



# Open-Domain Neural Dialogue Systems

YUN-NUNG (VIVIAN) CHEN 陳縕儂  
[HTTP://VIVIANCHEN.IDV.TW](http://vivianchen.idv.tw)



國立臺灣大學  
National Taiwan University



# Iron Man (2008)

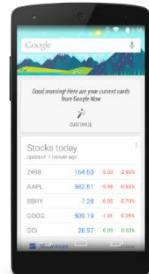
What can machines achieve now or in the future?

# Language Empowering Intelligent Assistants

3

Material: <http://opendialogue.miulab.tw>

Apple Siri (2011)

Google Now (2012)  
Google Assistant (2016)

Microsoft Cortana (2014)

[Message Us](#)[Send to Messenger](#)

Amazon Alexa/Echo (2014)

Facebook M &amp; Bot (2015)

Google Home (2016)

Apple HomePod (2017)

# Why and When We Need?

4

Material: <http://opendialogue.miulab.tw>

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

# Why and When We Need?

5

Material: <http://opendialogue.miulab.tw>

“I want to chat”

Turing Test (talk like a human)

“I have a question”

Information consumption

“I need to get this done”

Task completion

“What should I do?”

Decision support

# Why and When We Need?

6

Material: <http://opendialogue.miulab.tw>

“I want to chat”

Turing Test (talk like a human)

“I have a question”

Information consumption

“I need to get this done”

Task completion

“What should I do?”

Decision support

- *What is today's agenda?*
- *Which room is dialogue tutorial in?*
- *What does ISCSLP stand for?*

# Why and When We Need?

7

Material: <http://opendialogue.miulab.tw>

“I want to chat”

Turing Test (talk like a human)

“I have a question”

Information consumption

“I need to get this done”

Task completion

“What should I do?”

Decision support

- *Book me the train ticket from Taoyuan to Taipei*
- *Reserve a table at Din Tai Fung for 5 people, 7PM tonight*
- *Schedule a meeting with Vivian at 10:00 tomorrow*

# Why and When We Need?

8

Material: <http://opendialogue.miulab.tw>

“I want to chat”

Turing Test (talk like a human)

“I have a question”

Information consumption

“I need to get this done”

Task completion

“What should I do?”

Decision support

- *Is this bubble tea worth to try?*
- *Is the ISCSLP conference good to attend?*

# Why and When We Need?

9

Material: <http://opendialogue.miulab.tw>

“I want to chat”

Turing Test (talk like a human)

Social Chit-Chat

“I have a question”

Information consumption

“I need to get this done”

Task completion

“What should I do?”

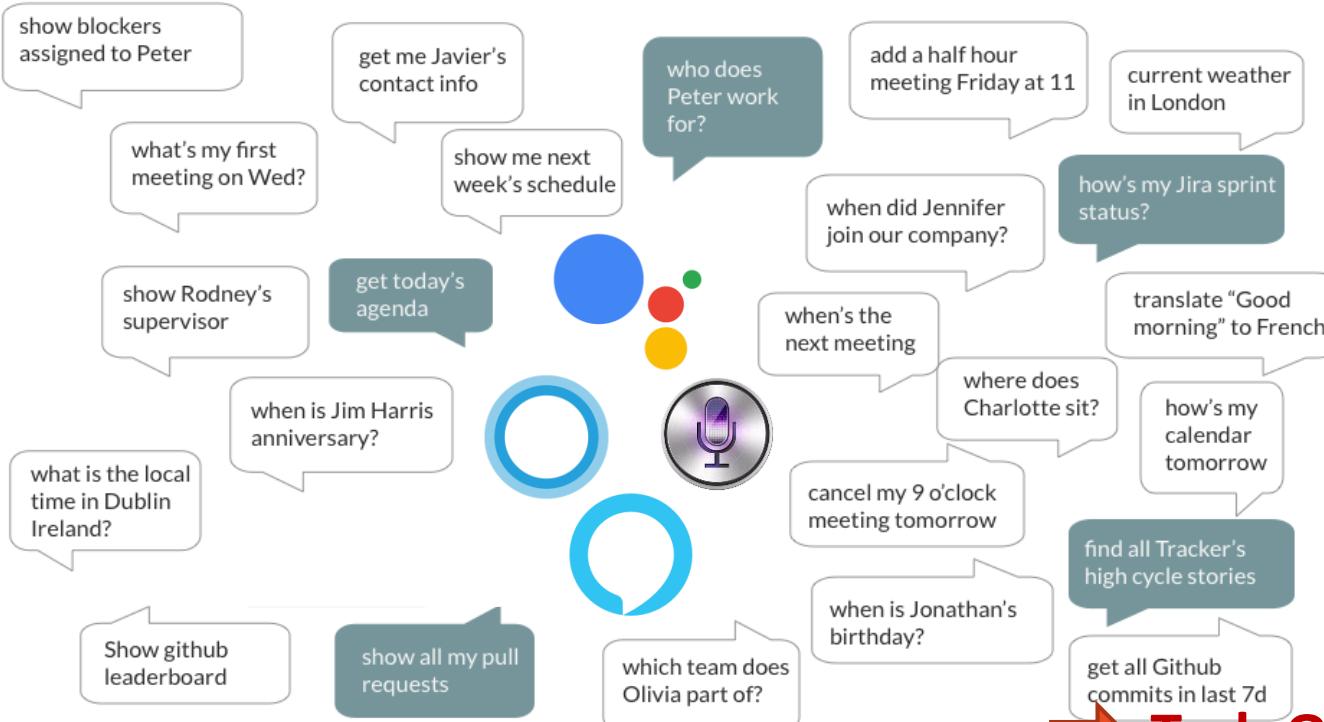
Decision support

} Task-Oriented Dialogues

# Intelligent Assistants

10

Material: <http://opendialogue.miulab.tw>



**Task-Oriented**

# Conversational Agents

11

Material: <http://opendialogue.miulab.tw>

Chit-Chat



Task-Oriented



# Task-Oriented Dialogue Systems



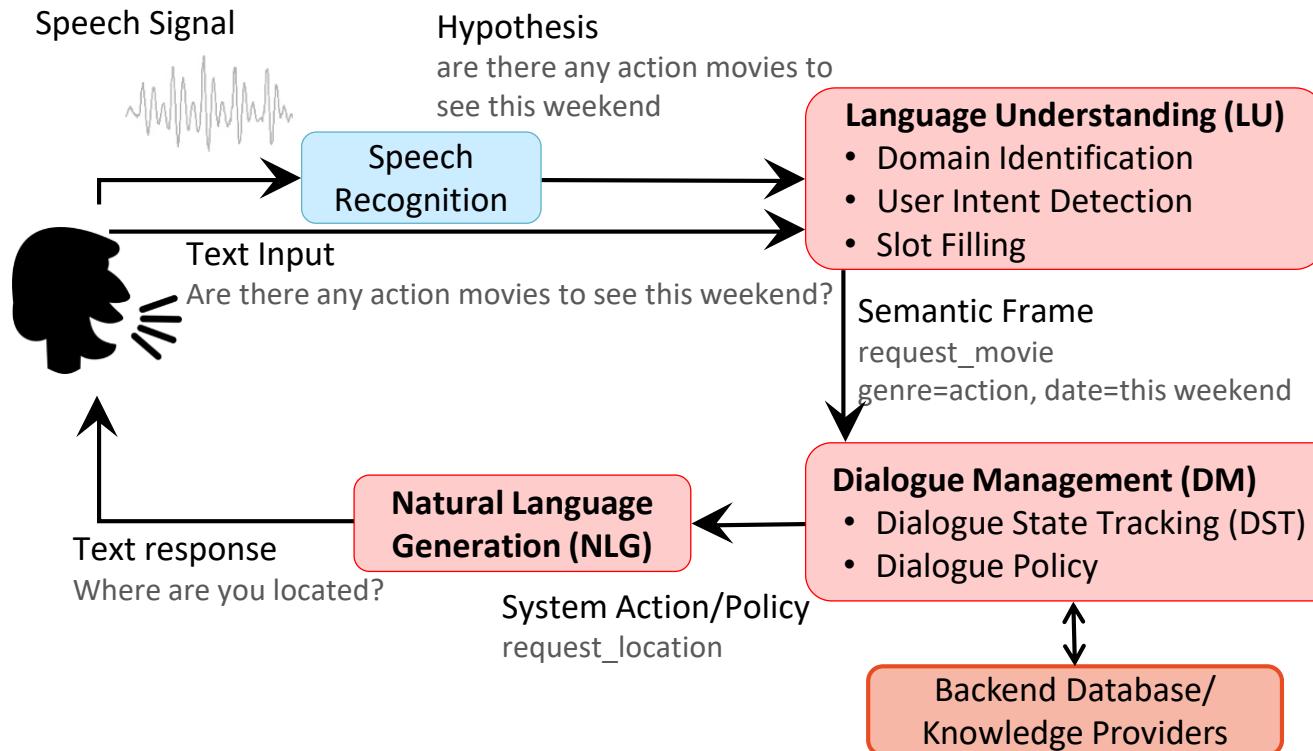
JARVIS – Iron Man's Personal Assistant



Baymax – Personal Healthcare Companion

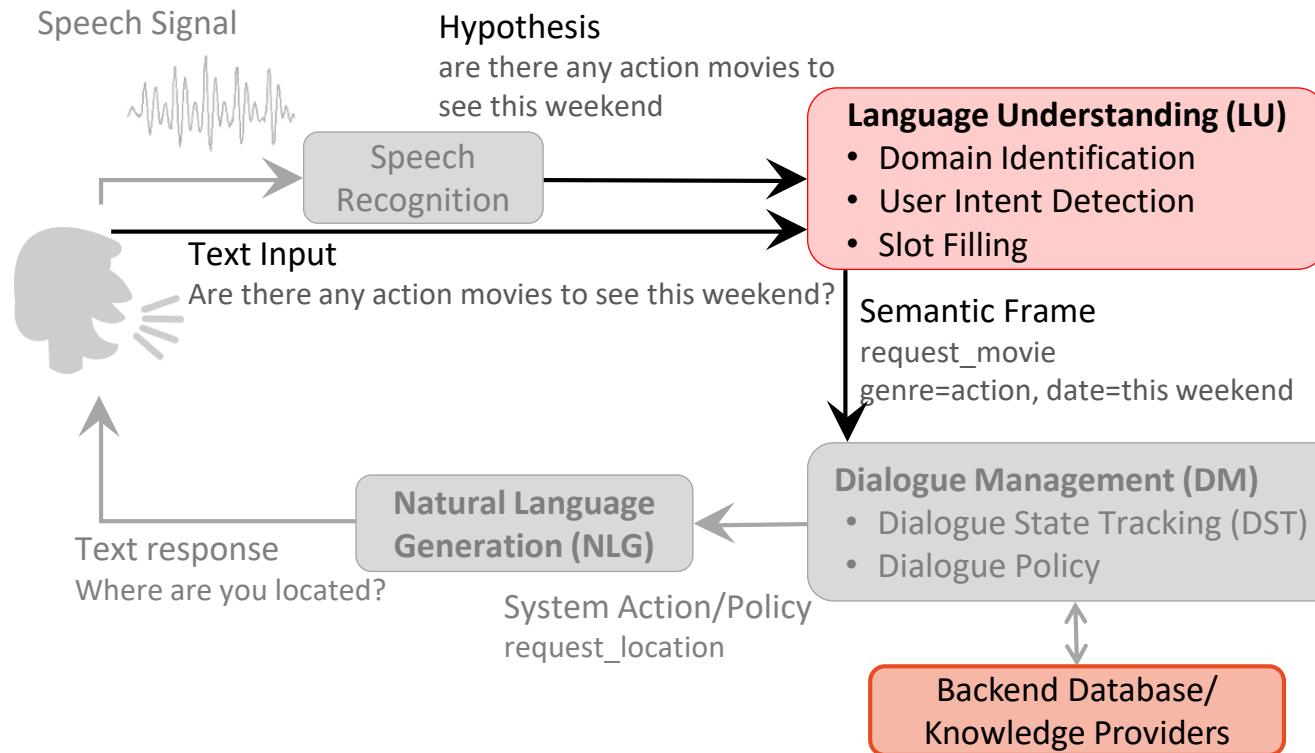
# Task-Oriented Dialogue Systems ([Young, 2000](#))

13

Material: <http://opendialogue.miulab.tw>

# Task-Oriented Dialogue Systems ([Young, 2000](#))

14

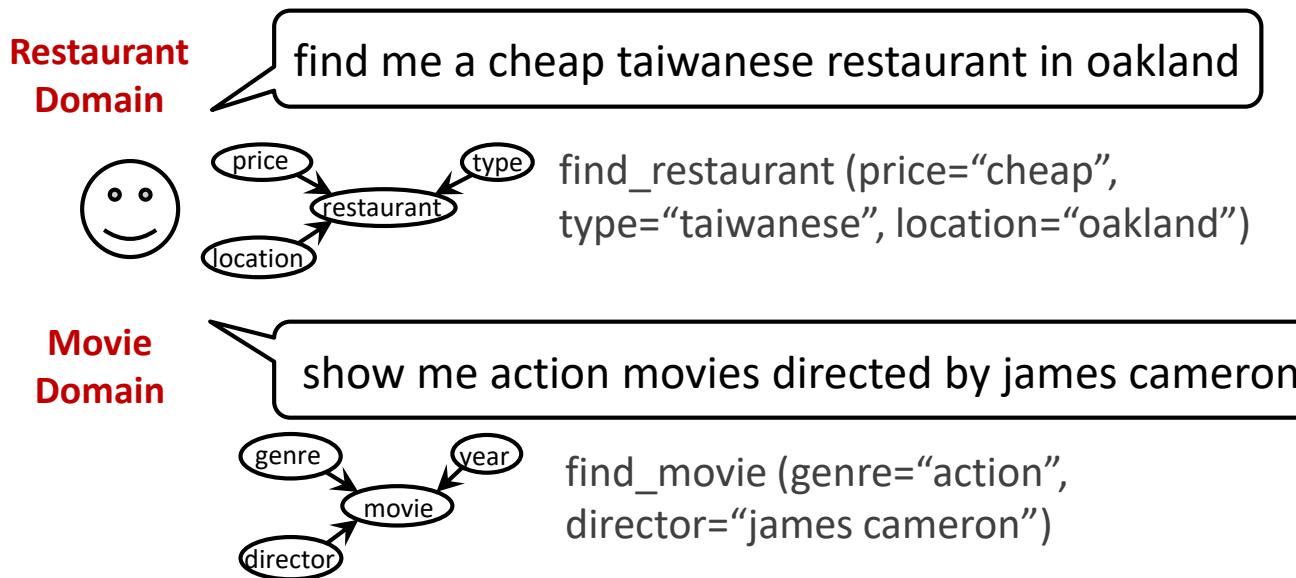
Material: <http://opendialogue.miulab.tw>

# Semantic Frame Representation

15

Material: <http://opendialogue.miulab.tw>

- Requires a domain ontology: early connection to **backend**
- Contains **core concept (intent, a set of slots with fillers)**

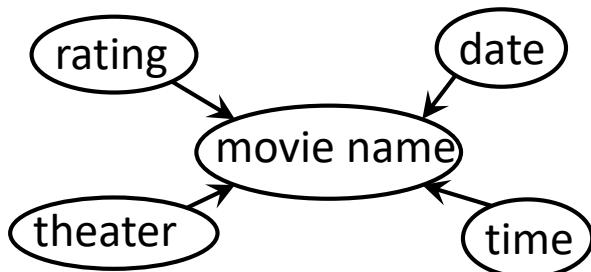


# Backend Database / Ontology

16

Material: <http://opendialogue.miulab.tw>

- Domain-specific table
  - ▣ Target and attributes
- Functionality
  - ▣ Information access: find specific entries
  - ▣ Task completion: find the row that satisfies the constraints

