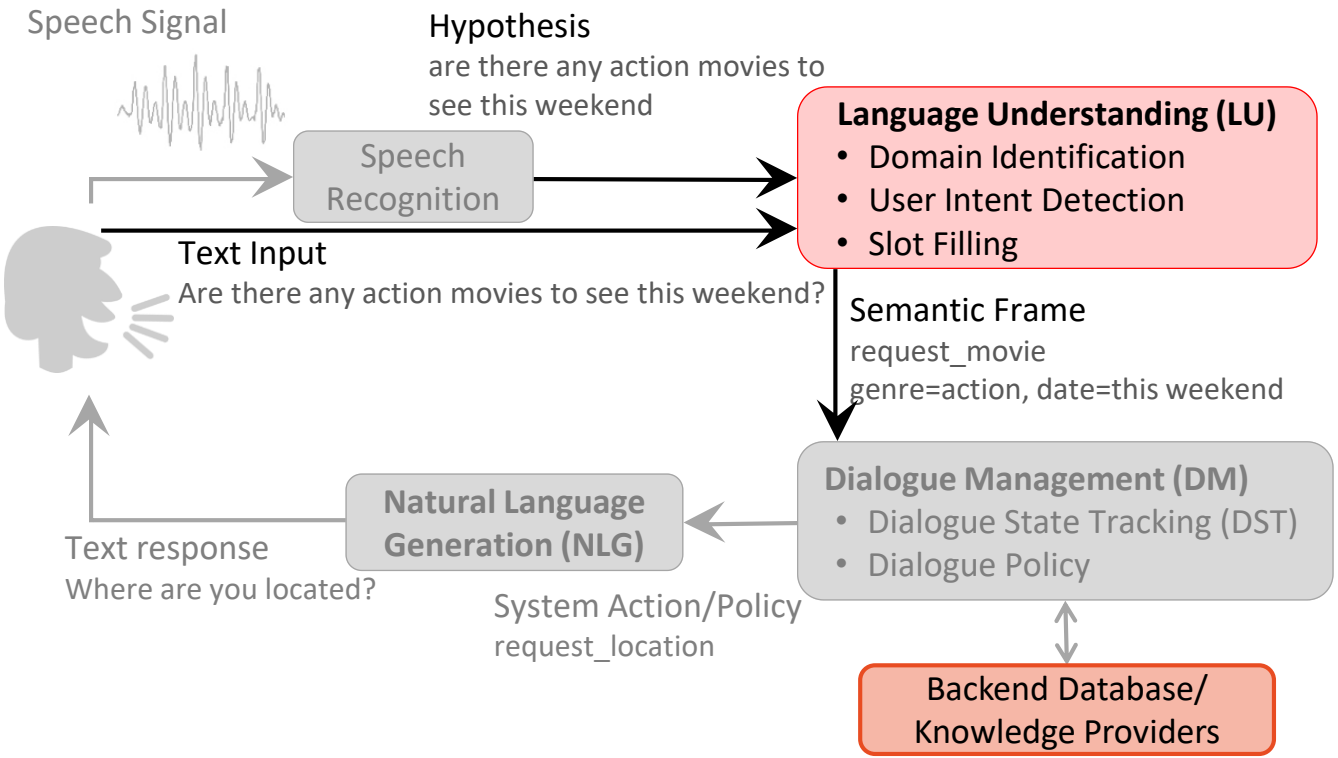
JARVIS – Iron Man's Personal Assistant

Baymax – Personal Healthcare Companion

Task-Oriented Dialogue Systems (Young, 2000)

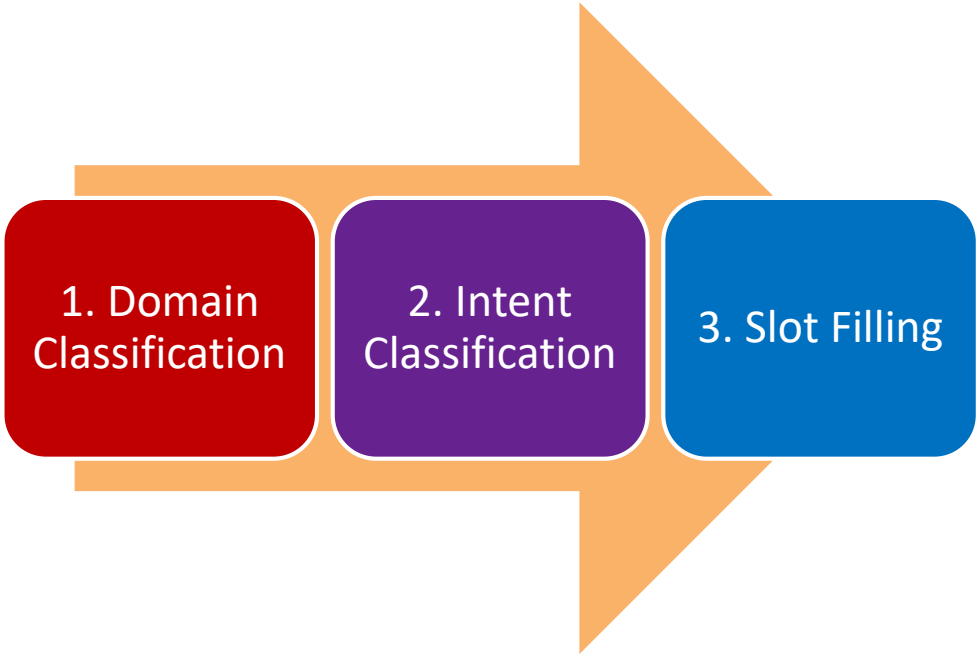


Task-Oriented Dialogue Systems (Young, 2000)



Language Understanding (LU)

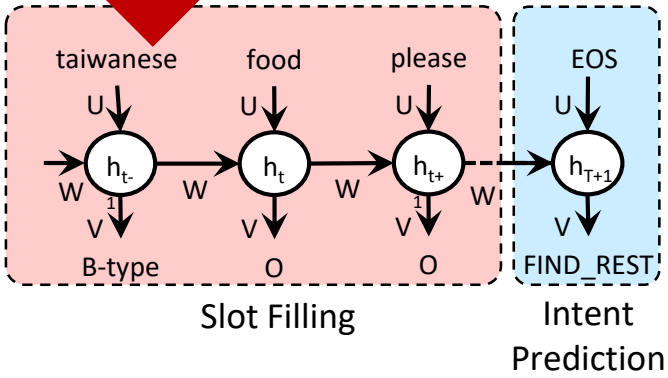
- Pipelined



Joint Semantic Frame Parsing

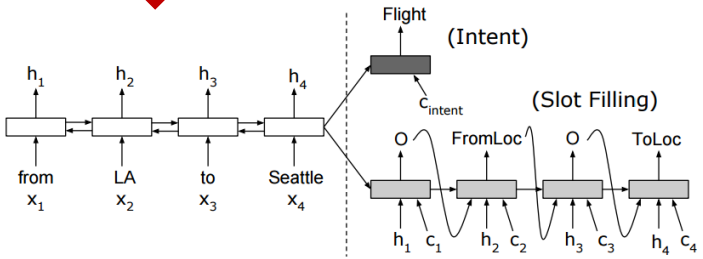
Sequence-based (Hakkani-Tur et al., 2016)

- Slot filling and intent prediction in the same output sequence



Parallel (Liu and Lane, 2016)

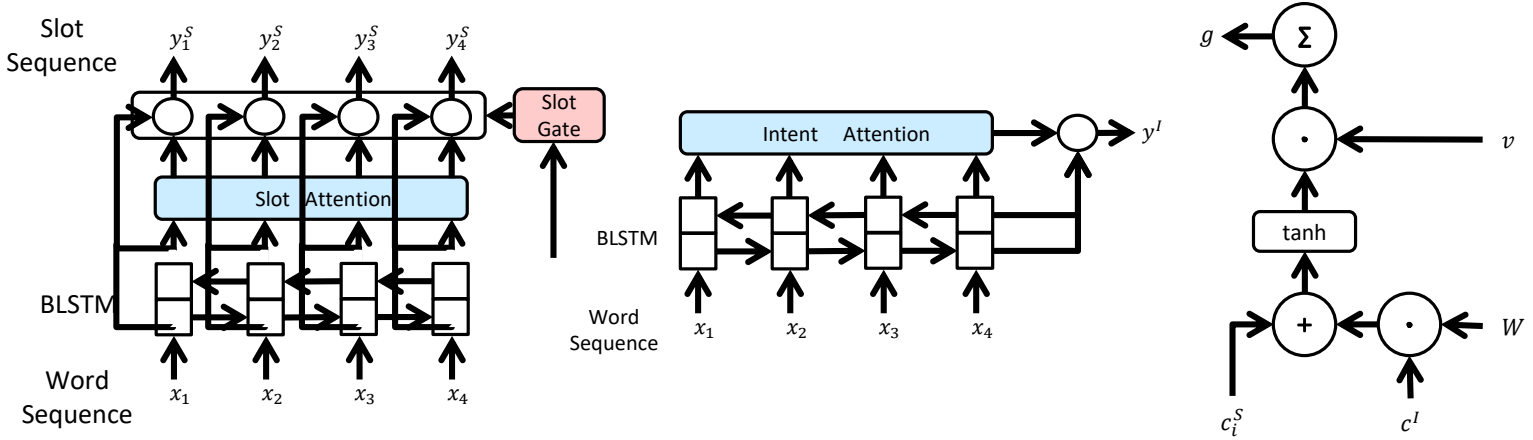
- Intent prediction and slot filling are performed in two branches



Joint Model Comparison

	Attention Mechanism	Intent-Slot Relationship
Joint bi-LSTM	X	Δ (Implicit)
Attentional Encoder-Decoder	√	Δ (Implicit)
Slot Gate Joint Model	√	√ (Explicit)

Slot-Gated Joint SLU (Goo et al., 2018)



- Slot Gate $g = \sum v \cdot \tanh(c_i^S + W \cdot c^I)$
- Slot Prediction $y_i^S = softmax(W^S(h_i + c_i^S) + b^S)$ ➡ $y_i^S = softmax(W^S(h_i + \boldsymbol{g} \cdot c_i^S) + b^S)$

$\boldsymbol{g}$ will be larger if slot and intent are better related

c_i^S: slot context vector
c^I : intent context vector
W: trainable matrix
v : trainable vector
g : scalar gate value

Contextual Language Understanding

Single Turn

U just sent email to bob about fishing this weekend

S O O O O O B-contact_name B-subject I-subject I-subject

→ send_email(contact_name="bob", subject="fishing this weekend")

Multi-Turn

U_1 send email to bob

S_1 B-contact_name

→ send_email(contact_name="bob")

U_2 are we going to fish this weekend

S_2 B-message I-message I-message I-message I-message I-message I-message

→ send_email(message="are we going to fish this weekend")

E2E MemNN for Contextual LU (Chen et al., 2016)

U: "i d like to purchase tickets to see deepwater horizon" → 0.69

S: "for which theatre"

U: "angelika"

S: "you want them for angelika theatre?"

U: "yes angelika"

S: "how many tickets would you like ?" → 0.13

U: "3 tickets for saturday"

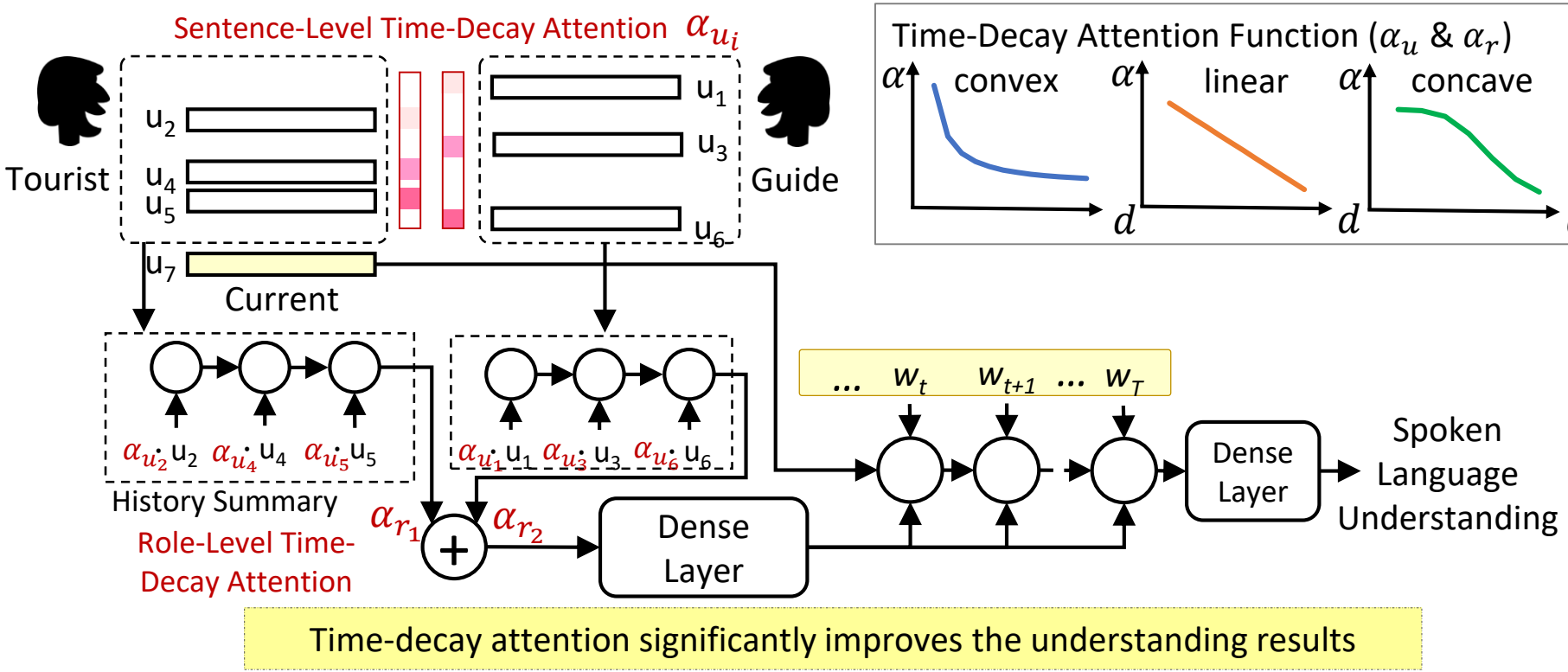
S: "What time would you like ?"

U: "Any time on saturday is fine"

S: "okay , there is 4:10 pm , 5:40 pm and 9:20 pm" → 0.16

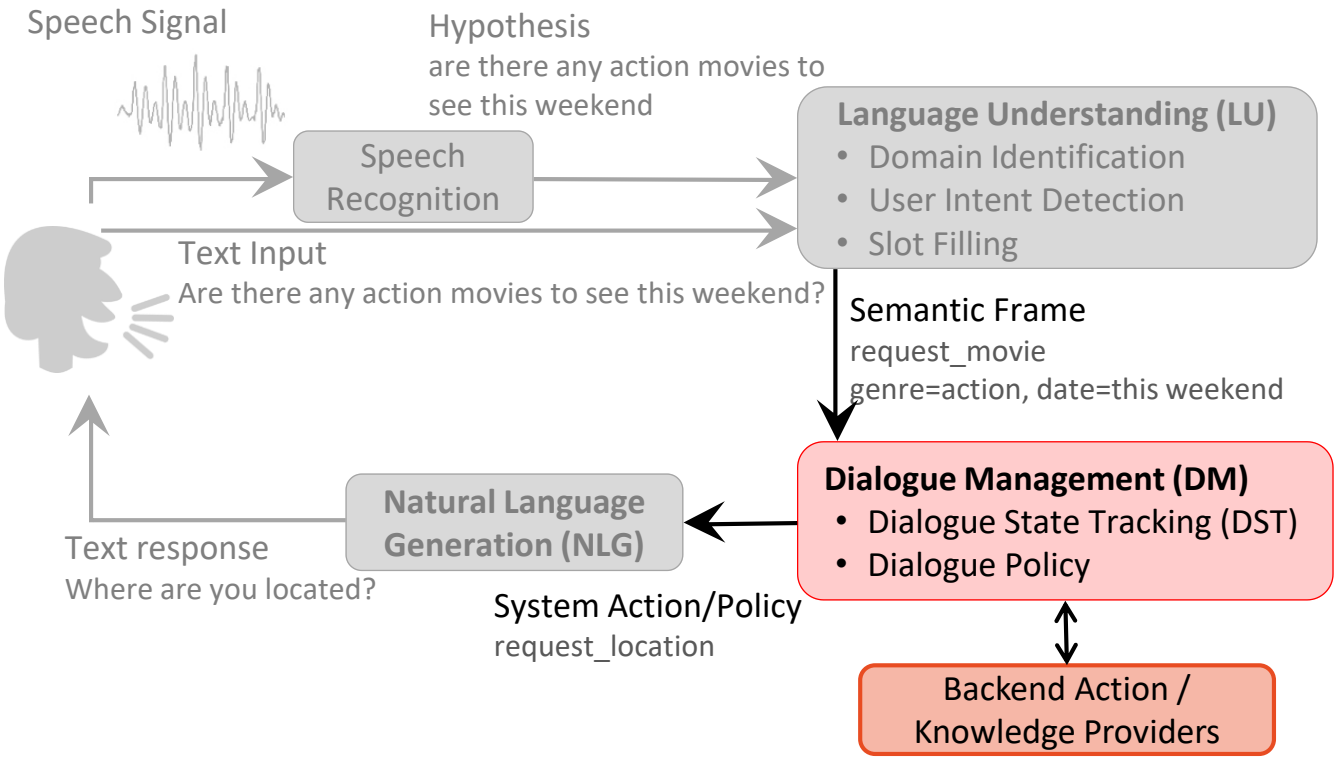
U: "Let's do 5:40"