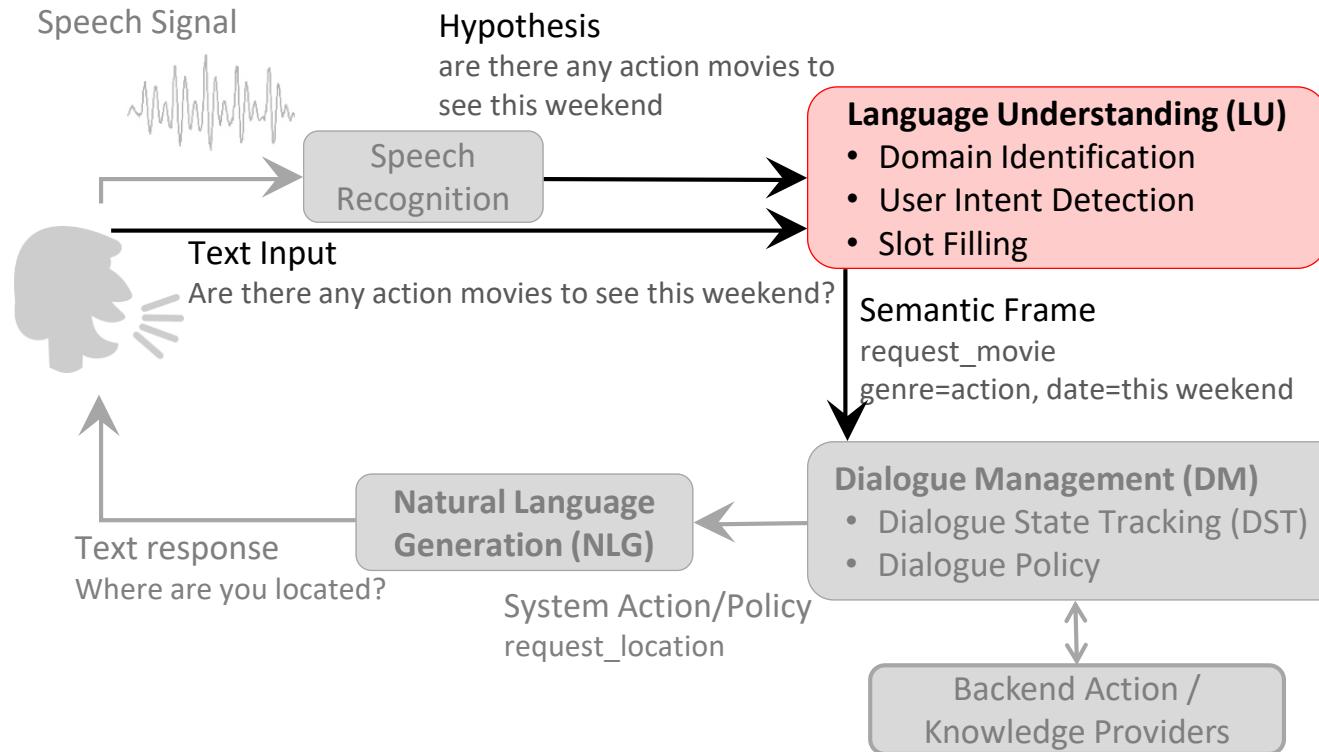


→

Movie Name	Theater	Rating	Date	Time
Iron Man Last	Taipei A1	8.5	2018/10/31	09:00
Iron Man Last	Taipei A1	8.5	2018/10/31	09:25
Iron Man Last	Taipei A1	8.5	2018/10/31	10:15
Iron Man Last	Taipei A1	8.5	2018/10/31	10:40

# Task-Oriented Dialogue Systems ([Young, 2000](http://opendialogue.miulab.tw))

17

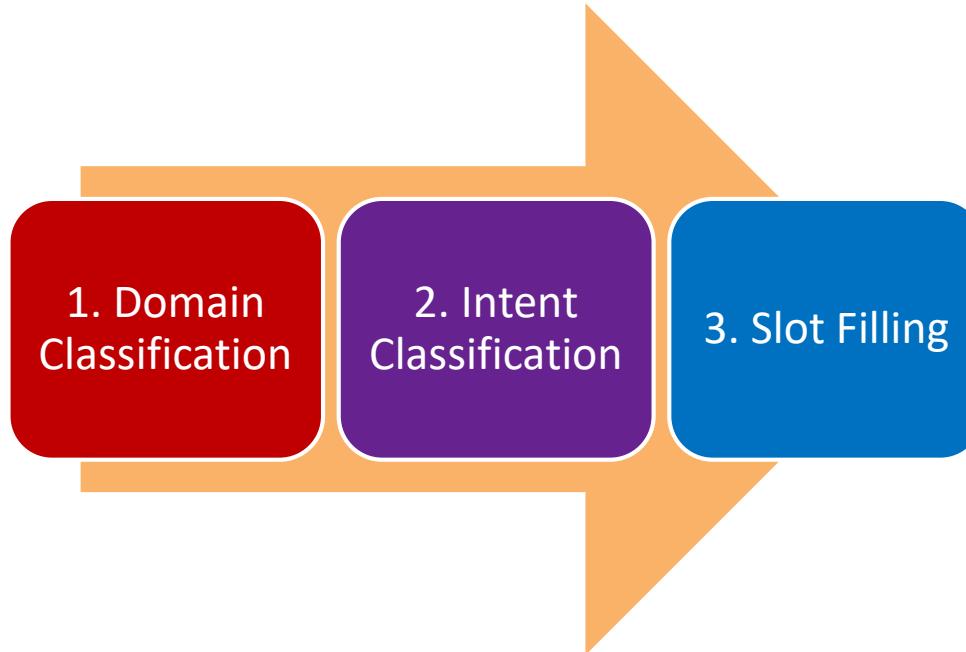
Material: <http://opendialogue.miulab.tw>

# Language Understanding (LU)

18

Material: <http://opendialogue.miulab.tw>

## □ Pipelined



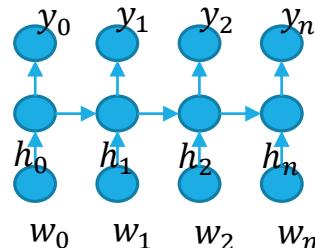
# RNN for Slot Tagging – I

Material: <http://opendialogue.miulab.tw>

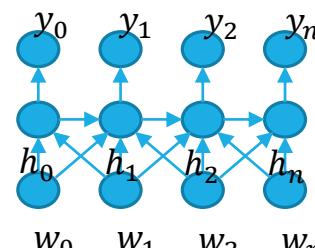
19

## □ Variations:

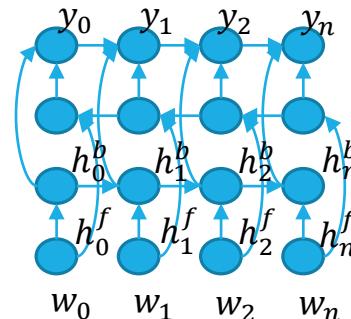
- a. RNNs with LSTM cells
- b. Input, sliding window of n-grams
- c. Bi-directional LSTMs



(a) LSTM



(b) LSTM-LA



(c) bLSTM

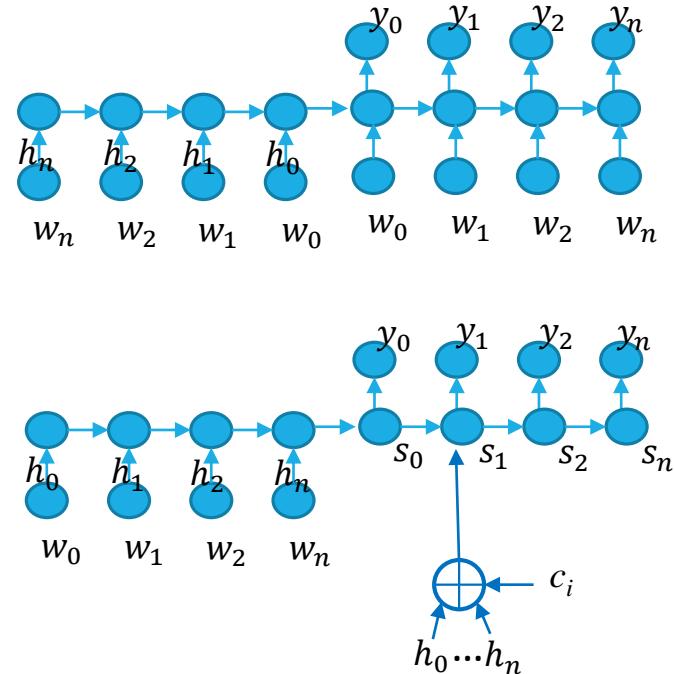
# RNN for Slot Tagging – II

([Kurata et al., 2016](#); Simonnet et al., 2015)

20

Material: <http://opendialogue.miulab.tw>

- Encoder-decoder networks
  - ▣ Leverages sentence level information
  
- Attention-based encoder-decoder
  - ▣ Use attention (as in MT) in the encoder-decoder network

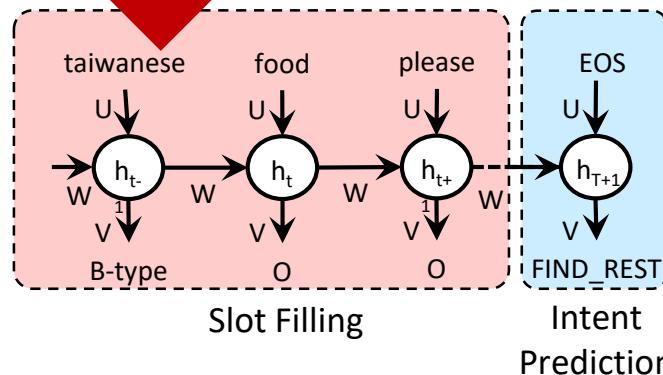


# Joint Semantic Frame Parsing

21

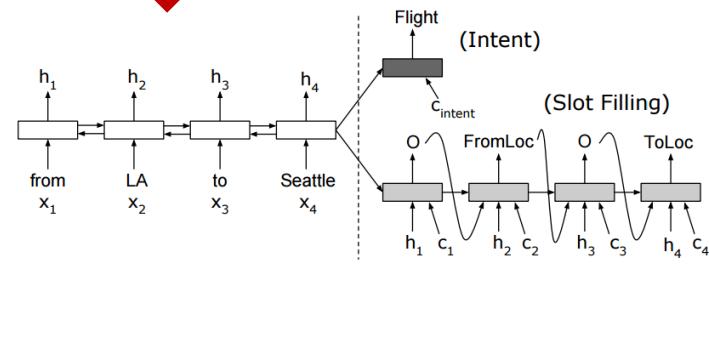
Sequence-based  
(Hakkani-Tur et al., 2016)

- 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

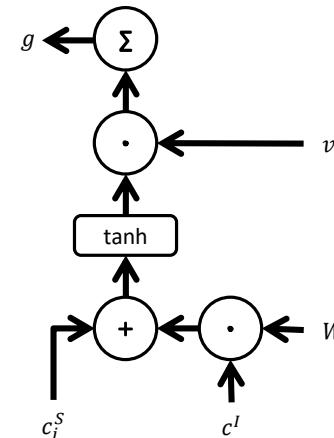
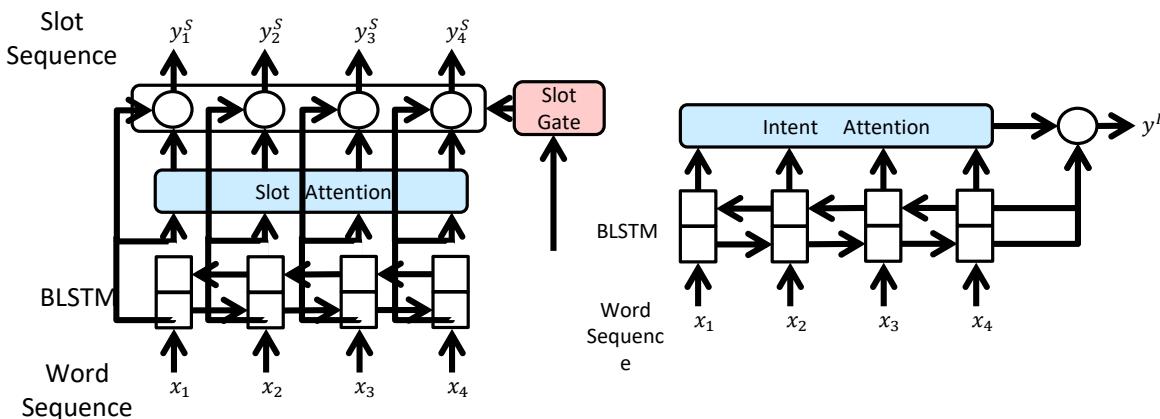
22

Material: <http://opendialogue.miulab.tw>

	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)

23

Material: <http://opendialogue.miulab.tw>

□ Slot Gate 
$$g = \sum v \cdot \tanh(c_i^S + W \cdot c^I)$$

$W^S$ : matrix for output layer  
 $b^S$  : bias for output layer

$c_i^S$ : slot context vector  
 $c^I$  : intent context vector

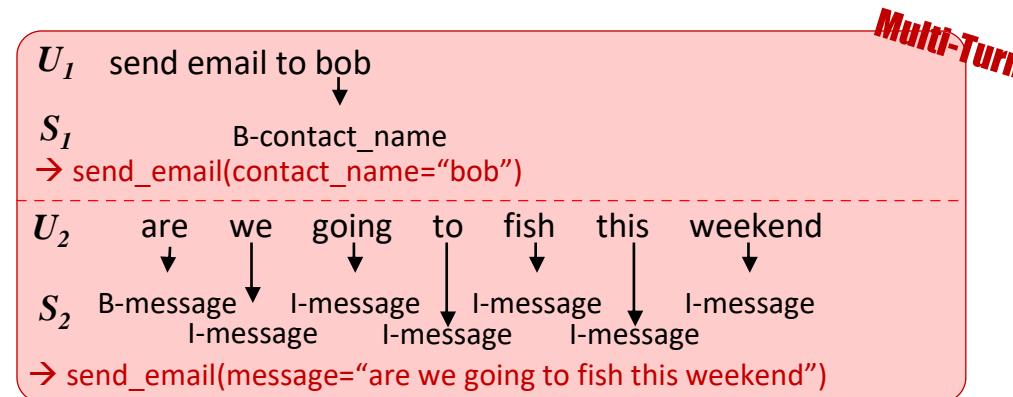
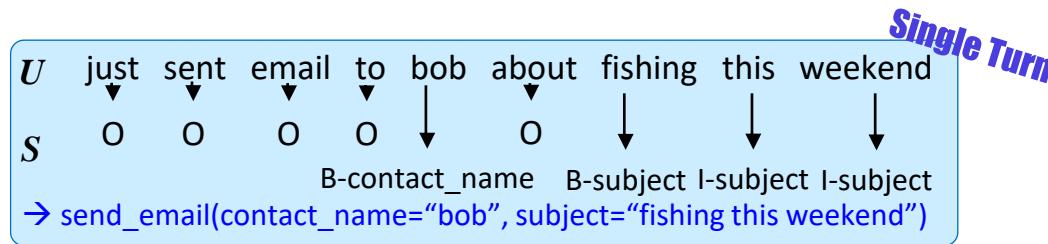
□ Slot Prediction  $y_i^S = \text{softmax}(W^S(h_i + c_i^S) + b^S) \rightarrow y_i^S = \text{softmax}(W^S(h_i + g \cdot c_i^S) + b^S)$

$g$  will be larger if slot and intent are better related

$W$ : trainable matrix  
 $v$  : trainable vector  
 $g$  : scalar gate value

# Contextual Language Understanding

24

Material: <http://opendialogue.miulab.tw>

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

25

Material: <http://opendialogue.miulab.tw>

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 ?"

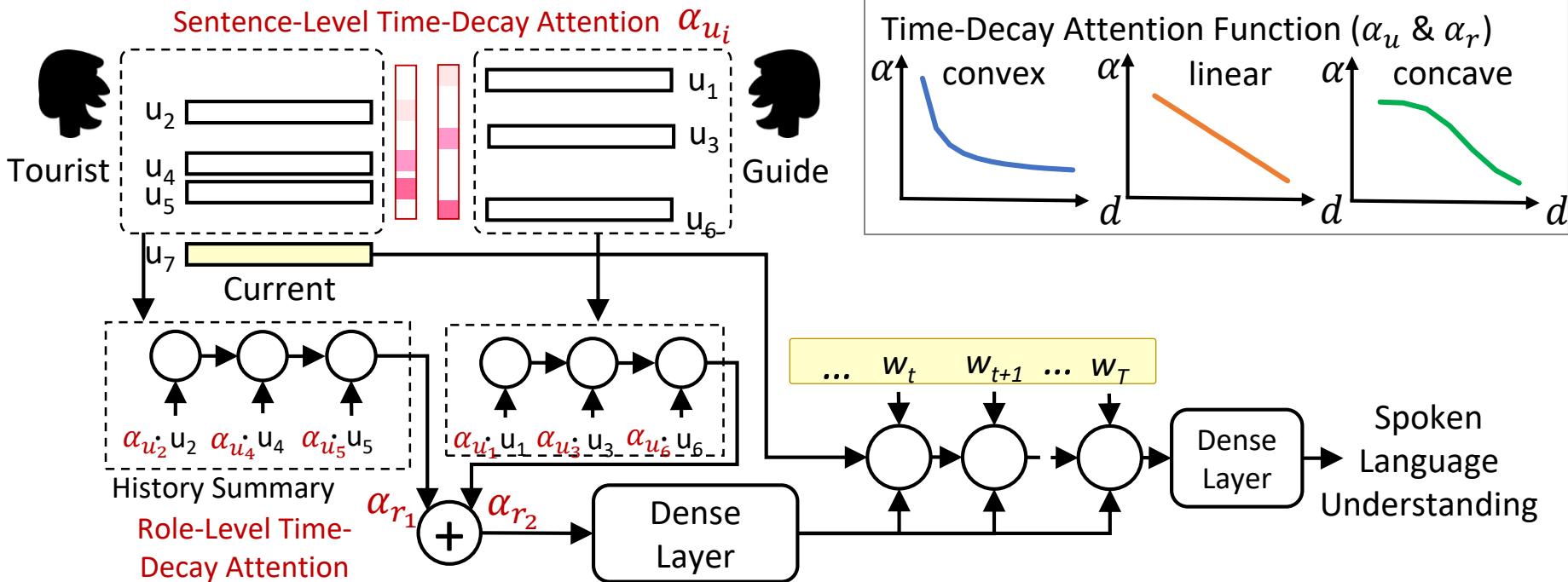
U: "Any time on saturday is fine"

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

U: "Let's do 5:40"

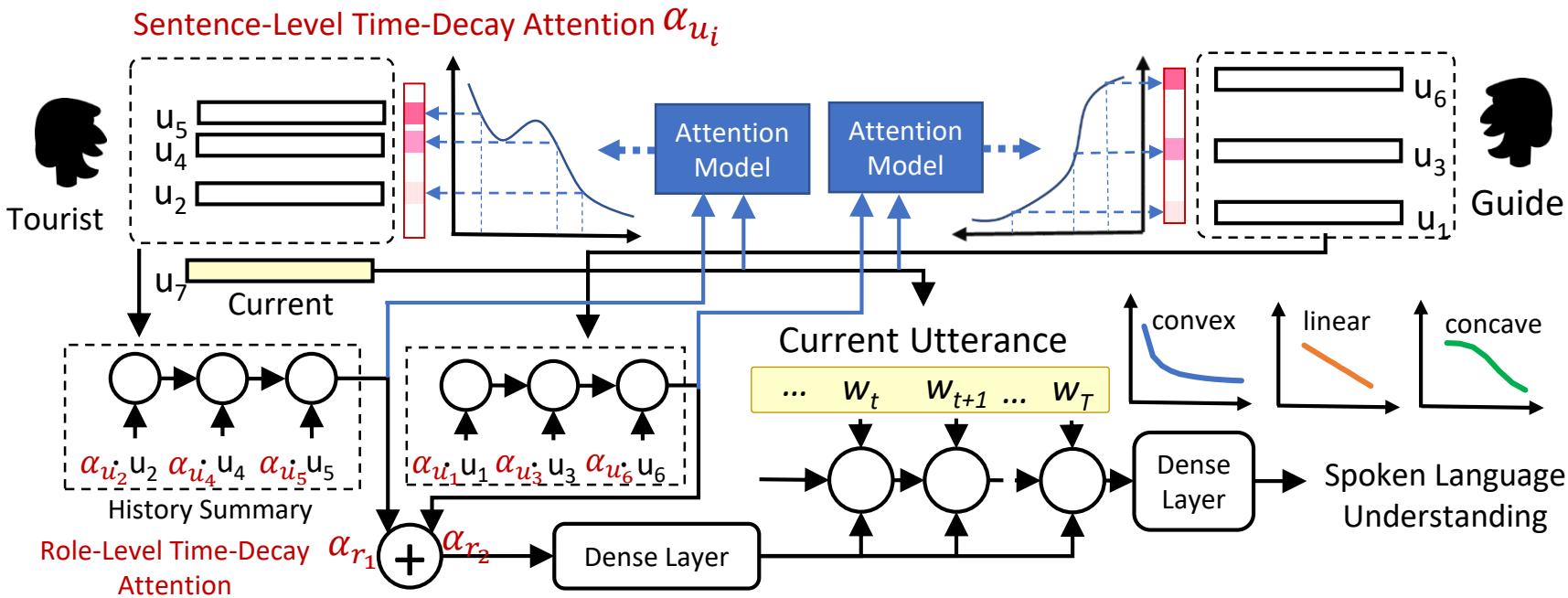
# Role-Based & Time-Aware Attention (Su et al., 2018)

26

Material: <http://opendialogue.miulab.tw>

# Context-Sensitive Time-Decay Attention (Su et al., 2018)

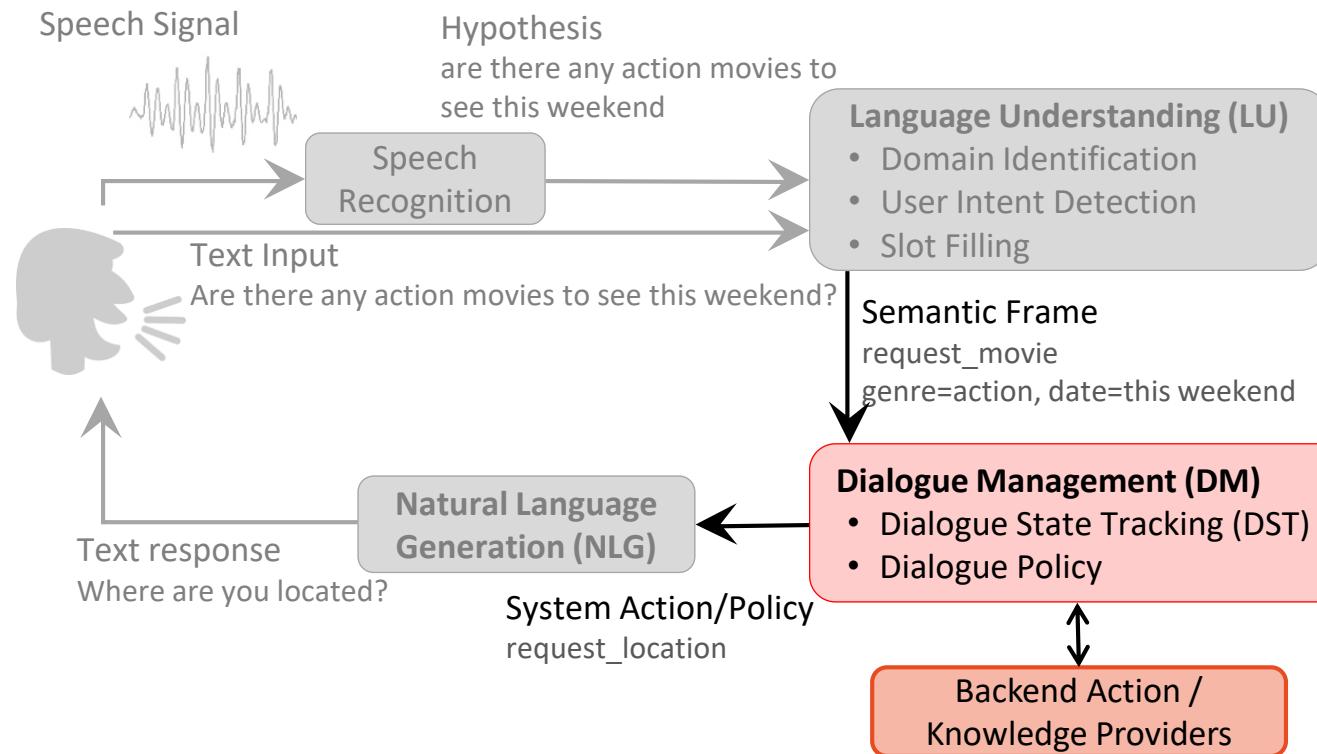
27

Material: <http://opendialogue.miulab.tw>

Time-decay attention significantly improves the understanding results

# Task-Oriented Dialogue Systems (Young, 2000)

28

Material: <http://opendialogue.miulab.tw>

# Dialogue State Tracking

29

Material: <http://opendialogue.miulab.tw>

Hello, how may I help you?

I'm looking for a Thai restaurant.

request (restaurant; foodtype=Thai)

What part of town do you have in mind?

Something in the centre.

inform (area=centre)

Bangkok city is a nice place, it is in the centre of town and it serves Thai food.

What's the address?

request (address)

Bangkok city is a nice place, their address is 24 Green street.

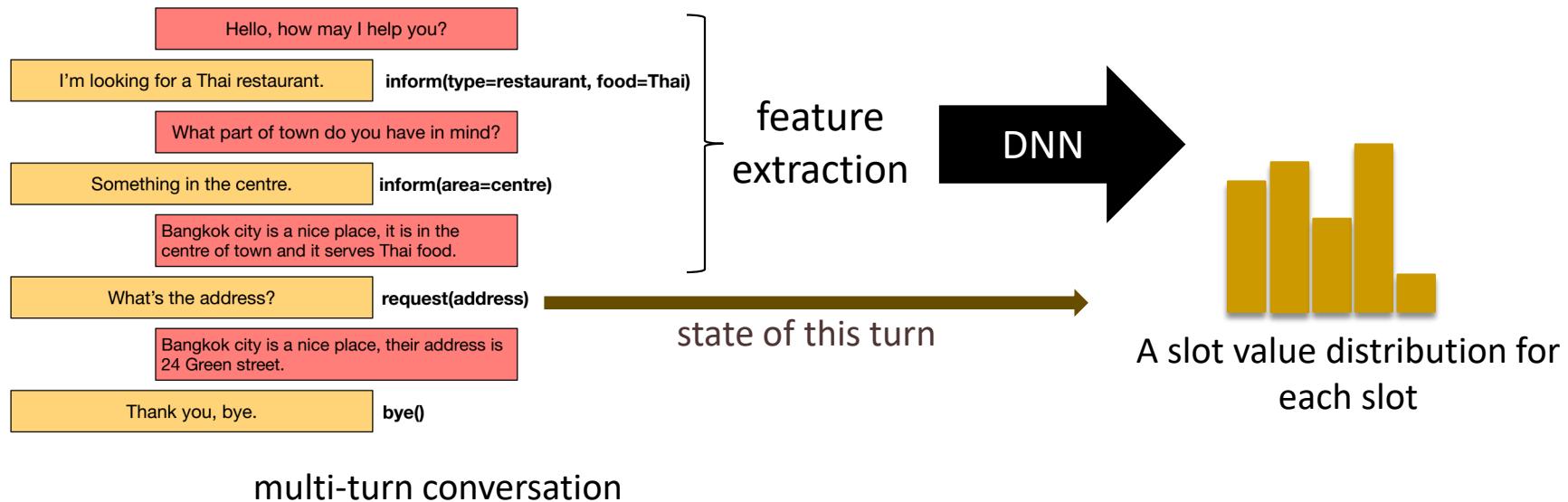
Thank you, bye.

bye ()

# DNN for DST

30

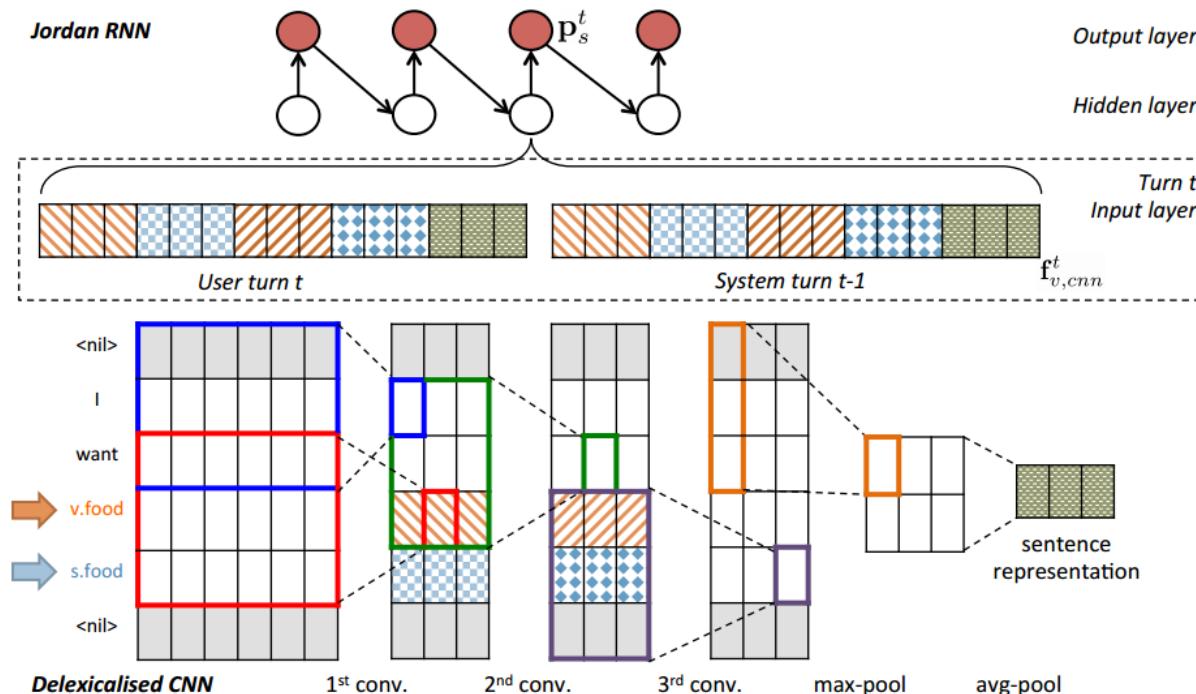
Material: <http://opendialogue.miulab.tw>



# RNN-CNN DST (Wen et al., 2016)

31

Material: <http://opendialogue.miulab.tw>



(Figure from Wen et al, 2016)

# Dialogue Policy Optimization

32

Material: <http://opendialogue.miulab.tw>

Hello, how may I help you?

greeting ()

I'm looking for a Thai restaurant.

request (restaurant; foodtype=Thai)

What part of town do you have in mind?

request (area)

Something in the centre.

inform (area=centre)

Bangkok city is a nice place, it is in the centre of town and it serves Thai food.

inform (restaurant=Bangkok  
city, area=centre of town,  
foodtype=Thai)

What's the address?

request (address)

Bangkok city is a nice place, their address is 24 Green street.

inform (address=24 Green street)

Thank you, bye.