Role-Based & Time-Aware Attention (Su et al., 2018)



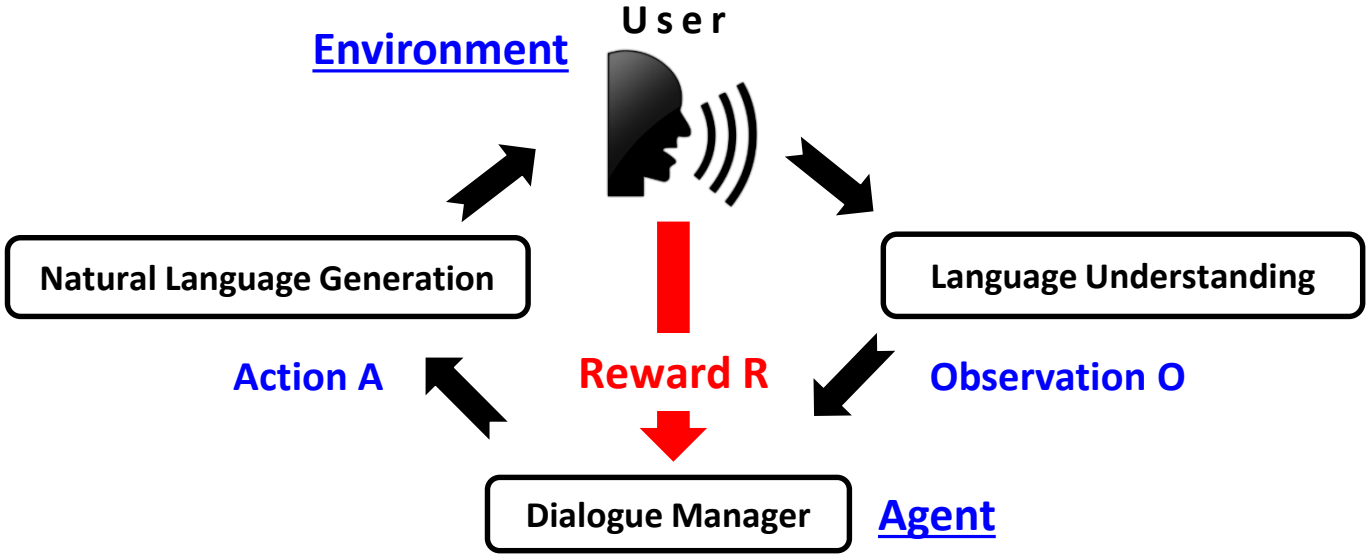
Time-decay attention significantly improves the understanding results

Task-Oriented Dialogue Systems (Young, 2000)



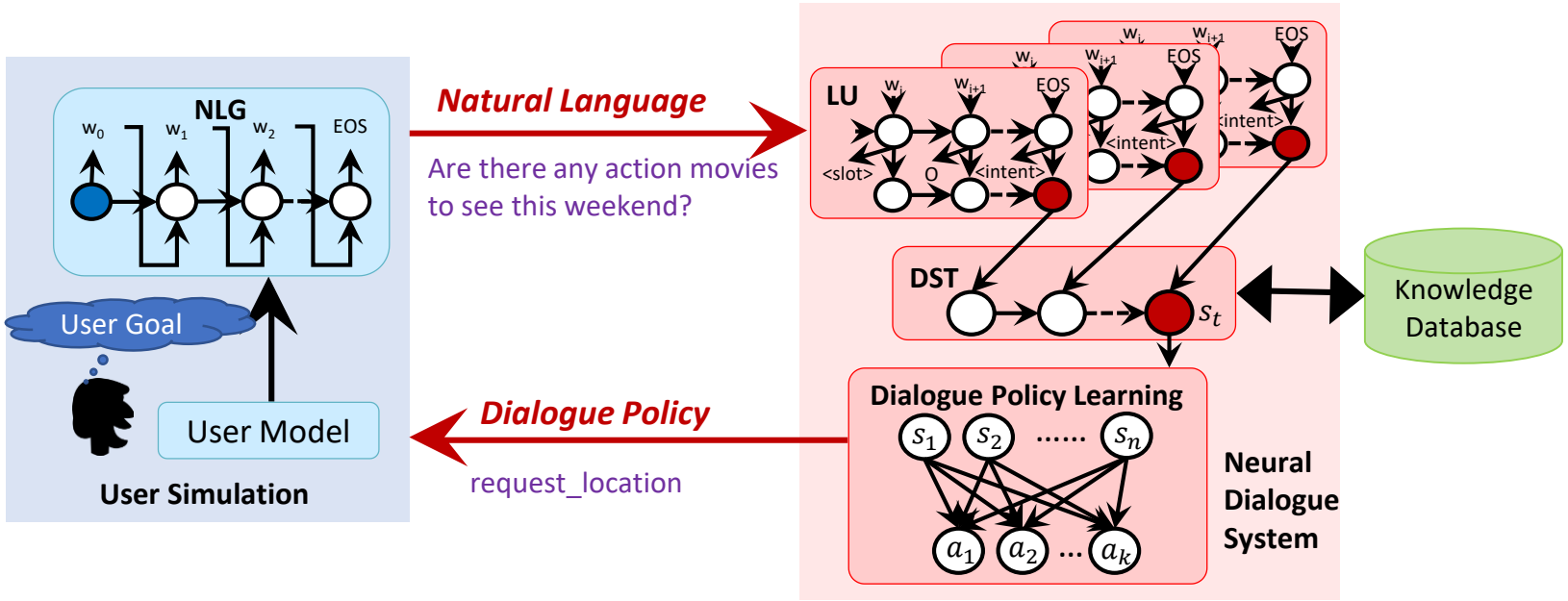
Dialogue Policy Optimization

- Dialogue management in a RL framework



Select the best action that **maximizes the future reward**

E2E Task-Completion Bot (TC-Bot) (Li et al., 2017)



Idea: SL for each component and RL for end-to-end training the neural dialogue system

Reinforcement Learning Agent (Li et al., 2017)

User goal – ***Two tickets for "Deadpool" tomorrow 9PM at AMC Pacific Place 11 theater, Seattle.***

RULE BASED SYSTEM

User: Do you have any ticket for movie 'Deadpool' at 9:00 pm?
Agent: Which theater do you want?
User: AMC pacific place 11 theater, please!
Agent: How many tickets do you want?
User: Two!
Agent: What date?
User: Tomorrow!
Agent: Which city do you want?
User: Seattle!
Agent: Do you have any specification for zip code?
User: No, I do not care.
Agent: Do you have any distance constraint?
User: No, I do not care.
Agent: How about video format? 2d? Imax 3d?
User: Any kind is okay.
Agent: Do you have any constraint to price?
User: No!
Agent: Okay, I will issue 2 tickets for you, tomorrow 9:00 pm at AMC pacific place 11 theater, Seattle, movie 'Deadpool'.
User: Okay, thanks.
Agent: Thanks!

no notion about what requests can be skipped

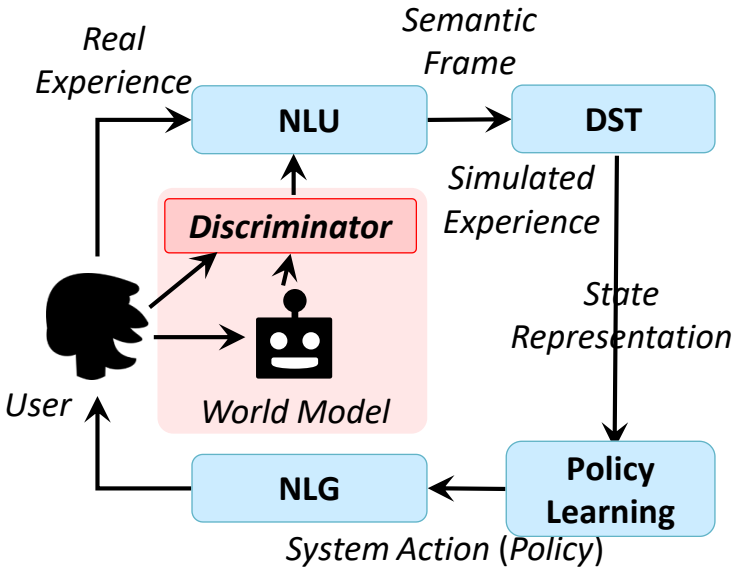
REINFORCEMENT LEARNING SYSTEM

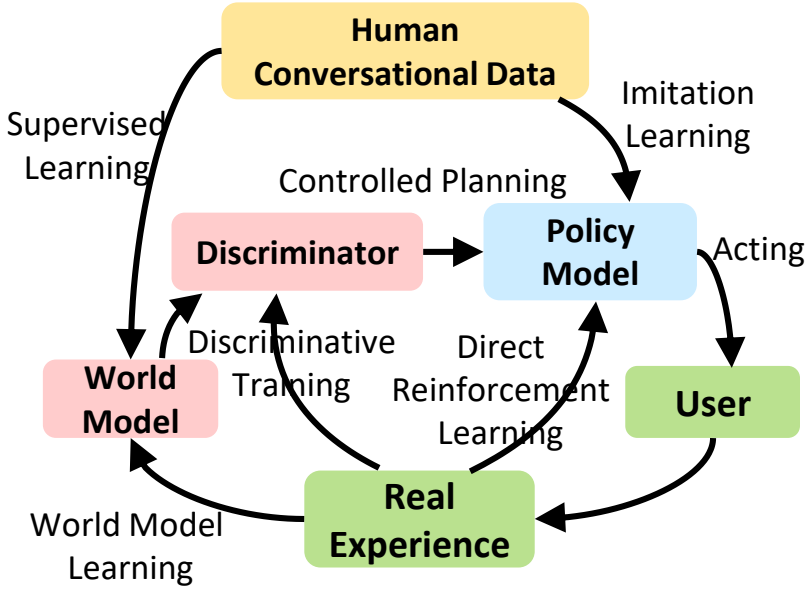
User: I want two tickets for movie 'Deadpool'.
Agent: What is the starttime you prefer?
User: 9:00 PM
Agent: Which day?
User: Tomorrow
Agent: Which city?
User: Seattle
Agent: Which theater do you prefer?
User: AMC Pacific Place 11 theater.
Agent: Okay, I will issue 2 tickets for you, tomorrow 9:00 pm at AMC pacific place 11 theater, Seattle, movie 'Deadpool'.
User: Okay, thanks.
Agent: Thanks!

Skip the requests the user may not care about to improve efficiency

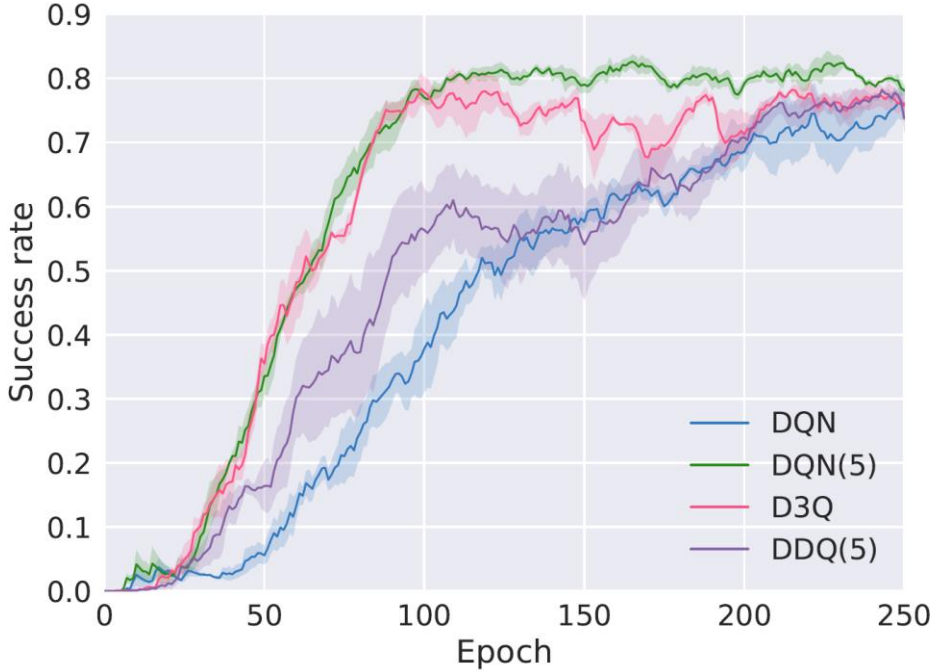
D3Q: Discriminative Deep Dyna-Q (Su et al., 2018)

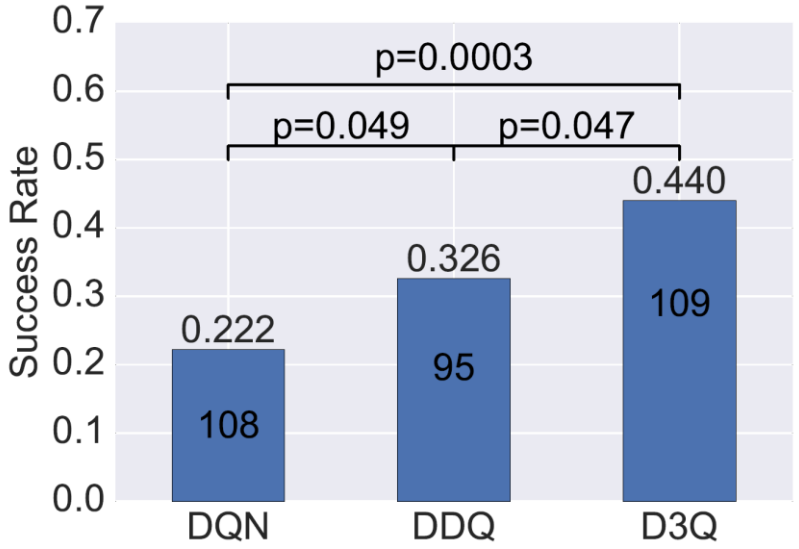
- Issue: real users are expensive, discrepancy between real users and simulators
- Idea: learning with real users with planning, *discriminator* to filter out bad experiences





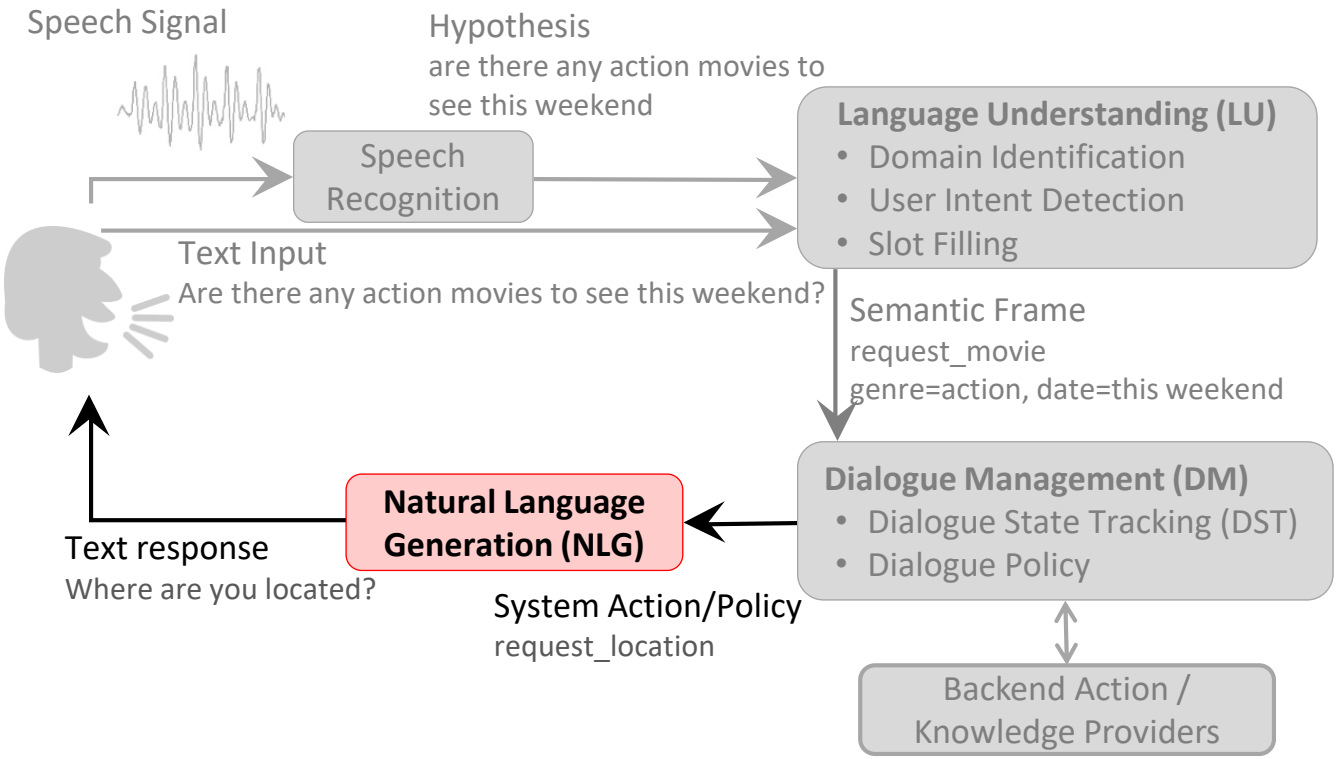
D3Q: Discriminative Deep Dyna-Q (Su et al., 2018)





The policy learning is more robust and shows the improvement in human evaluation

Task-Oriented Dialogue Systems (Young, 2000)