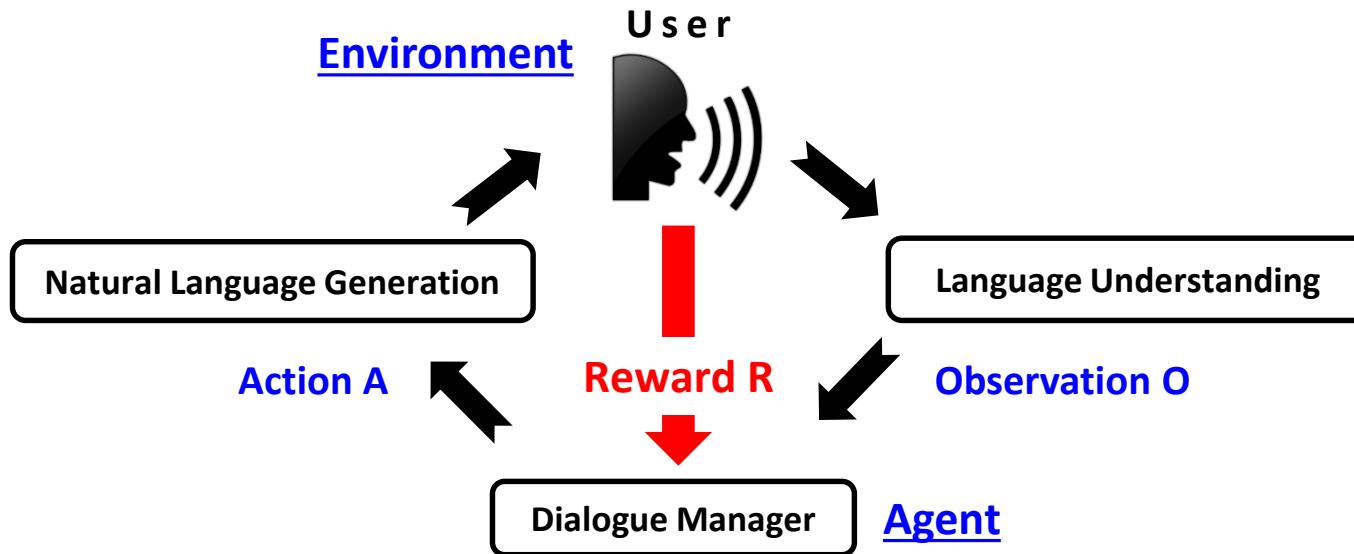
bye ()

# Dialogue Policy Optimization

33

Material: <http://opendialogue.miulab.tw>

- Dialogue management in a RL framework



Select the best action that maximizes the future reward

# Reward for RL $\approx$ Evaluation for System

34

Material: <http://opendialogue.miulab.tw>

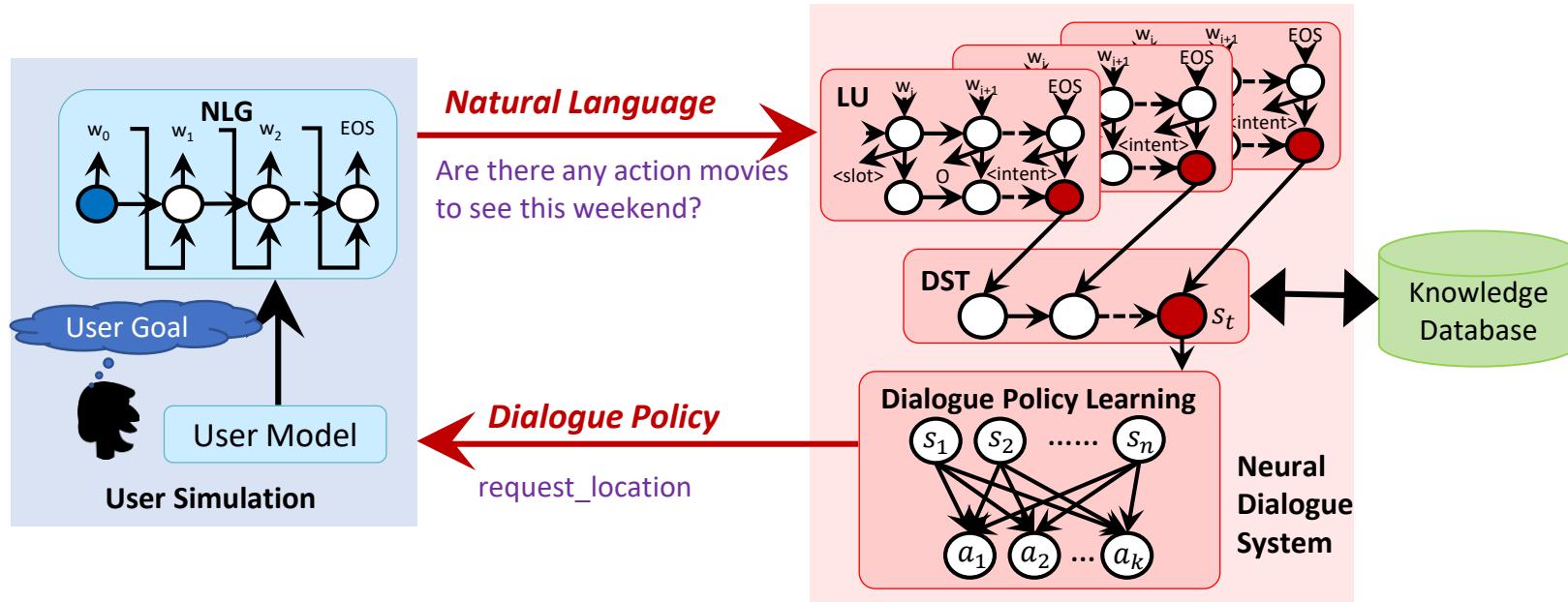
- Dialogue is a special RL task
  - Human involves in **interaction** and **rating (evaluation)** of a dialogue
  - Fully human-in-the-loop framework
- Rating: correctness, appropriateness, and adequacy

- Expert rating	high quality, <b>high cost</b>
- User rating	unreliable quality, <b>medium cost</b>
- Objective rating	Check desired aspects, <b>low cost</b>

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

35

Material: <http://opendialogue.miulab.tw>



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

# Reinforcement Learning Agent ([Li et al., 2017](#))

36

Material: <http://opendialogue.miulab.tw>

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!

**Issue 1: 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**

# Reinforcement Learning Agent ([Li et al., 2017](#))

37

Material: <http://opendialogue.miulab.tw>

User goal – ***Two tickets for "Hail Caesar" tomorrow 7:15PM at AMC Lowes Oak tree 6, Seattle.***

## RULE BASED SYSTEM

User: Do you any ticket for movie 'Hail Caesar' at theater  
AMC Lowes Oak Tree 6?

Agent: What start time do you prefer?

User: **7:15 pm** Tomorrow

Agent: Which city you are looking for?

User: Seattle

Agent: How many tickets do you want?

User: Two

Agent: Okay, I have issued two tickets for movie 'Hail Caeser'  
tomorrow **7:50 pm** at theater AMC Lowes Oak Tree 6,  
Seattle.

User: Oh, no.

Agent: Thanks!

**Issue 2: no principle way to deal  
with LU errors/uncertainty**

## REINFORCEMENT LEARNING SYSTEM

User: Do you any ticket for movie 'Hail Caesar' at theater AMC  
Lowes Oak Tree 6?

Agent: What start time do you prefer?

User: **7:15 pm** Tomorrow

Agent: Which city you are looking for?

User: Seattle

Agent: Okay, I will issue you tickets for movie 'Hail Caeser'  
tomorrow **7:50 pm** at theater AMC Lowes Oak Tree 6, Seattle.

User: Oh, no.

Agent: **It is 7:50 pm, right?**

User: Two tickets for 7:15 pm, please.

Agent: Okay, I will issue you two tickets for movie 'Hail Caeser'  
tomorrow **7:15 pm** at theater AMC Lowes Oak Tree 6, Seattle.

User: Thanks!

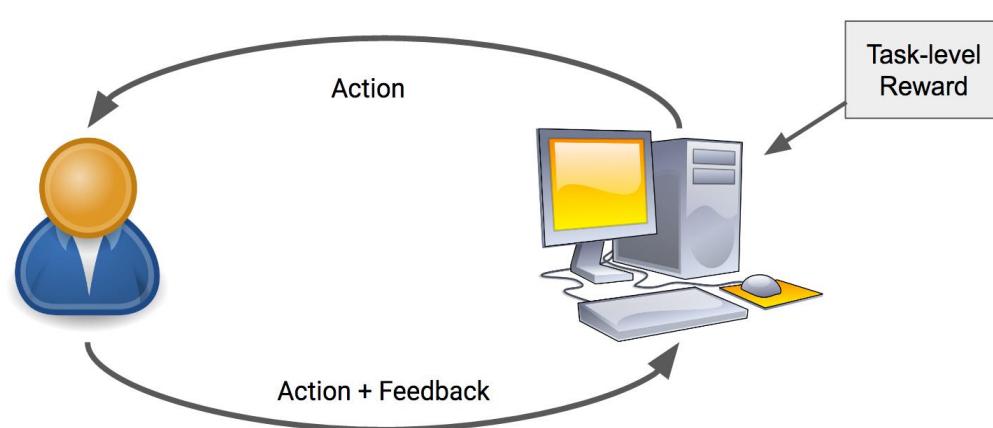
Agent: Thanks!

Automatically learn when and what to confirm for better robustness

# Interactive RL for DM ([Shah et al., 2016](#))

38

Material: <http://opendialogue.miulab.tw>



Immediate  
Feedback

Explicit



Implicit

- Is First Wok highly rated?
  - First Wok, Lucy's and Red Grill are good options.
  - No stupid, I am asking if First Wok is rated at least 3 stars?
- Frustration      Repetition

Use a third agent for providing interactive feedback to the DM

# RL in Dialogue Systems

39

Material: <http://opendialogue.miulab.tw>

1. Sample inefficient, hard to design reward function, local optima...
2. Real users are expensive
3. Discrepancy between real users and simulators

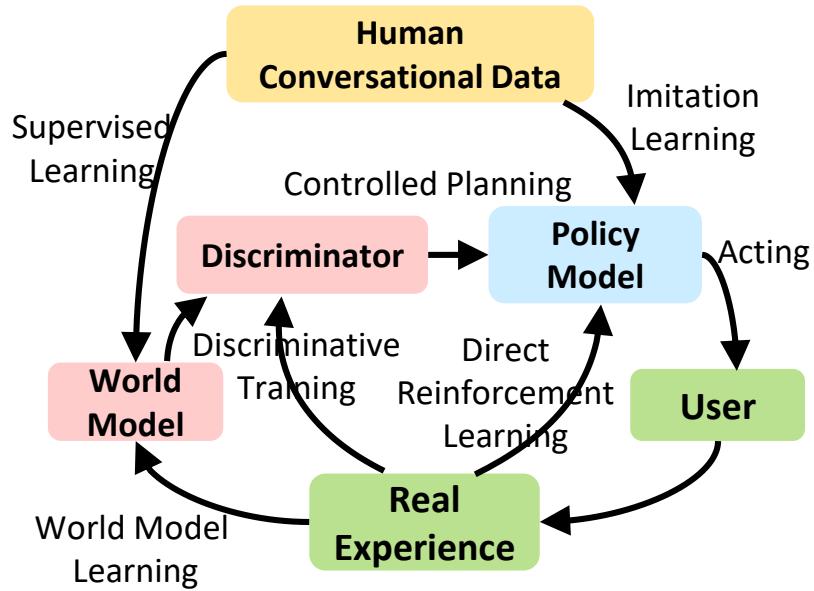
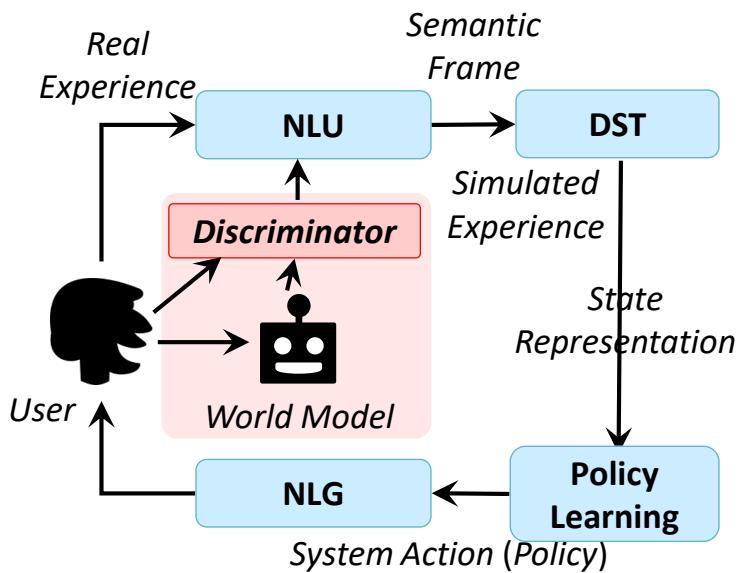
# D3Q: Discriminative Deep Dyna-Q (Su et al., 2018)

40

Material: <http://opendialogue.miulab.tw>

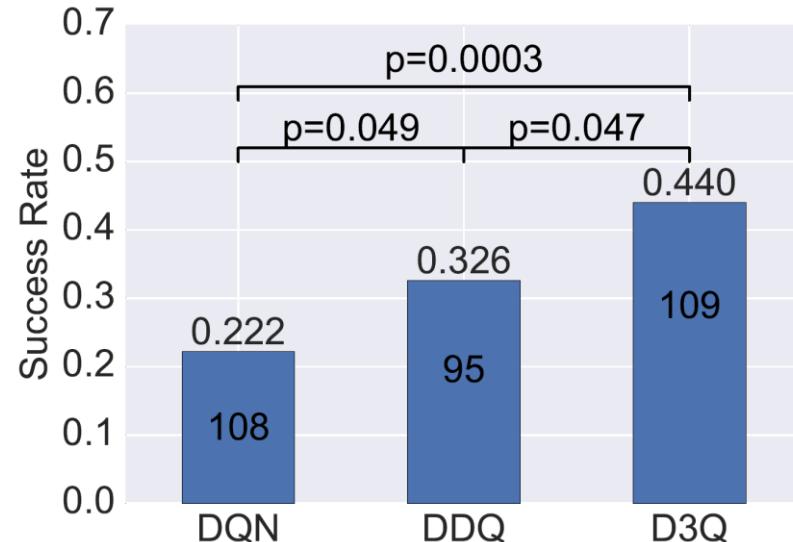
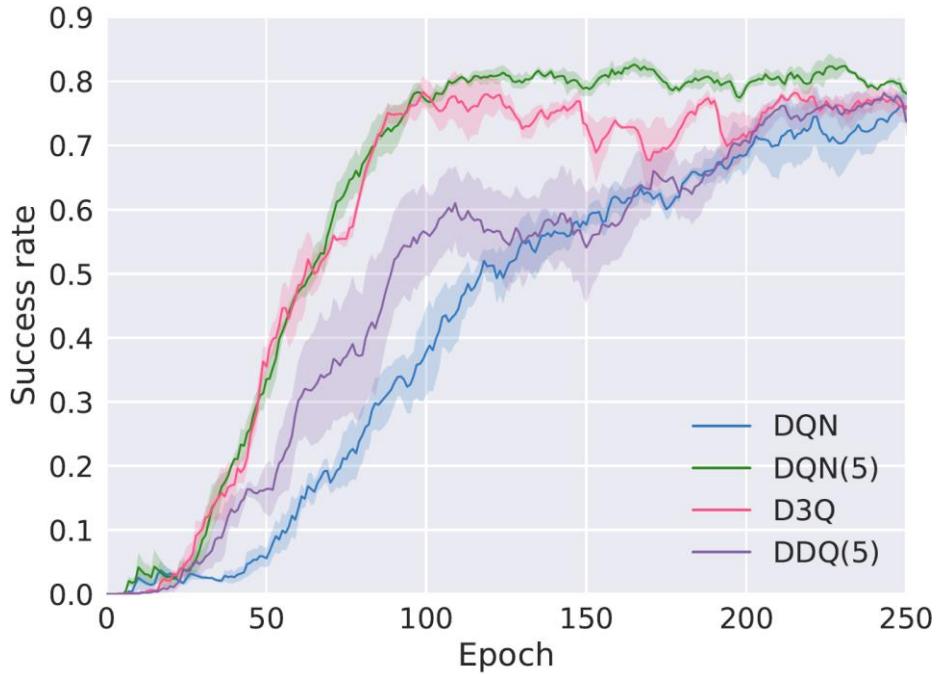
## □ Idea

- learning with real users with planning
- add a *discriminator* to filter out the bad experiences



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

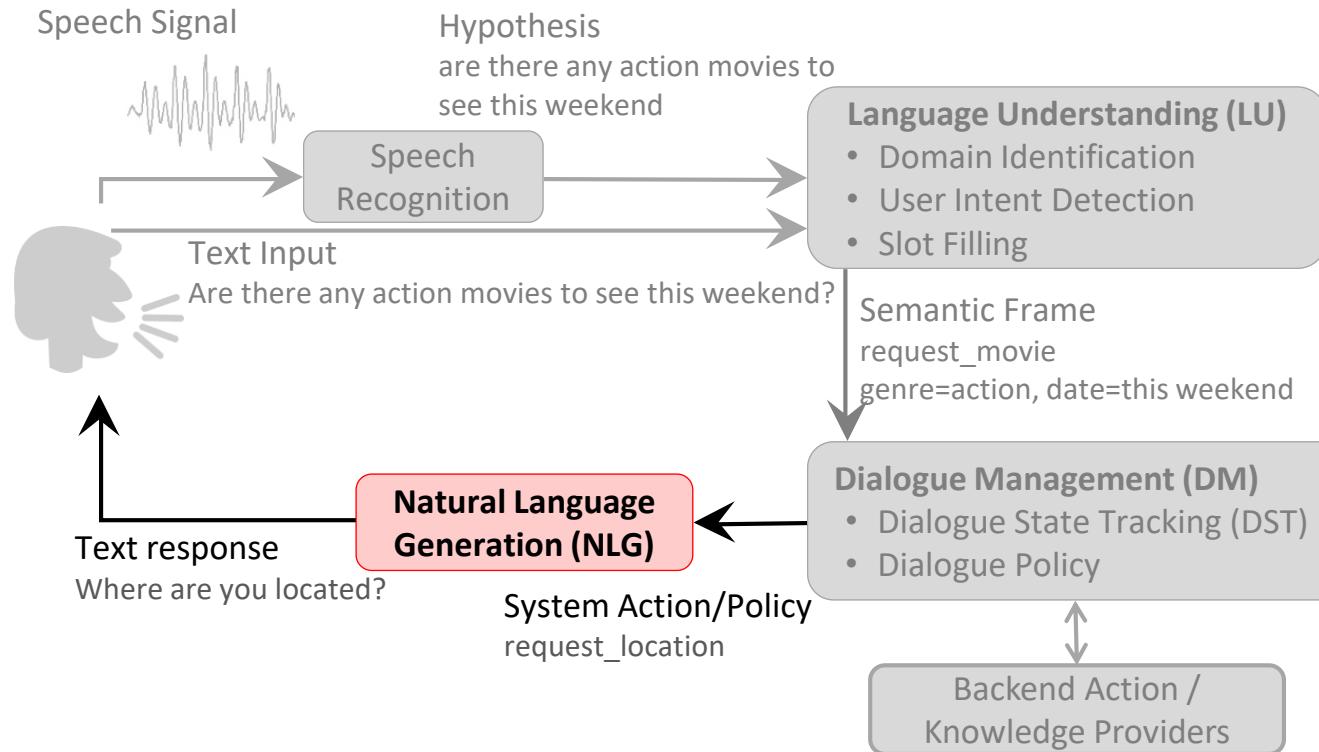
41

Material: <http://opendialogue.miulab.tw>

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

# Task-Oriented Dialogue Systems (Young, 2000)

42

Material: <http://opendialogue.miulab.tw>

# Natural Language Generation (NLG)

43

Material: <http://opendialogue.miulab.tw>

- Mapping dialogue acts into natural language

```
inform(name=Seven_Days, foodtype=Chinese)
```



Seven Days is a nice Chinese restaurant

# Template-Based NLG

44

Material: <http://opendialogue.miulab.tw>

- Define a set of rules to map frames to natural language

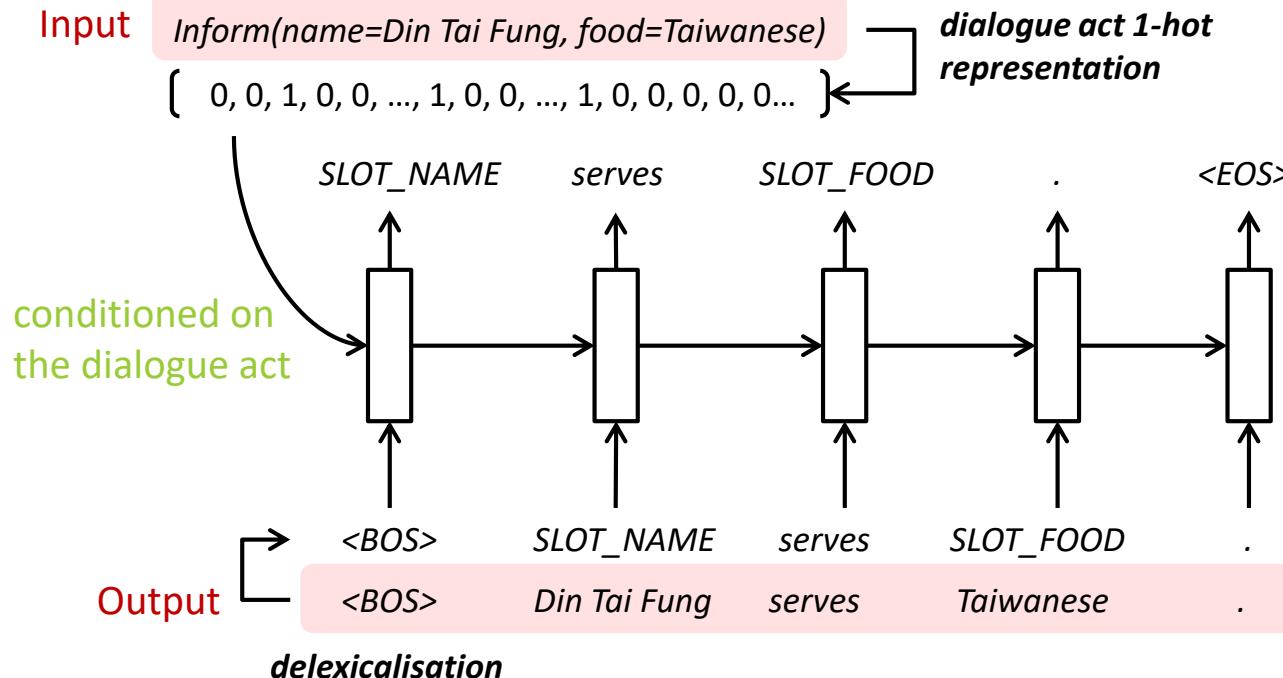
Semantic Frame	Natural Language
confirm()	“Please tell me more about the product your are looking for.”
confirm(area=\$V)	“Do you want somewhere in the \$V?”
confirm(food=\$V)	“Do you want a \$V restaurant?”
confirm(food=\$V,area=\$W)	“Do you want a \$V restaurant in the \$W.”

**Pros:** simple, error-free, easy to control

**Cons:** time-consuming, rigid, poor scalability

# RNN-Based LM NLG (Wen et al., 2015)

45

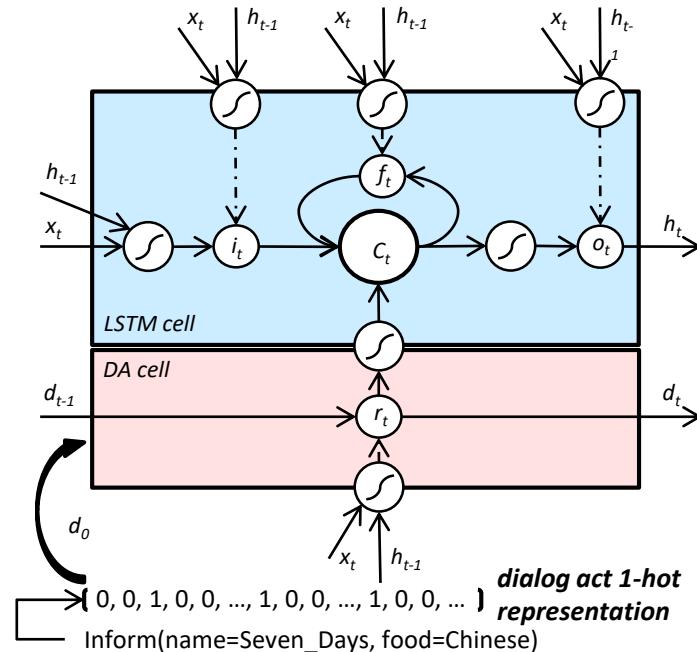
Material: <http://opendialogue.miulab.tw>

# Semantic Conditioned LSTM (Wen et al., 2015)

46

Material: <http://opendialogue.miulab.tw>

- Issue: semantic repetition
  - Din Tai Fung is a great Taiwanese restaurant that serves Taiwanese.
  - Din Tai Fung is a child friendly restaurant, and also allows kids.



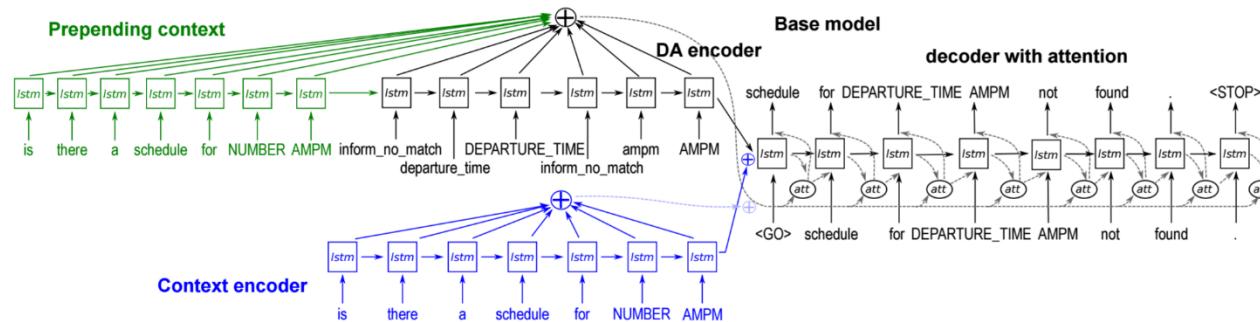
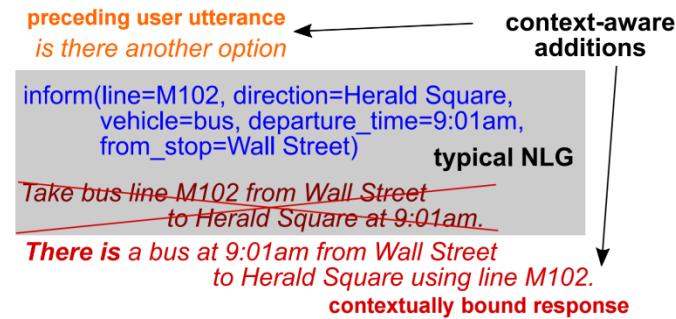
Idea: using gate mechanism to control the generated semantics (dialogue act/slots)

# Contextual NLG ([Dušek and Jurčíček, 2016](#))

47

Material: <http://opendialogue.miulab.tw>

- Goal: adapting users' way of speaking, providing context-aware responses
  - Context encoder
  - Seq2Seq model

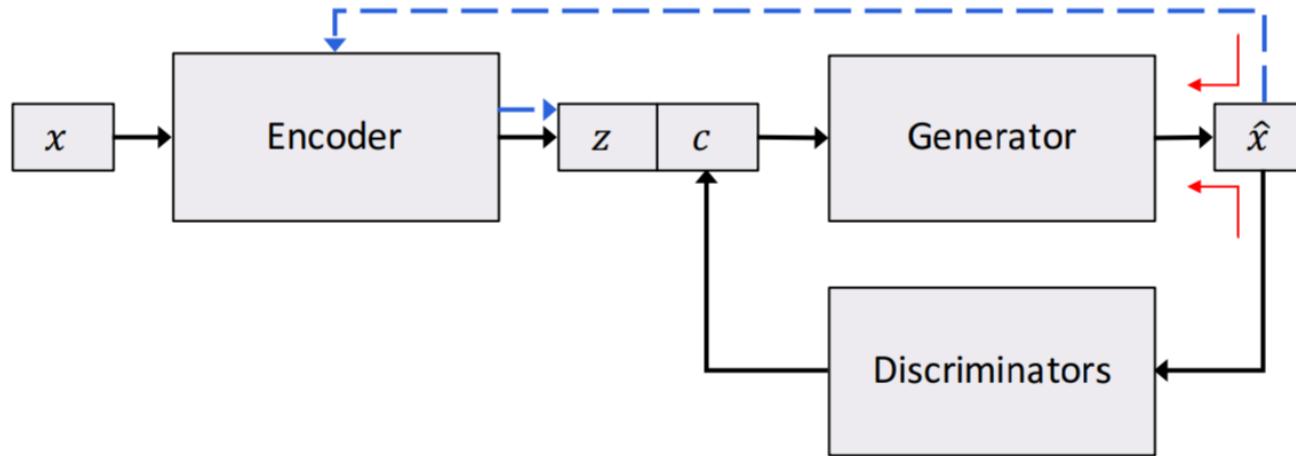


# Controlled Text Generation ([Hu et al., 2017](#))

48

Material: <http://opendialogue.miulab.tw>

- Idea: NLG based on generative adversarial network (GAN) framework
  - $c$ : targeted sentence attributes



# Issues in NLG

49

Material: <http://opendialogue.miulab.tw>

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

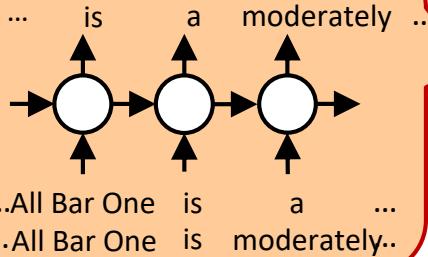
# Hierarchical NLG w/ Linguistic Patterns (Su et al., 2018)

50

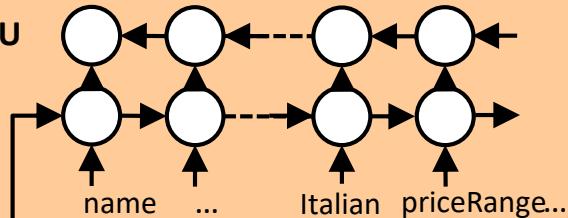
Material: <http://opendialogue.miulab.tw>

## GRU Decoder

1. Repeat-input
2. Inner-Layer Teacher Forcing
3. Inter-Layer Teacher Forcing
4. Curriculum Learning



## Bidirectional GRU Encoder



## Semantic 1-hot Representation

[ ... 1, 0, 0, 1, 0, ... ]

**Input** name[Midsummer House], food[Italian],  
**Semantics** priceRange[moderate], near[All Bar One]

ENCODER

$h_{enc}$

Near All Bar One is a moderately priced Italian place it is called Midsummer House