Natural Language Generation (NLG)

- Mapping dialogue acts into natural language

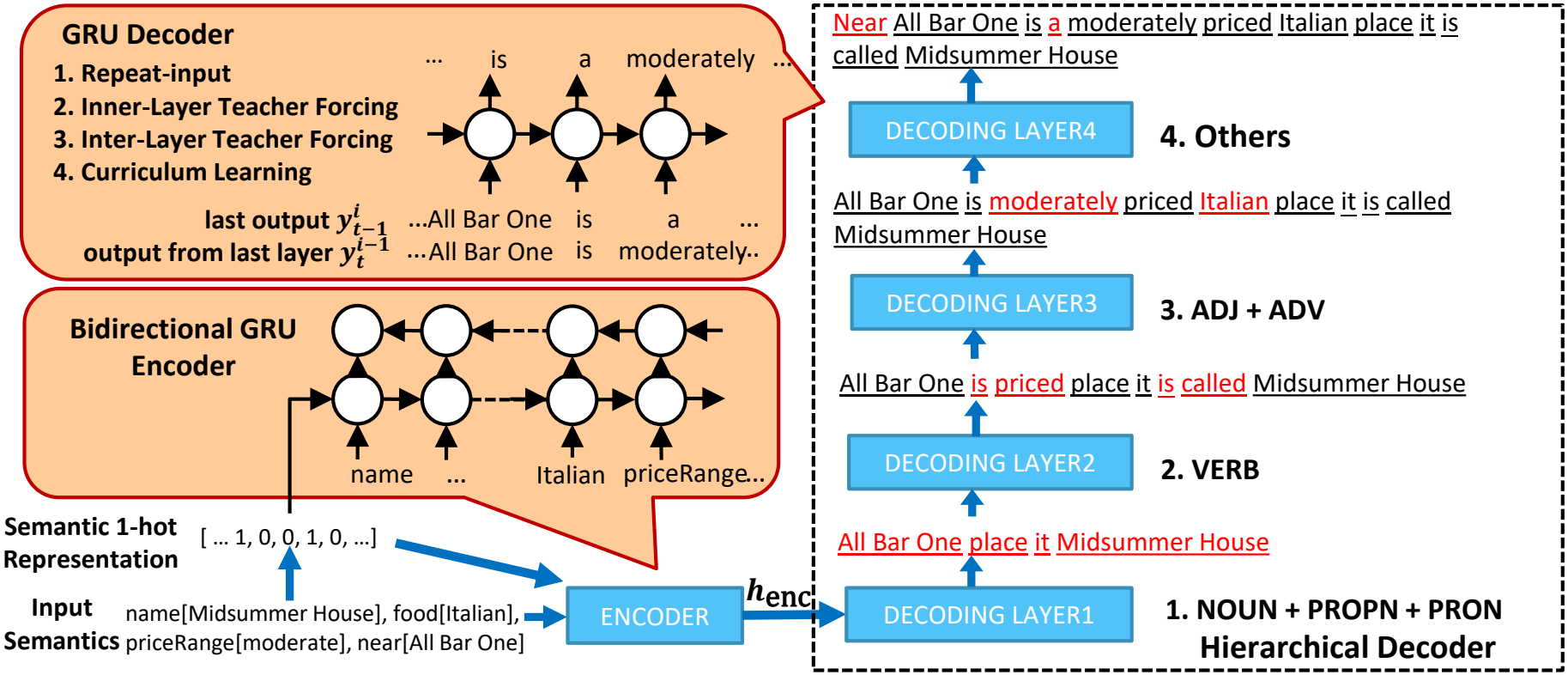
inform(name=Seven_Days, foodtype=Chinese)

Seven Days is a nice Chinese restaurant

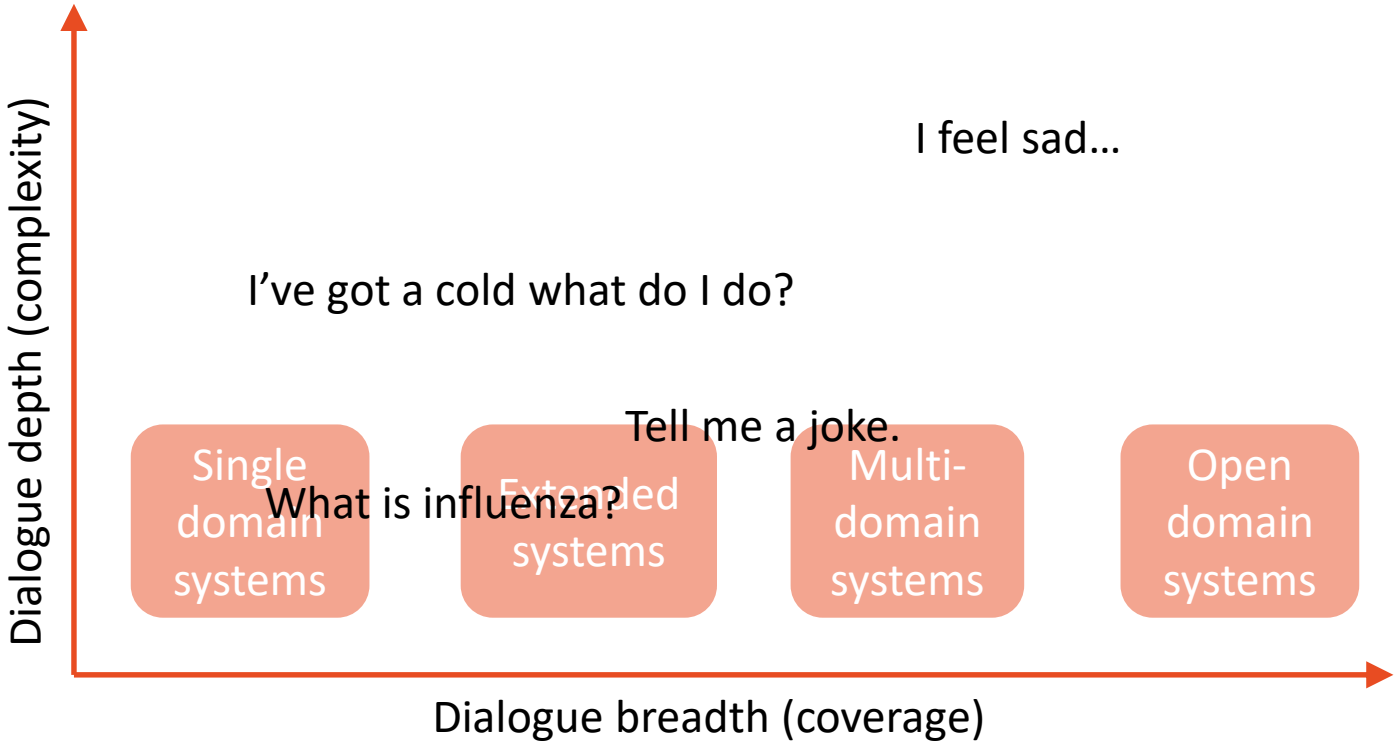
Issues in Neural NLG

- Issue
  - NLG tends to generate shorter sentences
  - NLG may generate grammatically-incorrect sentences
- Solution
  - Generate word patterns in a order
  - Consider linguistic patterns

Hierarchical NLG w/ Linguistic Patterns (Su et al., 2018)



Evolution Roadmap



Dialogue Systems

Task-Oriented Dialogue

input **x**

output **y**

Understanding (NLU)

State tracker

Generation (NLG)