## 4. Others

All Bar One is moderately priced Italian place it is called Midsummer House

## DECODING LAYER3

## 3. ADJ + ADV

All Bar One is priced place it is called Midsummer House

## DECODING LAYER2

## 2. VERB

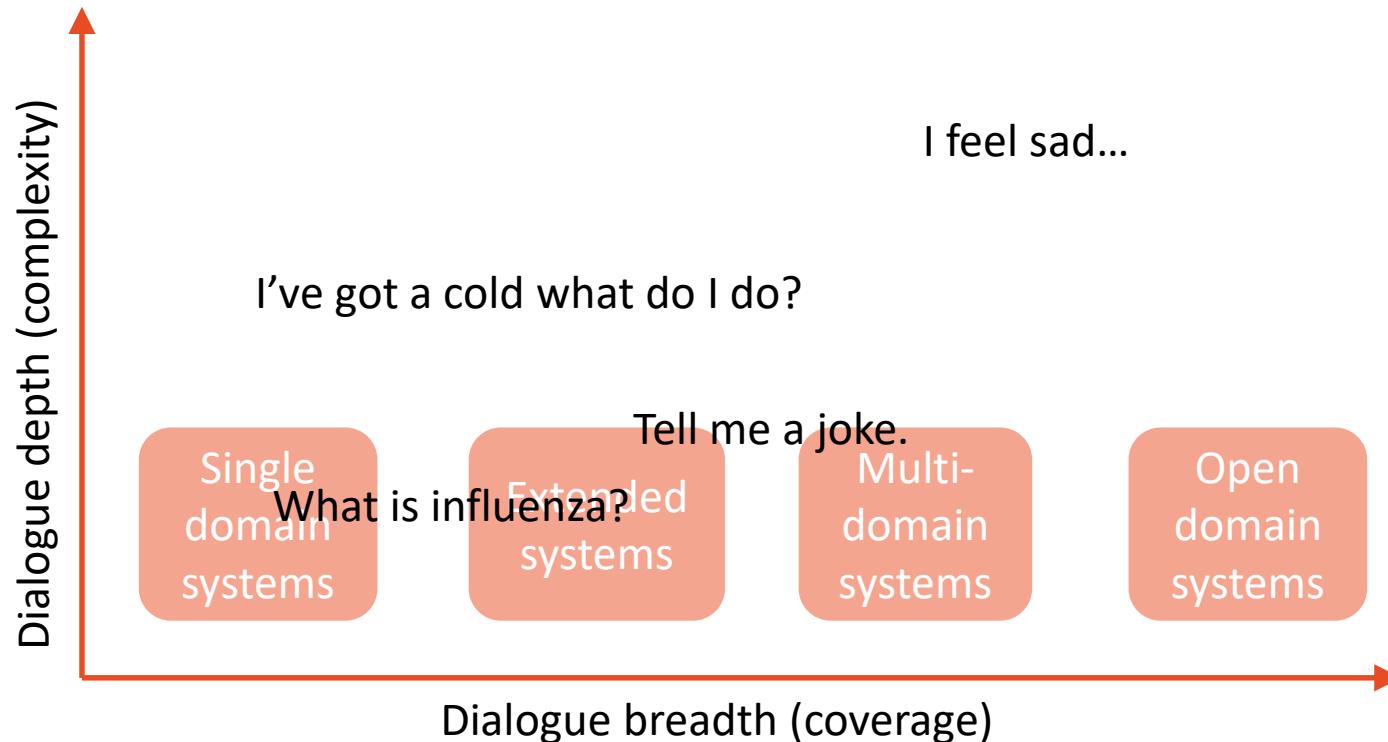
All Bar One place it Midsummer House

## DECODING LAYER1

## 1. NOUN + PROPN + PRON Hierarchical Decoder

# Evolution Roadmap

51

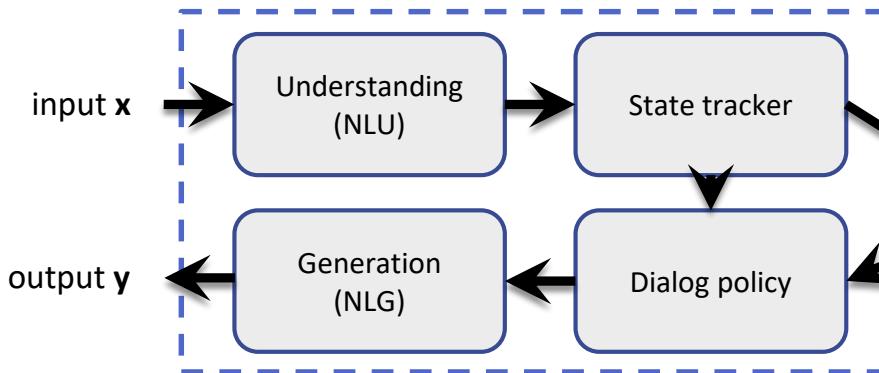
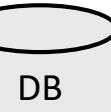
Material: <http://opendialogue.miulab.tw>

# Dialogue Systems

52

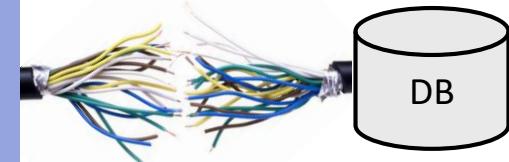
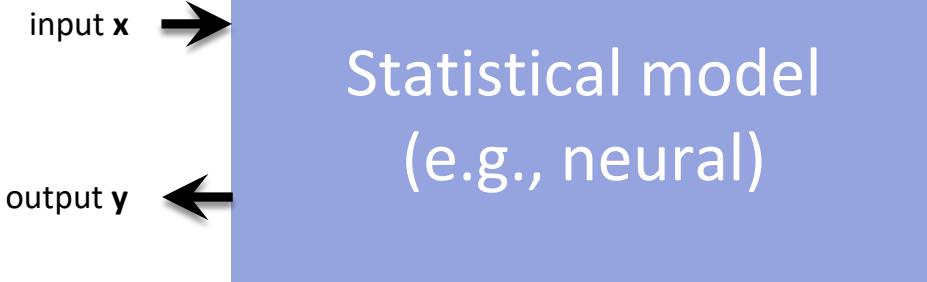
Material: <http://opendialogue.miulab.tw>

Task-Oriented Dialogue

Database  
MemoryThe  
Weather  
ChannelExternal knowledge  
 Wikipedia logo, showing a globe with puzzle pieces.  
 Tripadvisor logo, showing an owl.

tripadvisor®

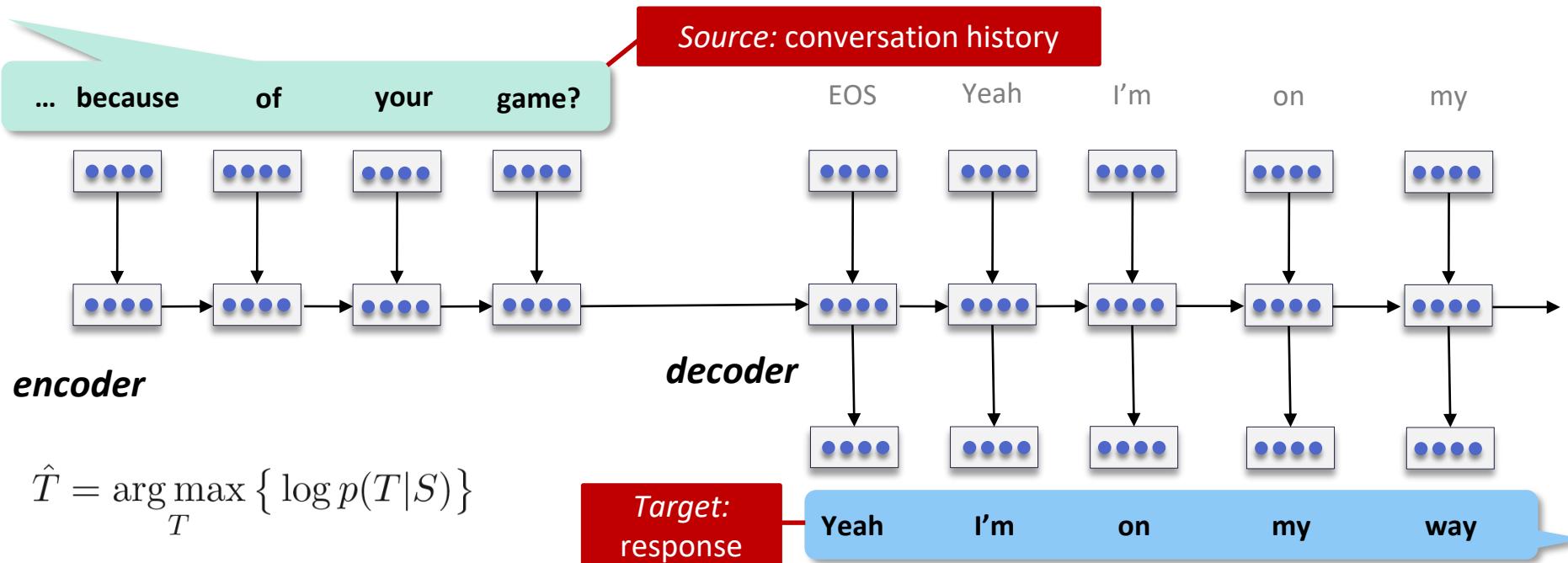
Fully Data-Driven



# Chit-Chat Social Bots

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

54

Material: <http://opendialogue.miulab.tw>

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

# Issue 1: Blandness Problem

55

Material: <http://opendialogue.miulab.tw>

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"*

# Mutual Information for Neural Generation

56

Material: <http://opendialogue.miulab.tw>

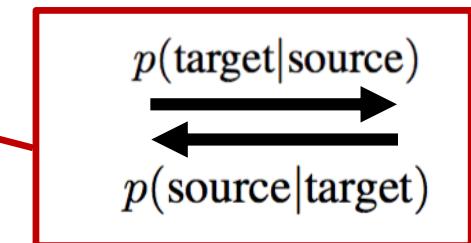
- Mutual information objective

$$\hat{T} = \arg \max_T \left\{ \log \frac{p(S, T)}{p(S)p(T)} \right\}$$

$$\hat{T} = \arg \max_T \left\{ \boxed{\log p(T|S)} - \boxed{\lambda \log p(T)} \right\} \quad \textit{Bayes' rule}$$

standard likelihood                              anti-LM

$$\hat{T} = \arg \max_T \left\{ (1 - \lambda) \log p(T|S) + \lambda \log p(S|T) \right\}$$



# Mutual Information for Neural Generation

57

Material: <http://opendialogue.miulab.tw>

$$\hat{T} = \arg \max_T \left\{ \log p(T|S) - \boxed{\lambda \log p(T)} \right\}$$

**antiLM:**  
large  $\lambda$  makes  
output ungrammatical

$$\hat{T} = \arg \max_T \left\{ (1 - \lambda) \log p(T|S) + \boxed{\lambda \log p(S|T)} \right\}$$

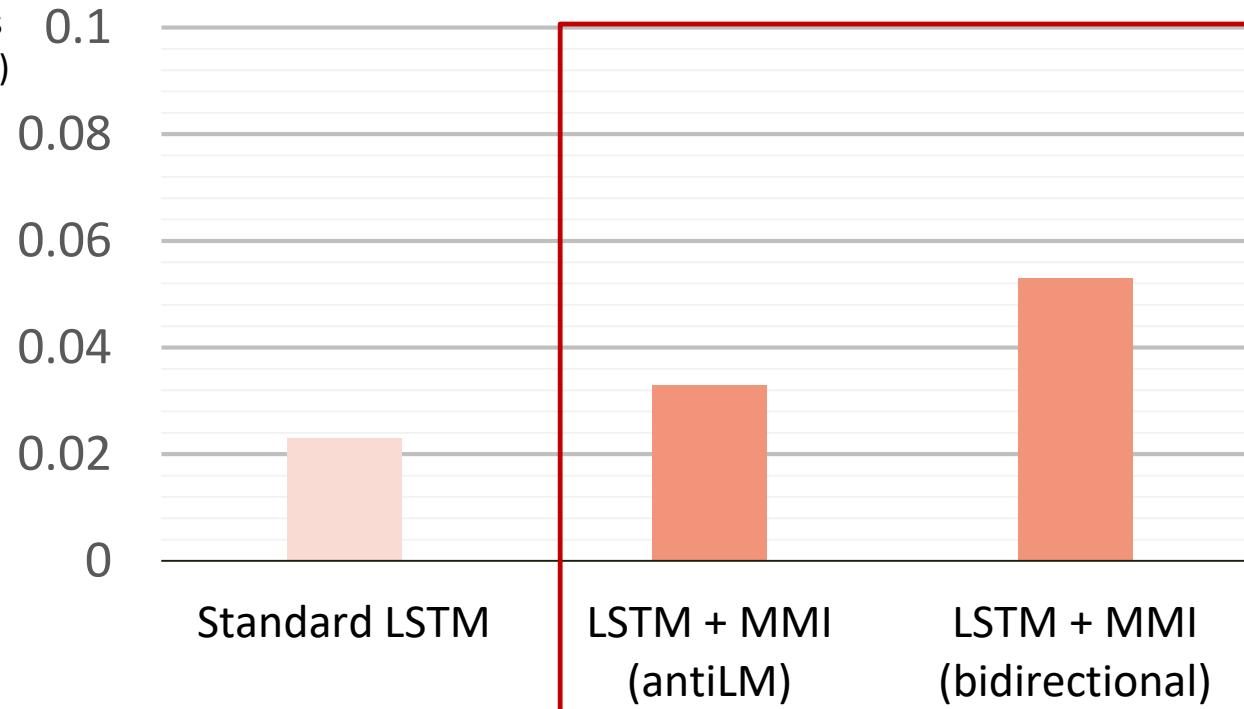
**used in rescoring  
(intractable otherwise)**

# Lexical Diversity

58

Material: <http://opendialogue.miulab.tw>

# Distinct Tokens in  
Generated Targets  
(Divided by Total #)



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

59

Material: <http://opendialogue.miulab.tw>

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?

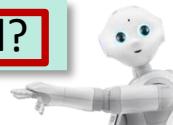
I think he'd be a good signing.

'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.



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

60

Material: <http://opendialogue.miulab.tw>

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.



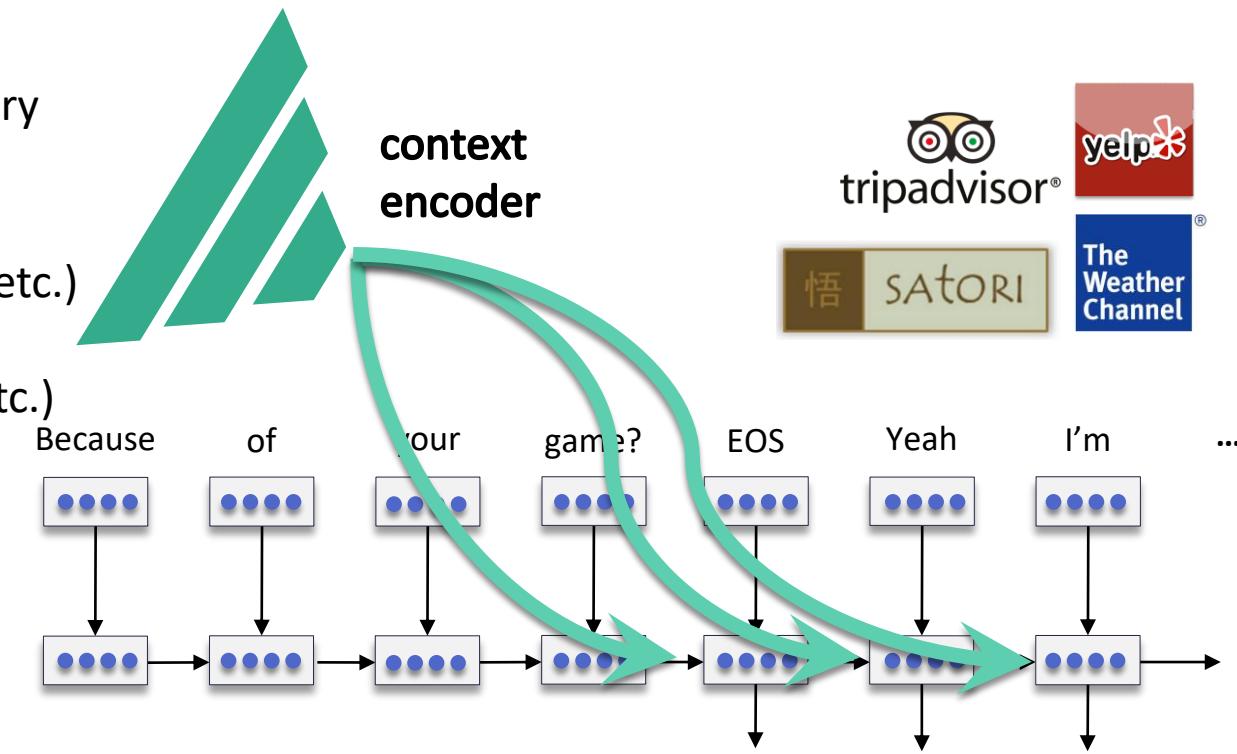
# Real-World Conversations

61

Material: <http://opendialogue.miulab.tw>

## □ Multimodality

- Conversation history
- Persona
- User profile data  
(bio, social graph, etc.)
- Visual signal  
(camera, picture etc.)
- Knowledge base
- Mood
- Geolocation
- Time



# Issue 2: Response Inconsistency

62

Material: <http://opendialogue.miulab.tw>

*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?