Dialog policy

DB

Database
Memory
External knowledge

Fully Data-Driven

input **x**

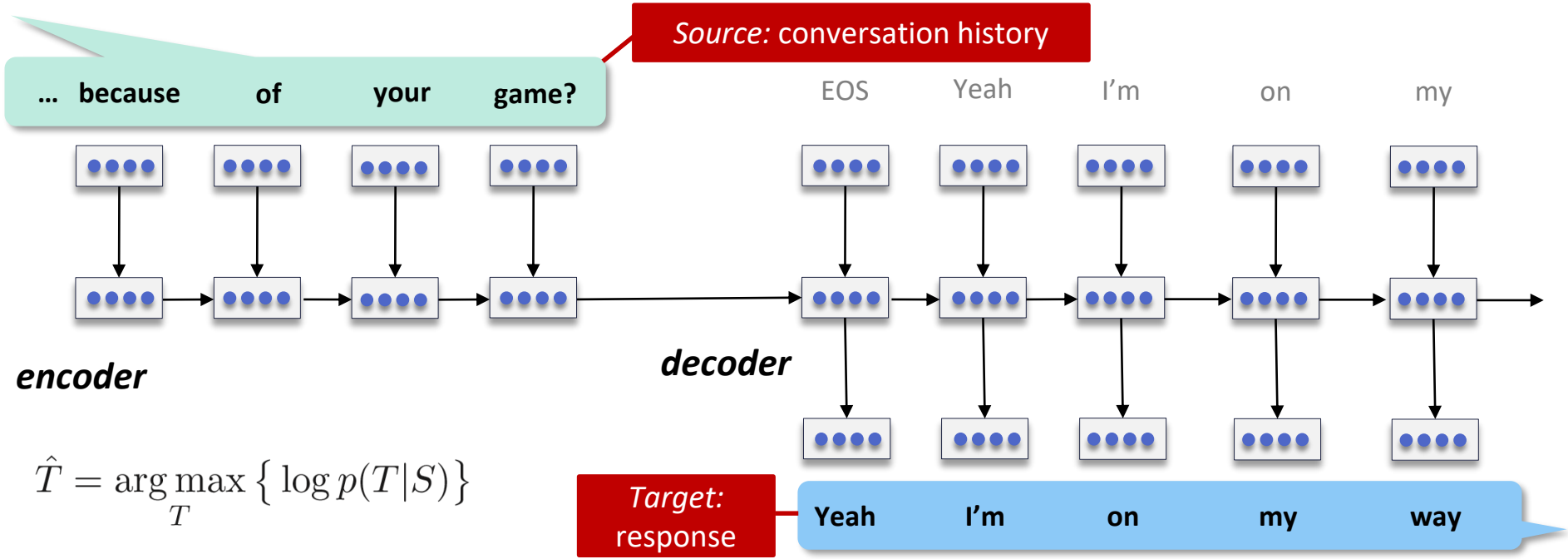
output **y**

Statistical model
(e.g., neural)

DB

Chit-Chat Social Bots

Neural Response Generation (Sordoni et al., 2015; Vinyals & Le, 2015)



$$\hat{T} = \arg\max_{T} \{ \log p(T|S) \}$$

Learns to generate dialogues from offline data (no state, action, intent, slot, etc.)

Issue 1: Blandness Problem

Wow sour starbursts really do make your mouth water... mm drool. Can I have one?

Of course!

Milan apparently selling Zlatan to balance the books... Where next, Madrid?

I don’t know.

‘tis a fine brew on a day like this! Stro

I'm not sure yet,

Well he was on in Bromley a wh

I don't even know what he's talking about.

32% responses are general and meaningless

“I don’t know”

“I don’t know what you are talking about”

“I don’t think that is a good idea”

“Oh my god”

MMI for Response Diversity (Li et al., 2016)

Wow sour starbursts really do make your mouth water... mm drool. Can I have one?

Of course you can! They're delicious!

Milan apparently selling Zlatan to balance the books... Where next, Madrid?

I think he'd be a good signing.

'tis a fine brew on a day like this! Strong though, how many is sensible?

Depends on how much you drink!

Well he was on in Bromley a while ago... still touring.

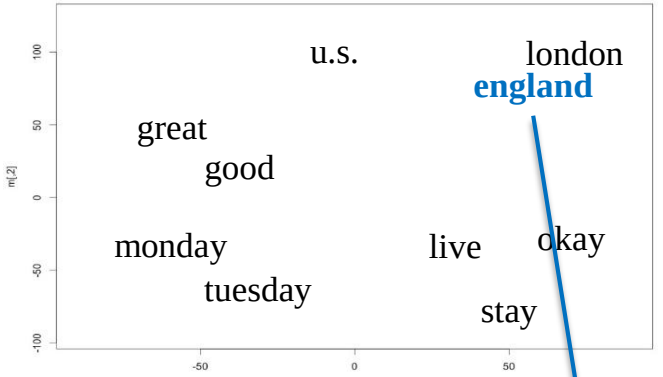
I've never seen him live.

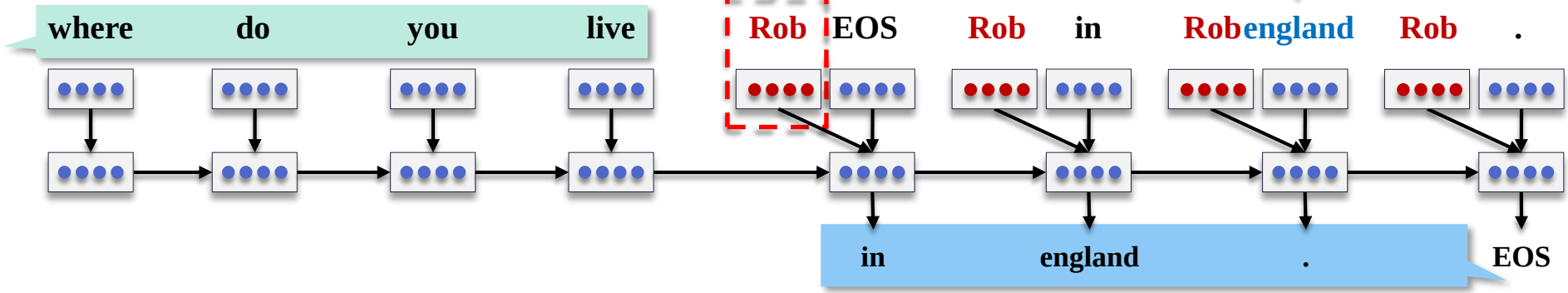
Issue 2: Response Inconsistency

message	Where do you live now?
response	I live in Los Angeles.
message	In which city do you live now?
response	I live in Madrid.
message	In which country do you live now?
response	England, you?
message	Where were you born?
response	I was born in Canada.
message	Where are you from?
response	England, you?
message	Where did you grow up?
response	I grew up in Texas.

message	How old are you?
response	16 and you?
message	What's your age?
response	18.
message	What is your major?
response	I'm majoring in psychology
message	What did you study in college?
response	English lit.

Personalized Response Generation (Li et al., 2016)





Persona Model for Speaker Consistency (Li et al., 2016)

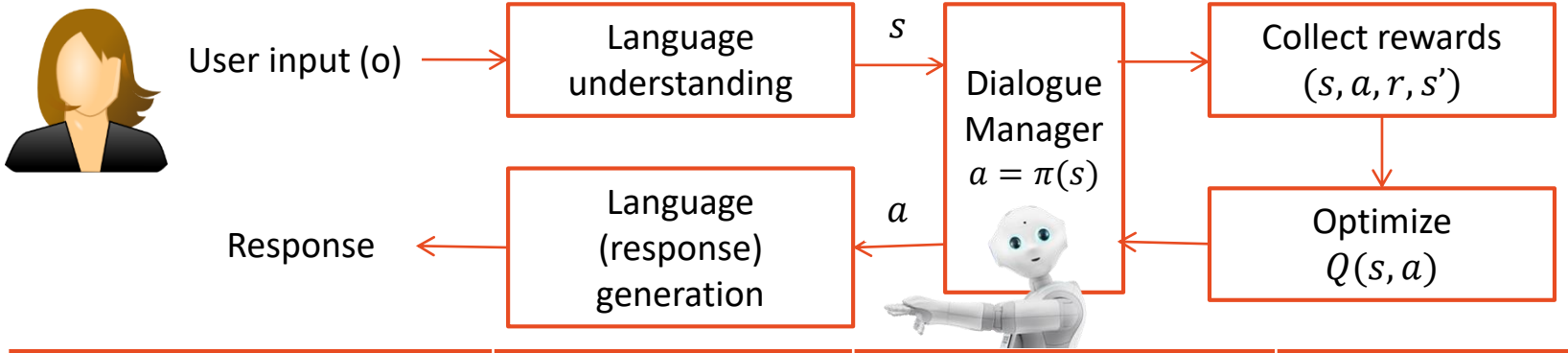
Baseline model → inconsistency

message	Where is your hometown?
response	I live in Indonesia.
message	Where are you from?
response	England, you?
message	In which city do you live now?
response	I live in Jakarta.
message	In which country do you live now?
response	I live in Indonesia.
message	What is your major?
response	Business, you?
message	What did you study in college?
response	Psychology, you?
message	How old are you?
response	I'm 18.
message	What is your age?
response	I'm 16.

Persona model using speaker embedding → consistency

message	Where is your hometown?
response	I'm from England.
message	Where are you from?
response	I'm from England.
message	In which city do you live now?
response	I live in London.
message	In which country do you live now?
response	I live in England.
message	What is your major?
response	Business. You?
message	What did you study in college?
response	I did business studies.
message	How old are you?
response	I'm 18.
message	What is your age?
response	I'm 18.

Issue 3: Dialogue-Level Optimization via RL



Application	State	Action	Reward
Task Completion Bots (Movies, Restaurants, …)	User input + Context	Dialog act + slot-value	Task success rate # of turns
Info Bots (Q&A bot over KB, Web etc.)	Question + Context	Clarification questions, Answers	Relevance of answer # of turns
Social Bot (XiaoIce)	Conversation history	Response	Engagement(?)

Deep RL for Response Generation (Li et al., 2016)

Input message	Supervised Learning Agent	Reinforcement Learning Agent
How old are you ?	I'm 16.	I'm 16, why are you asking ?
What's your full name ?	I have no idea .	What's yours ?
I don't want to go home tonight.	Really ?	Why ?
Do you have any feelings for me ?	I don't know what you are talking about.	I don 't want to hurt your feelings .
How much time do you have here?	Not long enough. Sorry, sir.	Ten seconds .
Shall we get started ?	Of course !	Yes . We 've got a lot of work to do here .
Do you play football ?	No, i don't	Yes. I love football !
We'd have to talk to him.	I mean, he's a good guy	About what ?
How come you never say it?	Because I don't want to hurt you .	I don 't think it 's a good idea to say it .

- RL agent generates more *interactive* responses
- RL agent tends to end a sentence *with a question* and hand the conversation over to the user

Issue 4: No Grounding (Sordoni et al., 2015; Li et al., 2016)

Neural model learns the general shape of conversations, and the system output is situationally appropriate and coherent.

No grounding into a real calendar, but the "shape" of the conversation is fluent and plausible.

H: hi
M: how are you?
H: not bad
M: what's wrong?
H: nothing really
M: wanna talk about it? i 'm here if you wanna talk
H: well, i'm just tired
M: me too, i'm here if you wanna talk

H: would thursday afternoon work sometime?
M: yeah , sure . just let me know when you're free.
H: after lunch is probably the best time
M: okay, sounds good . just let me know when you're free.
H: would 2 pm work for you?
M: works for me.
H: well let's say 2 pm then i 'll see you there
M: sounds good.

Chit-Chat v.s. Task-Oriented

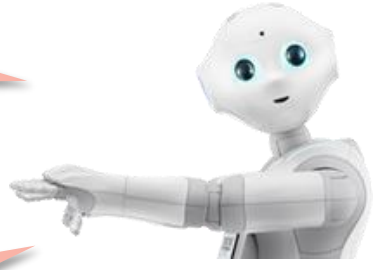
Social Chat
Engaging, Human-Like Interaction
(Ungrounded)

Task-Oriented
Task Completion, Decision Support
(Grounded)

The weather is so depressing these days.

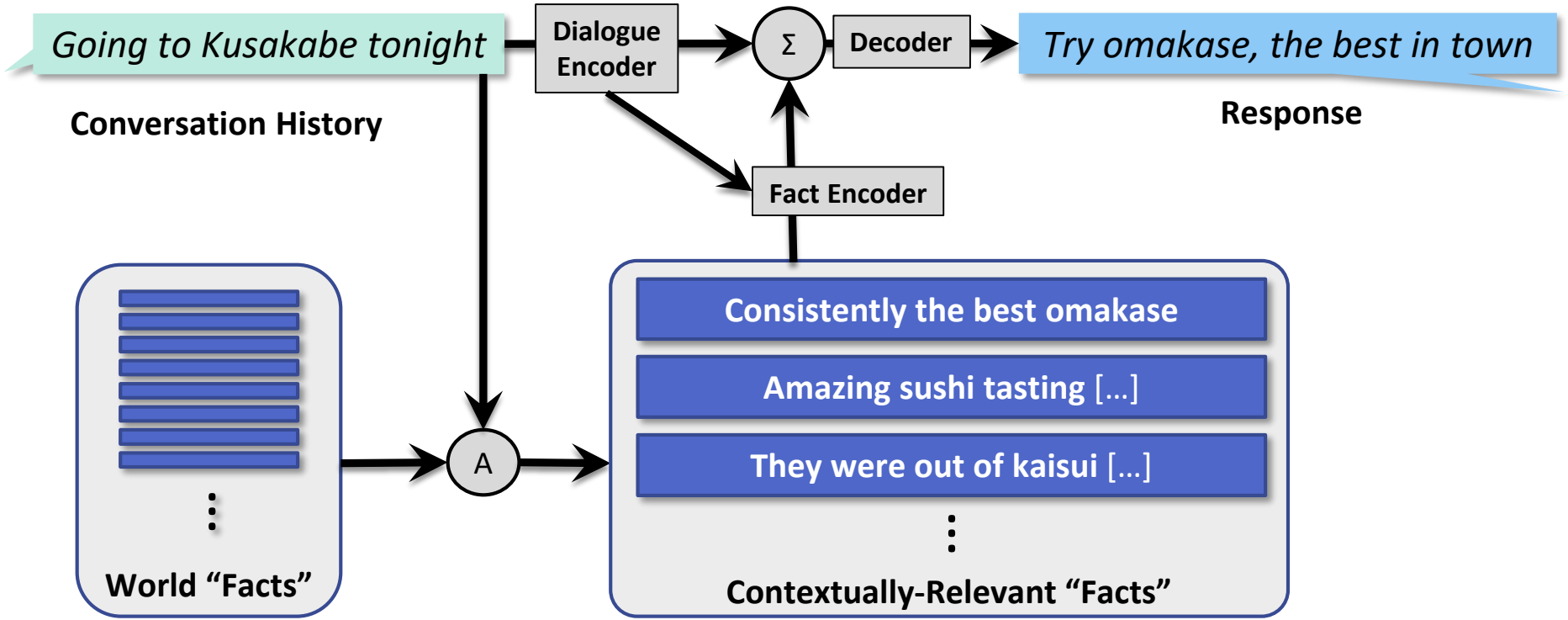
I know, I dislike rain too.
What about a day trip to eastern Washington?

Any recommendation?



Try Dry Falls, it's spectacular!

Knowledge-Grounded Responses (Ghazvininejad et al., 2017)