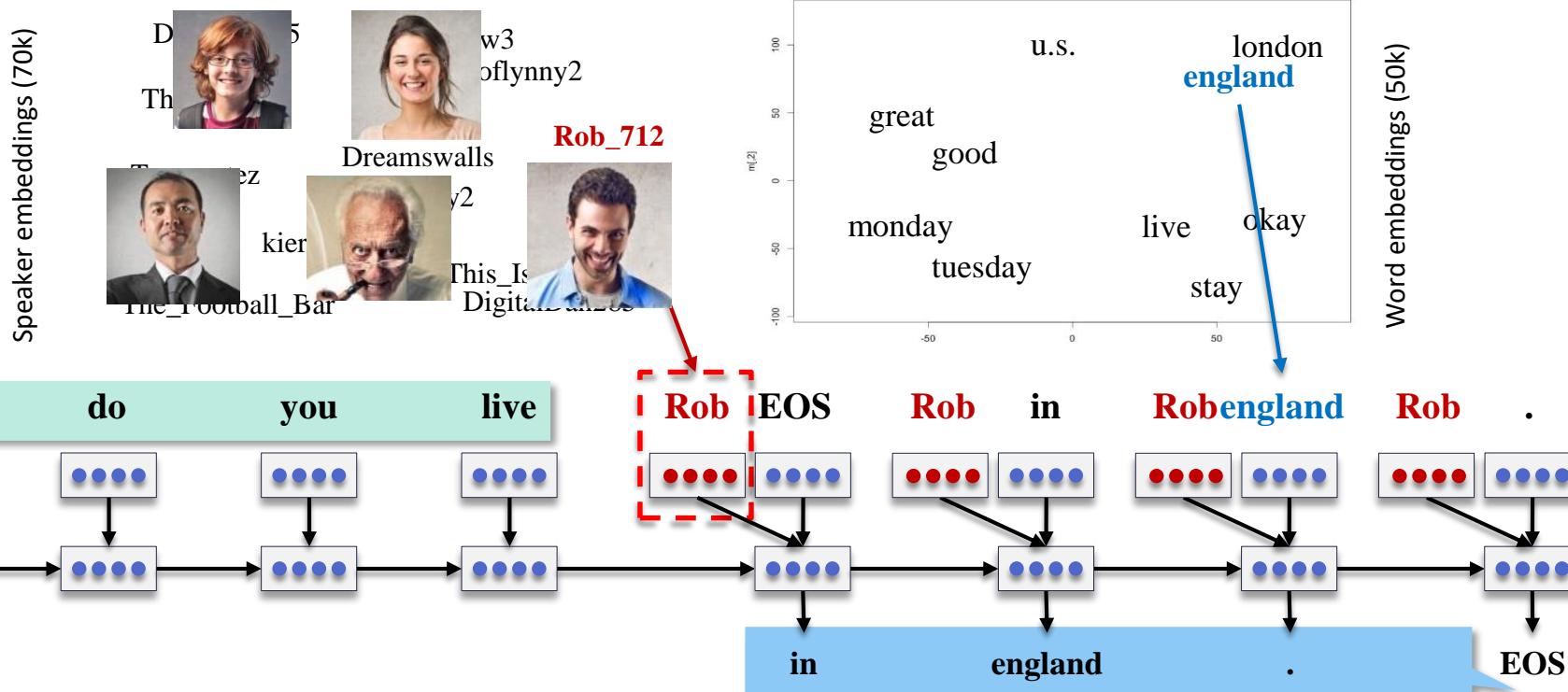
*response* I'm majoring in psychology

*message* What did you study in college?

*response* English lit.

# Personalized Response Generation (Li et al., 2016)

63

Material: <http://opendialogue.miulab.tw>

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

64

Material: <http://opendialogue.miulab.tw>

Baseline model → inconsistency

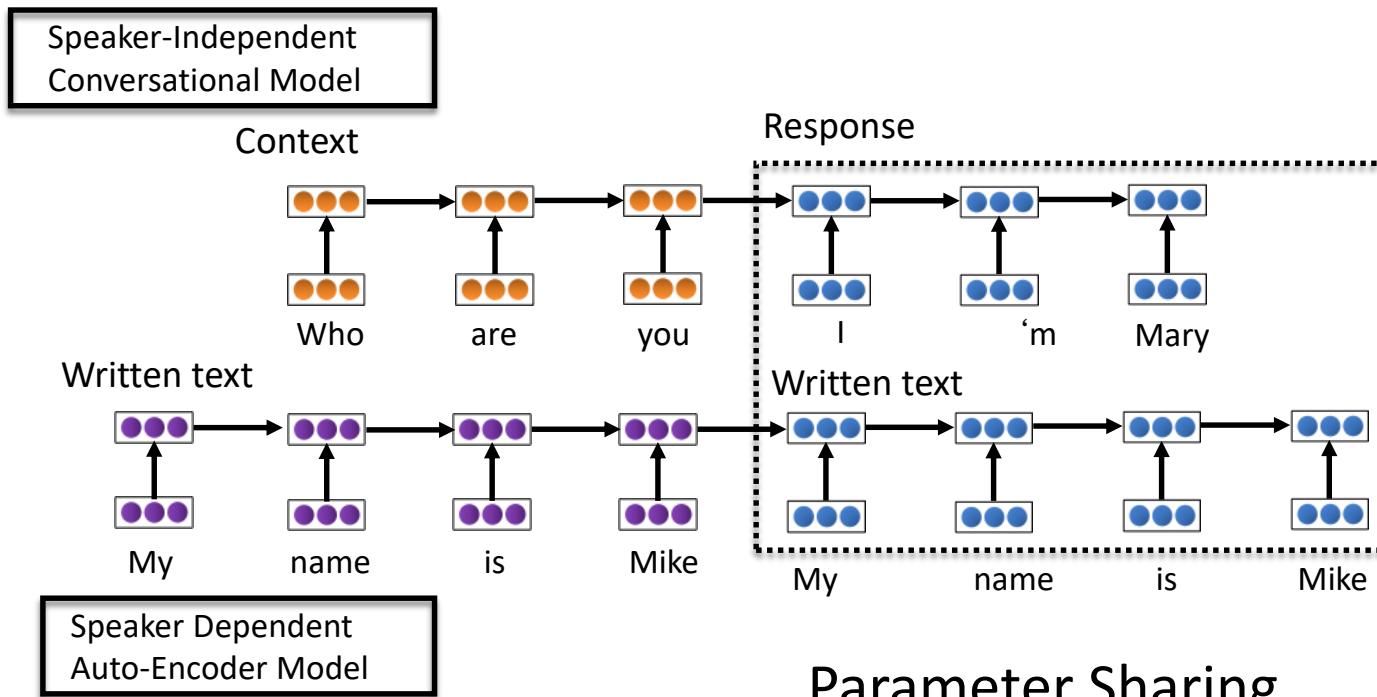
<i>message</i>	Where is your hometown?
<i>response</i>	I live in Indonesia.
<i>message</i>	Where are you from?
<i>response</i>	England, you?
<i>message</i>	In which city do you live now?
<i>response</i>	I live in Jakarta.
<i>message</i>	In which country do you live now?
<i>response</i>	I live in Indonesia.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

Persona model using speaker embedding → consistency

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

# Speak-Role Aware Response (Luan et al., 2017)

65

Material: <http://opendialogue.miulab.tw>

# Speak-Role Aware Response (Luan et al., 2017)

66

Material: <http://opendialogue.miulab.tw>

- Speaker role: support person
- Domain expertise: technical
- Speaking style: polite

**User input:** I am getting a loop back to login page.  
**Baseline model:** Ah, ok. Thanks for the info.  
**Our model:** I'm sorry to hear that. Have you tried clearing your cache and cookies?

*Context: have you tried uninstalling nike+ connec? how about docking it to a different computer?*

*Input: Tried both of those techniques. But nothing works. I can try the other computer again.*

**Baseline:** Let me know if there's anything else I can help you with. Thanks for the help.

**Multitask:** Have you tried using a different browser? Let us know if you have any other questions.

*Context: What do you see when you have a failed login attempt? Never ending spinner? Loop you back to login page?*

*Input: I am getting a loop back to login page, windows.*

**Baseline:** Ah, ok. Thanks for the info. I'll check it out. Thanks for the feedback.

**Multitask:** I'm sorry to hear that. Have you tried clearing your cache and cookies?

*Context: To clarify: your first run after resetting your connection will not post to Facebook. Does this help? Let us know.*

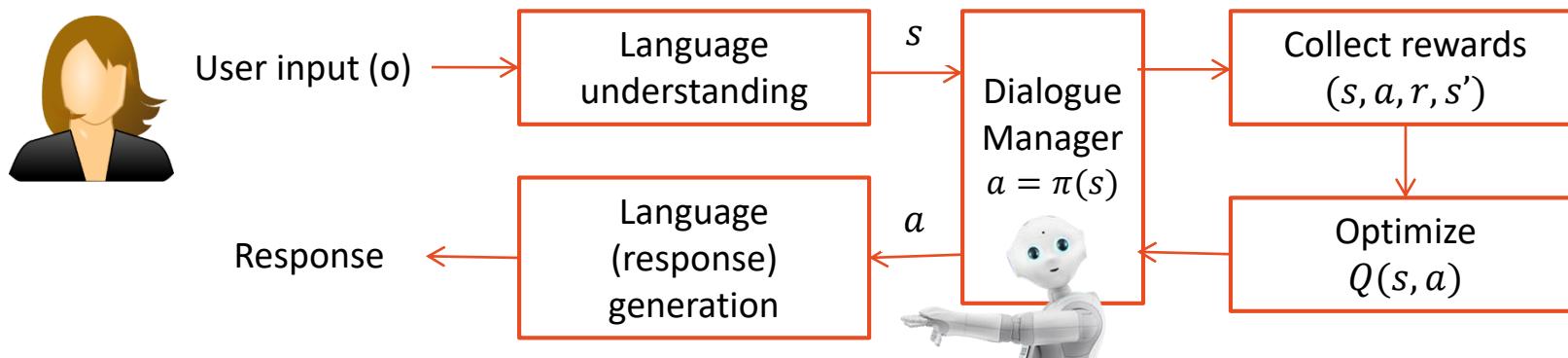
*Input: That's correct. I reset it twice! It still doesn't work.*

**Baseline:** Let me know if there's anything else I can help you with. Thanks for the update.

**Multitask:** I'm sorry to hear that. Are you receiving any error messages? Let us know.

# Issue 3: Dialogue-Level Optimization via RL

67

Material: <http://opendialogue.miulab.tw>

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)

68

Material: <http://opendialogue.miulab.tw>

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](#))

69

Material: <http://opendialogue.miulab.tw>

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

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

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

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

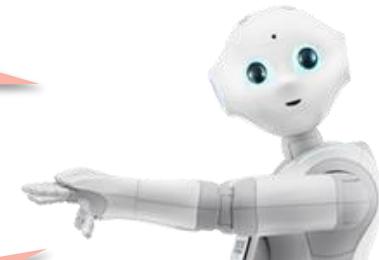
70

Material: <http://opendialogue.miulab.tw>

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!

# Image-Grounded Dialogues ([Mostafazadeh et al., 2017](#))

71

Material: <http://opendialogue.miulab.tw>

Neural model, trained on 250K Twitter convos:

- Turn 1: tweet + accompanying photo
- Turn 2: tweet is a question



Tweet: some of the damage done  
NeuroCon: **Is that at your house?**



Tweet: my poor baby moms saying it 's probably totaled so rip Bessie  
NeuroCon: **How did you manage to do that?**

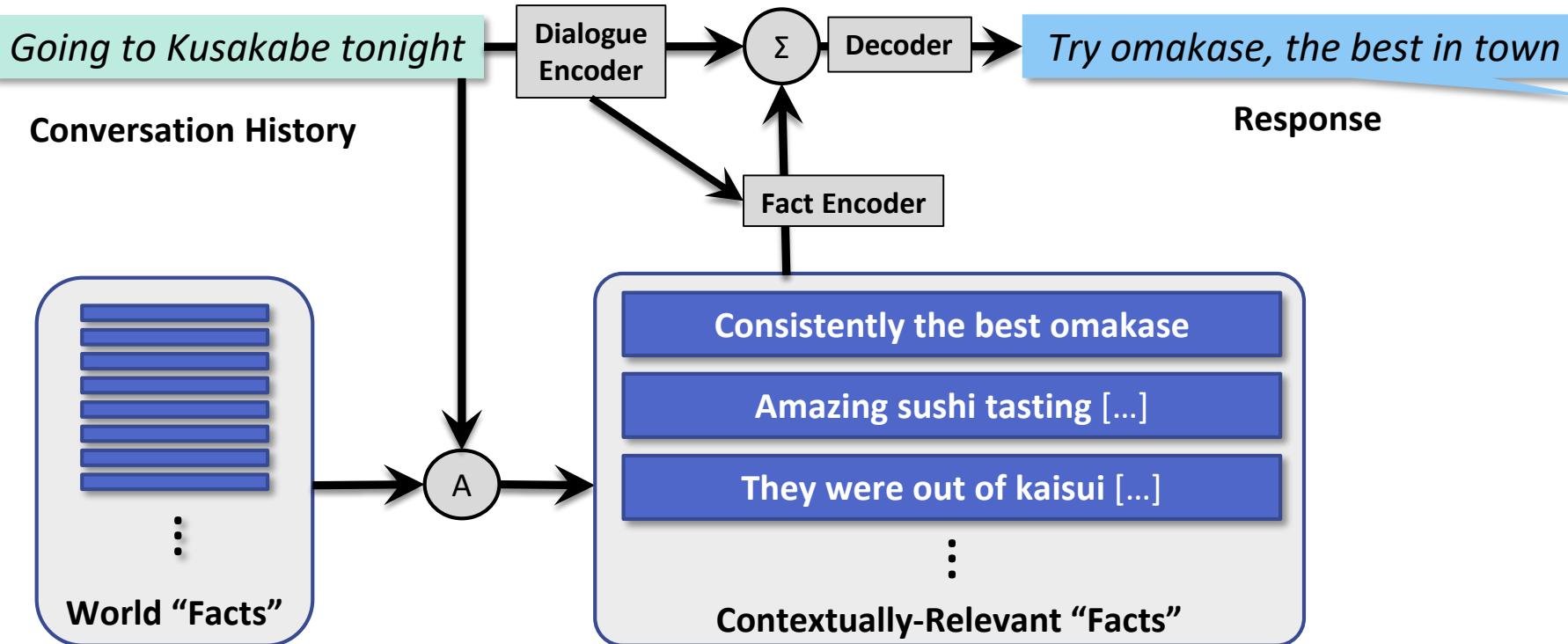
Given data, other signals can be exploited in the same way: program state, database info, geolocation, etc.