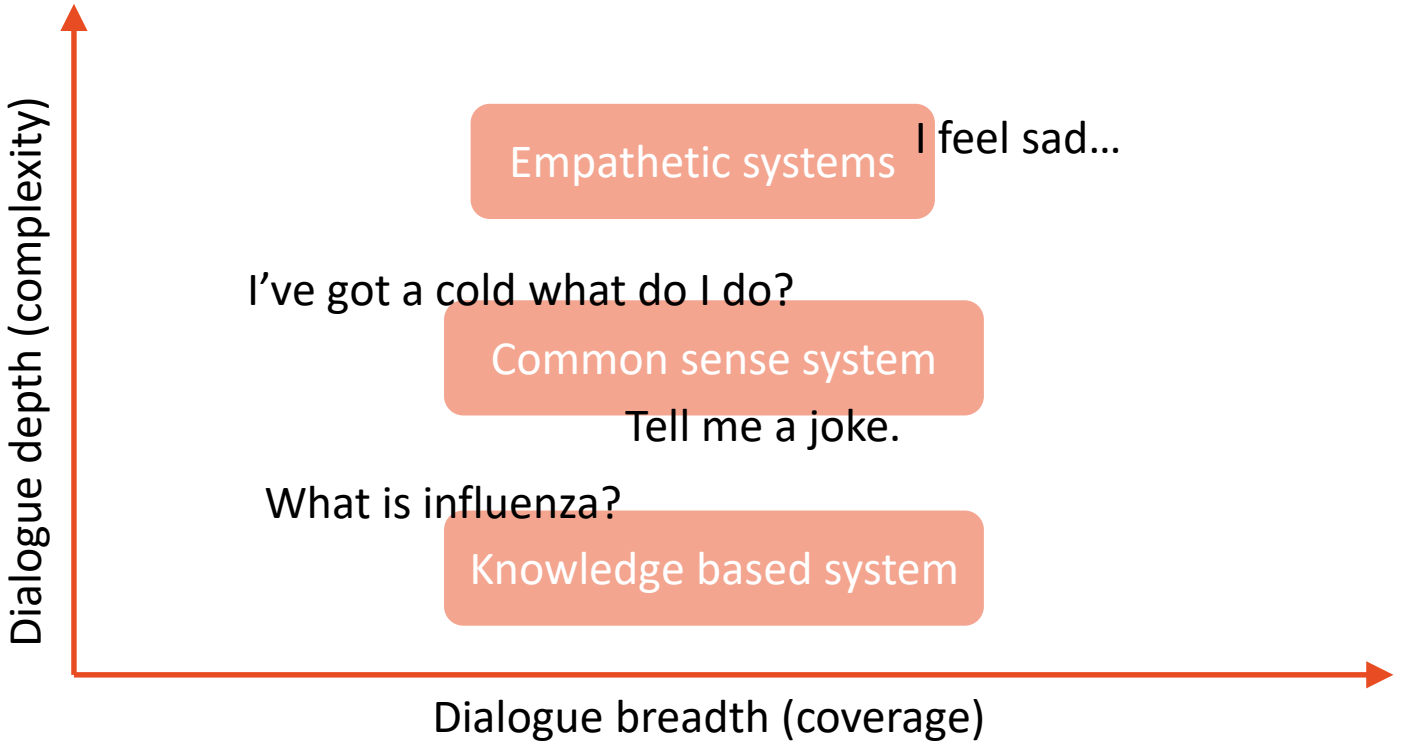


Conversational Agents

Chit-Chat

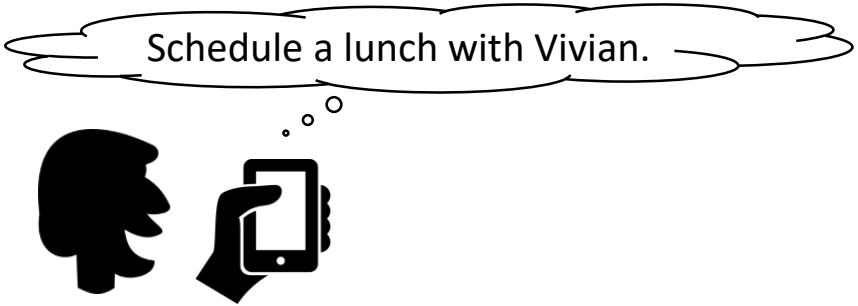
Task-Oriented

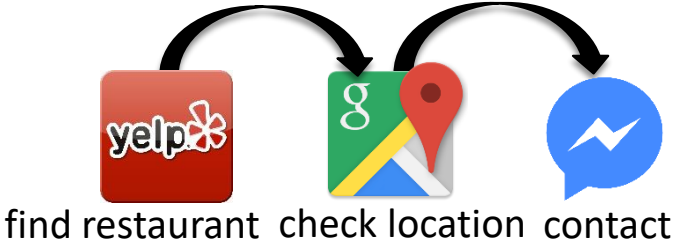
Evolution Roadmap

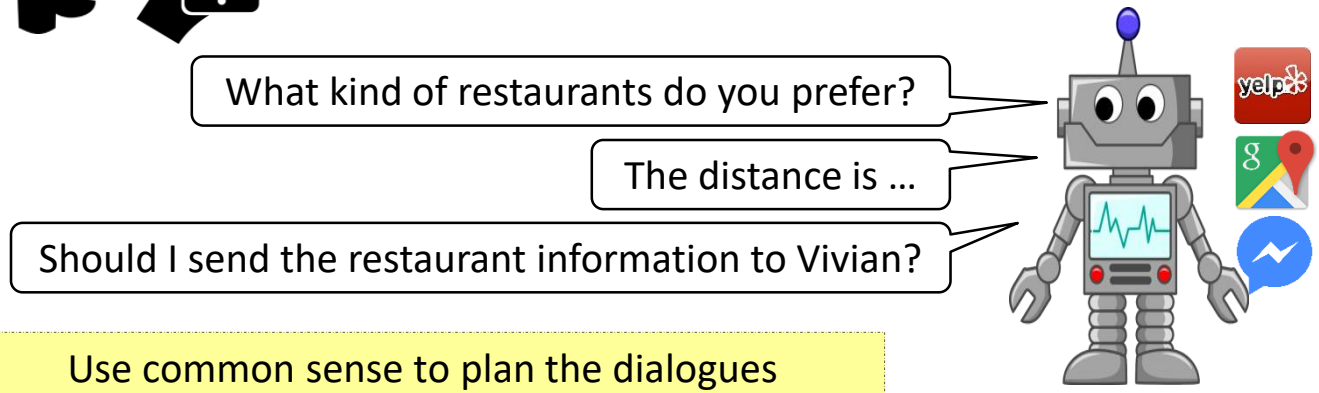


High-Level Intention Learning (Sun et al., 2016; Sun et al., 2016)

- High-level intention may span several domains



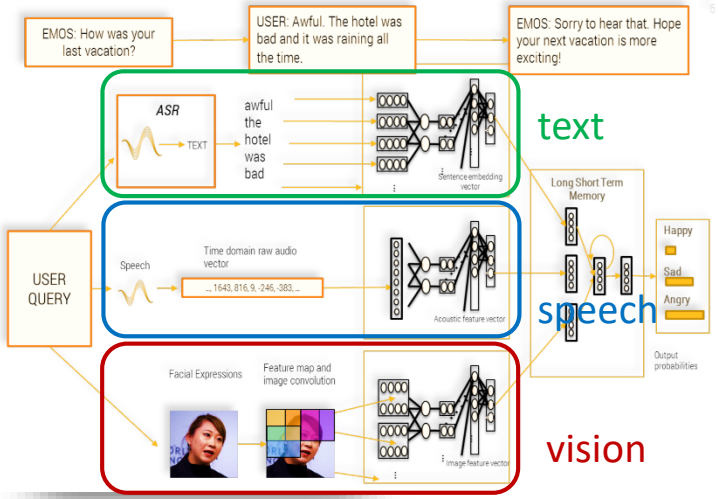




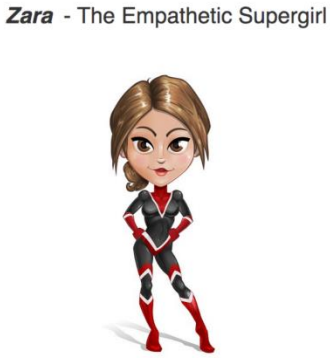
Use common sense to plan the dialogues

Empathy in Dialogue System (Fung et al., 2016)

- Embed an empathy module
  - Recognize emotion using multimodality
  - Generate emotion-aware responses



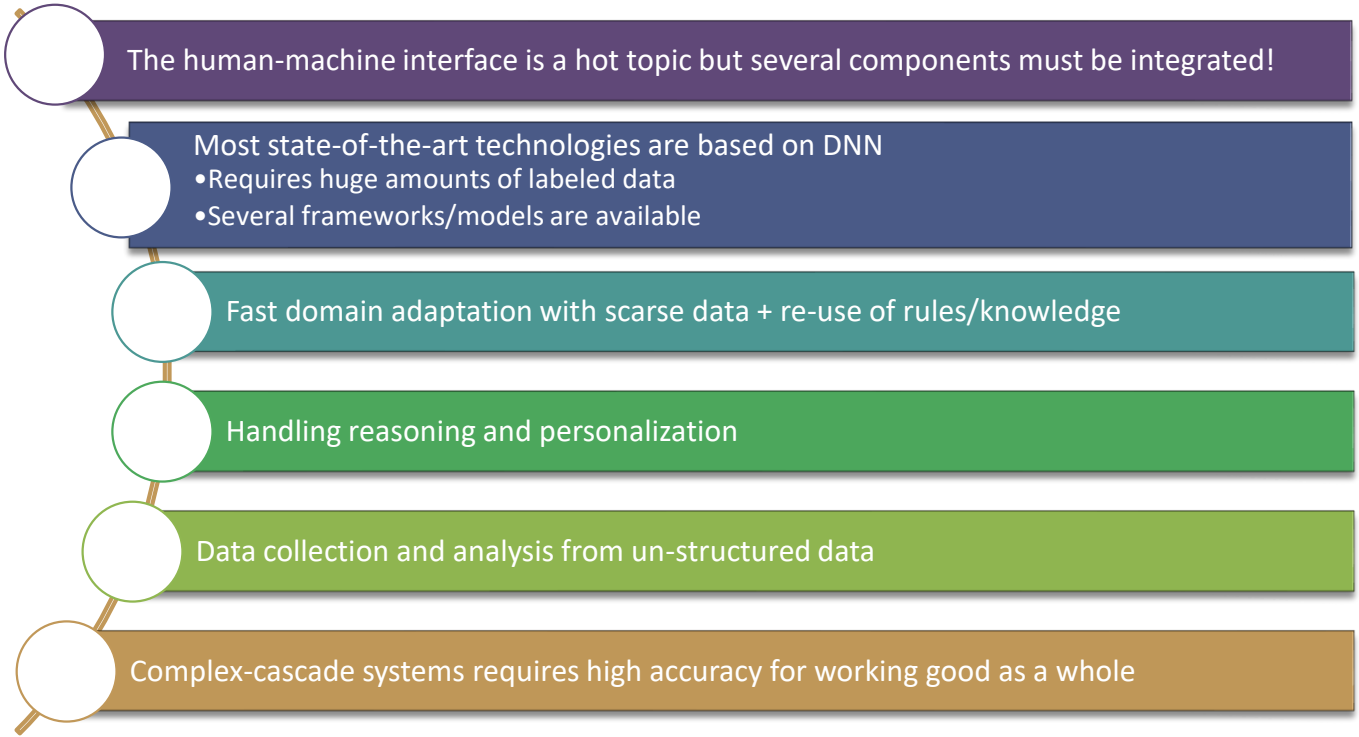
Emotion Recognizer




```
Face recognition output                                    (index):1728
  {                                                        (index):1729
    "recognition": "Race: Asian Confidence: 65.42750000000001 Smiling:
3.95896 Gender: Female Confidence: 88.9369",
    "race": "Asian",
    "race_confidence": "65.42750000000001",
    "smiling": "3.95896",
    "gender": "Female",
    "gender_confidence": "88.9369"
  }
```

Challenges and Conclusions

Challenge Summary



Her (2013)

What can machines achieve now or in the future?

Thanks all people for providing materials for this presentation!

Thanks for Your Attention!

Q & A

Yun-Nung (Vivian) Chen
Assistant Professor
National Taiwan University
y.v.chen@ieee.org / http://vivianchen.idv.tw