Tweet: as penitence I offer this  
NeuroCon: **can I have one of those?**

# Knowledge-Grounded Responses ([Ghazvininejad et al., 2017](http://opendialogue.miulab.tw))

72

Material: <http://opendialogue.miulab.tw>

# Conversation and Non-Conversation Data

73

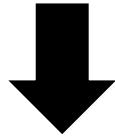
Material: <http://opendialogue.miulab.tw>

*You know any good **A** restaurant in **B**?*



*Try **C**, one of the best **D** in the city.*

Conversation Data



*You know any good **Japanese** restaurant in **Seattle**?*

*Try **Kisaku**, one of the best **sushi restaurants** in the city.*



**Kisaku**

515 reviews [Details](#)

\$ Sushi Bars, Japanese [Edit](#)



2101 N 55th St  
Ste 100  
Seattle, WA 98103  
bt 56th St & N Kenwood Pl  
Wallingford

[Get Directions](#)  
(206) 545-9050  
[kisaku.com](#)

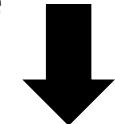


various rolls with tunas by Ken L



"Kisaku is one of the best sushi restaurants in Seattle and located in the heart of Wallingford. In 25 reviews"

Knowledge Resource



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

74

Material: <http://opendialogue.miulab.tw>

A: Visiting the celebs at Los Angeles International Airport (LAX) - [...] w/ 70 others

B: Nice airport terminal. Have a safe flight.

A: Is that [...] in your photos? It's on my list of places to visit in NYC.

B: Don't forget to check out the 5th floor, while you are here, it's a great view.

A: Live right now on [...] Tune in!!!!

B: Listen to Lisa Paige

A: Been craving Chicken Pot Pie-who has the best? Trying [...] at [...] Must be Change of weather!

B: Love the pasta trattoria.

A: So [...] is down to one copy of Pound Foolish. I'm curious to see if they are re-ordering it.

B: Check out the video feed on 6 and take a picture of the Simpsons on the 3rd floor.

A: I wish [...] would introduce another vegetarian option besides the shroomburger. It's delicious but kind of ridiculous.

B: This is the best j.crew in the world. Try the lemonade!

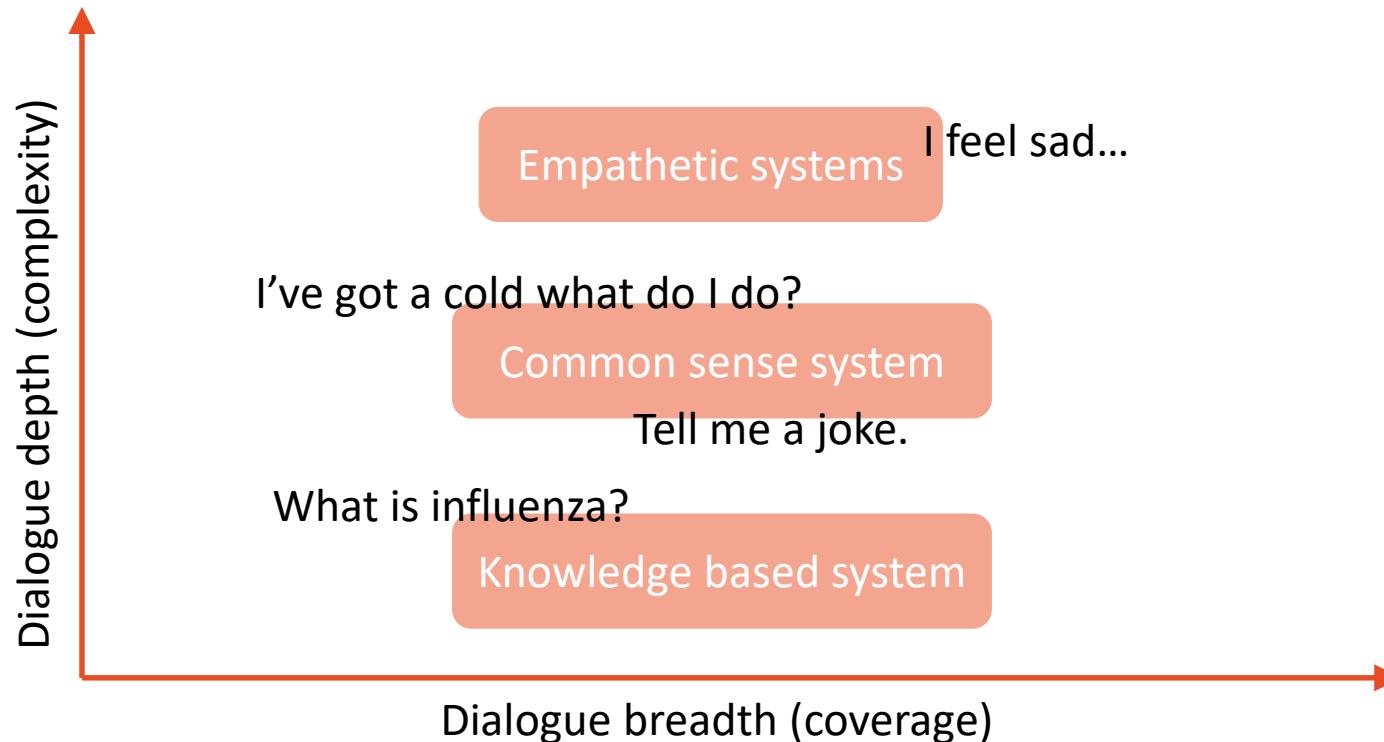
A: Just had an awesome dinner at [...] Great recommendation [...]

B: One of my favorite places I've ever been to in NYC. The food is great and the service is lackluster.

Results (23M conversations) outperforms competitive neural baseline (human + automatic eval)

# Evolution Roadmap

75

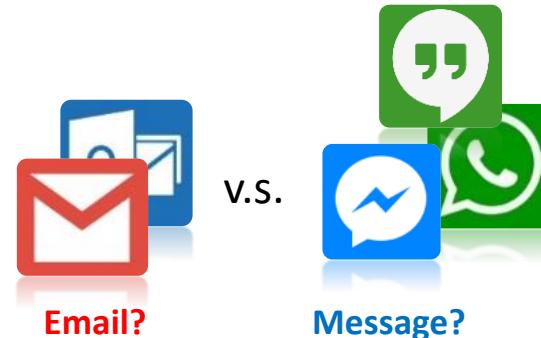
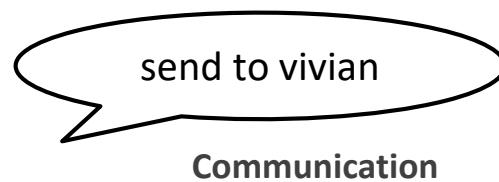
Material: <http://opendialogue.miulab.tw>

# Multimodality & Personalization ([Chen et al., 2018](http://opendialogue.miulab.tw))

76

Material: <http://opendialogue.miulab.tw>

- Task: user intent prediction
- Challenge: language ambiguity



## ① User preference

- ✓ Some people prefer “Message” to “Email”
- ✓ Some people prefer “Ping” to “Text”

## ② App-level contexts

- ✓ “Message” is more likely to follow “Camera”
- ✓ “Email” is more likely to follow “Excel”

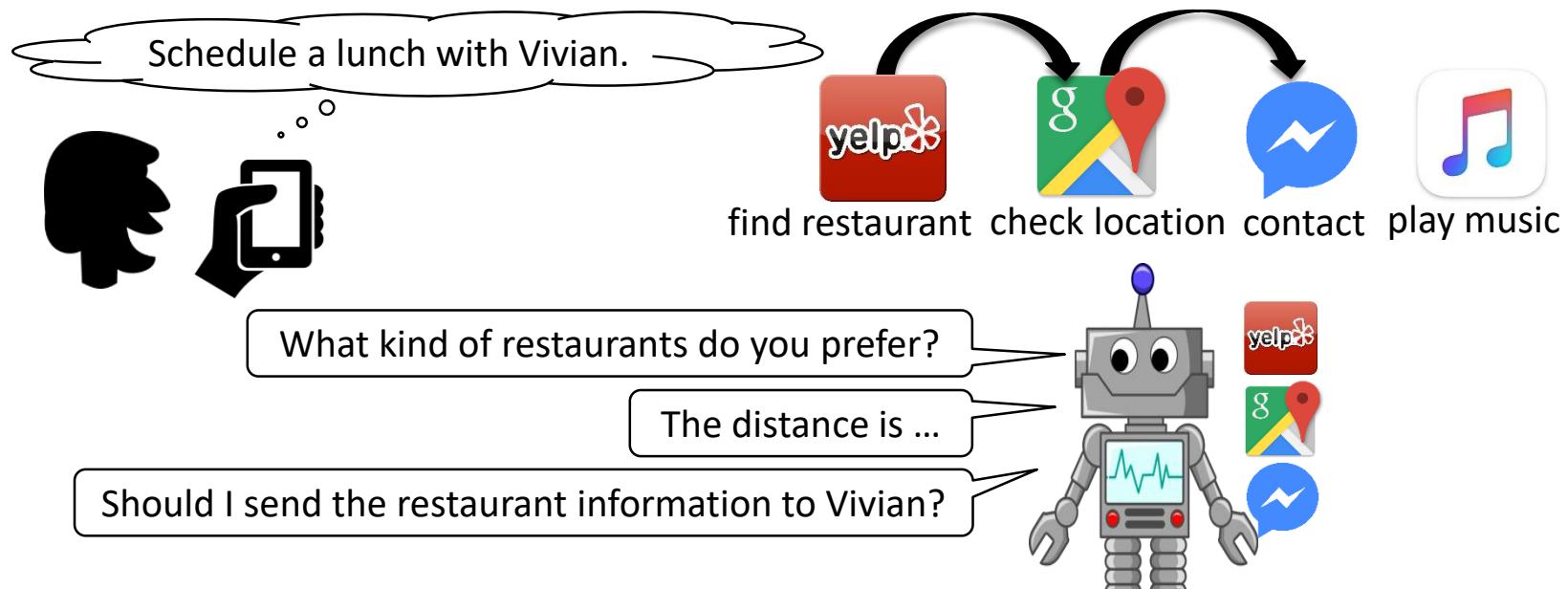
Behavioral patterns in history helps intent prediction.

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

77

Material: <http://opendialogue.miulab.tw>

- High-level intention may span several domains



Users interact via high-level descriptions and the system learns how to plan the dialogues

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

78

Material: <http://opendialogue.miulab.tw>

**Zara** - The Empathetic Supergirl



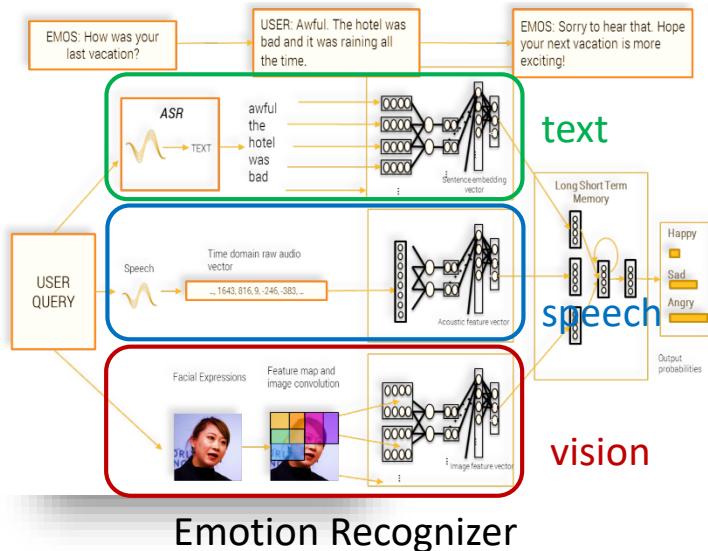
Made with love by Ivo Technologies in collaboration with Hong Kong University of Science and Technology



Face recognition output

```
{
  "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"
}
```

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



# Cognitive Behavioral Therapy (CBT)

79

Material: <http://opendialogue.miulab.tw>

## Mood Tracking



## Pattern Mining



## Depression Reduction



Daily lessons and check-ins

Quick conversations to feel better

## Content Providing



## Always Be There



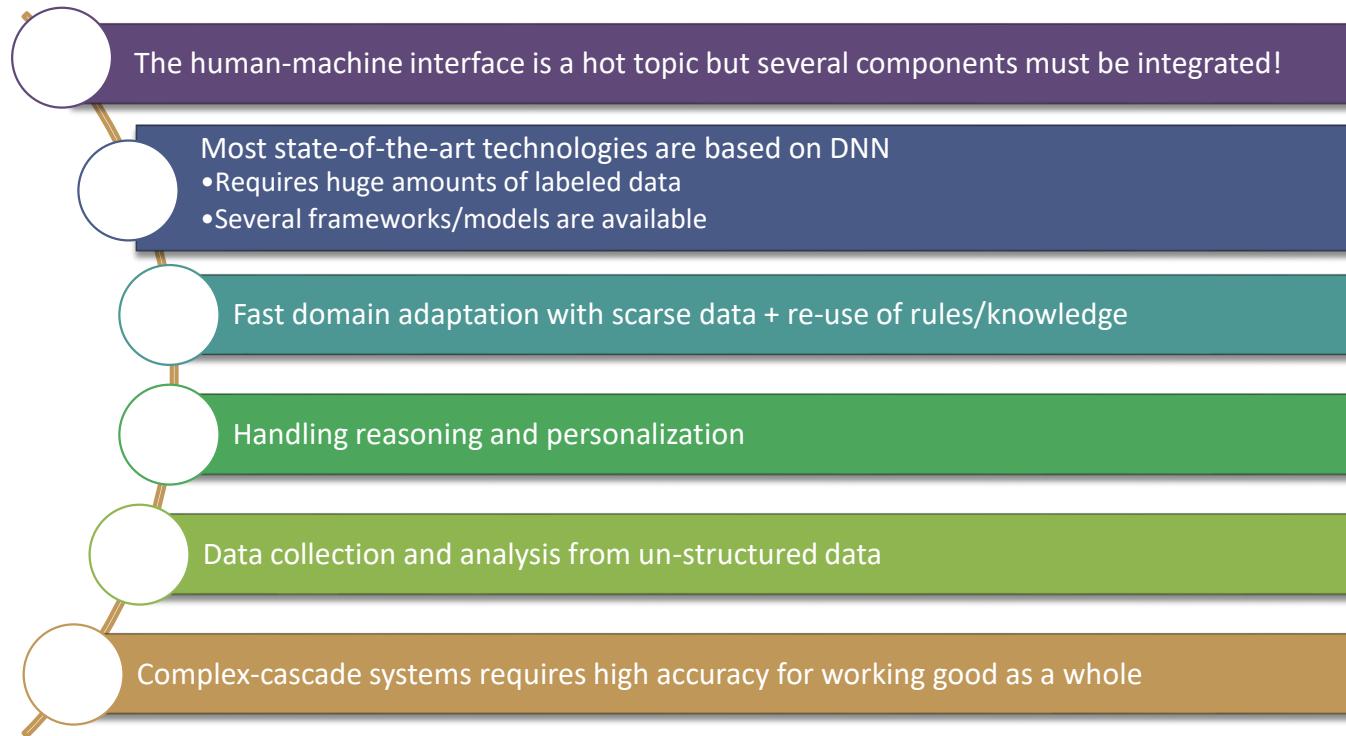
## Know You Well



## Challenges and Conclusions

# Challenge Summary

81

Material: <http://opendialogue.miulab.tw>

A man with glasses and a mustache, wearing a red shirt, sits at a desk in a dark room. He is looking at a computer monitor which displays a red screen with a white infinity symbol. The room has a window showing a city skyline at night. On the desk are books, papers, and a lamp.

# Her (2013)

What can machines achieve now or in the future?

# Thanks for Your Attention!

## Q & A



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