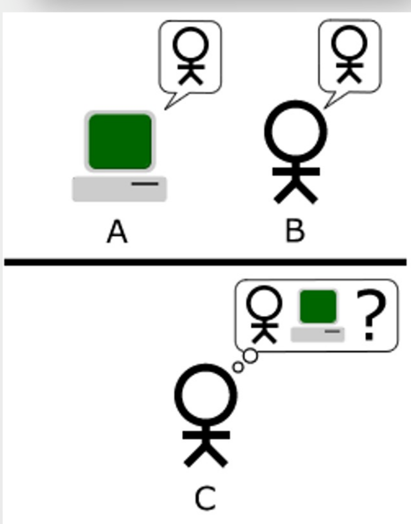
Dialogue

CSC401/2511 – Natural Language Computing – Spring 2022
Lecture 12
University of Toronto

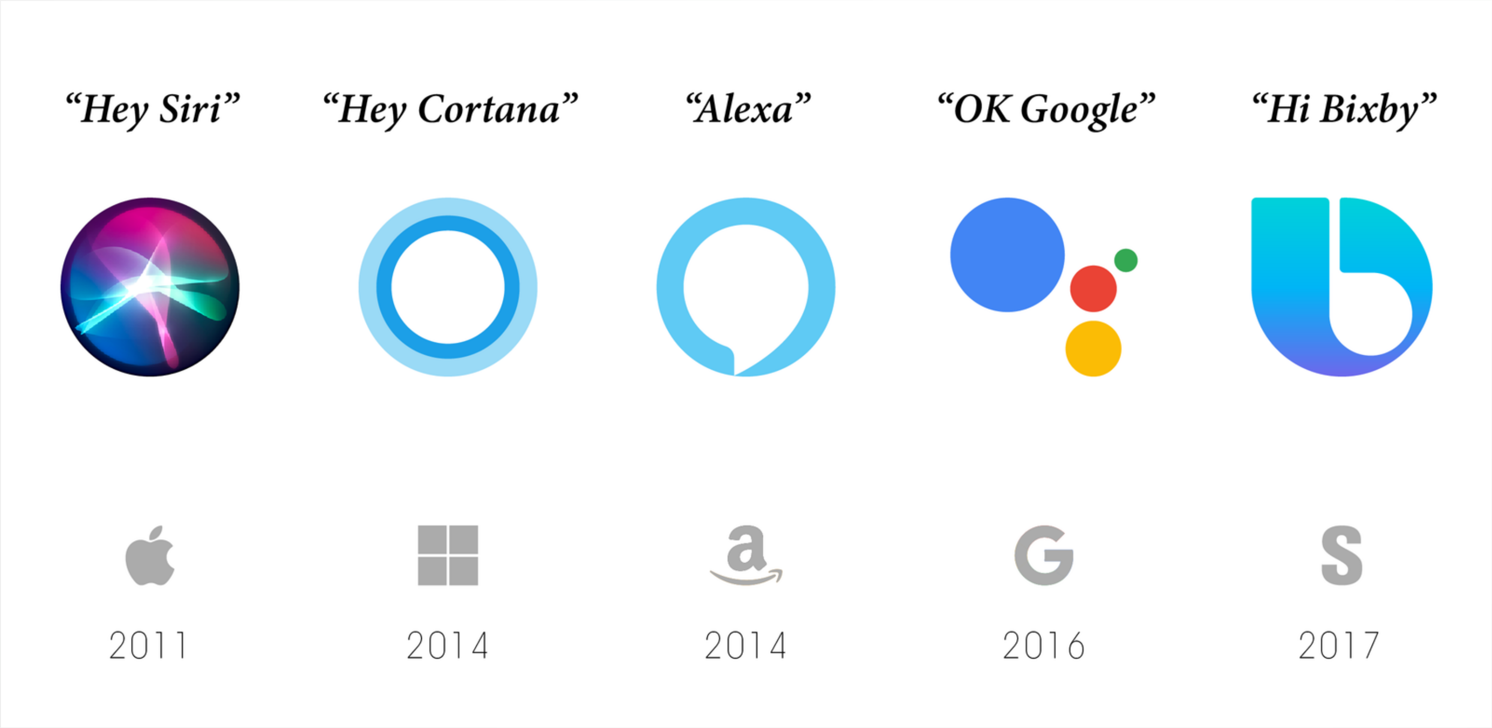
Dialogue – the final frontier



- Human-like dialogue with a machine was literally the ***first* task** proposed in the field of artificial intelligence.
- It remains the **most elusive**.

- To succeed, our agents must:
  1. Understand the world and task, *and*
  2. Respond realistically and consistently.

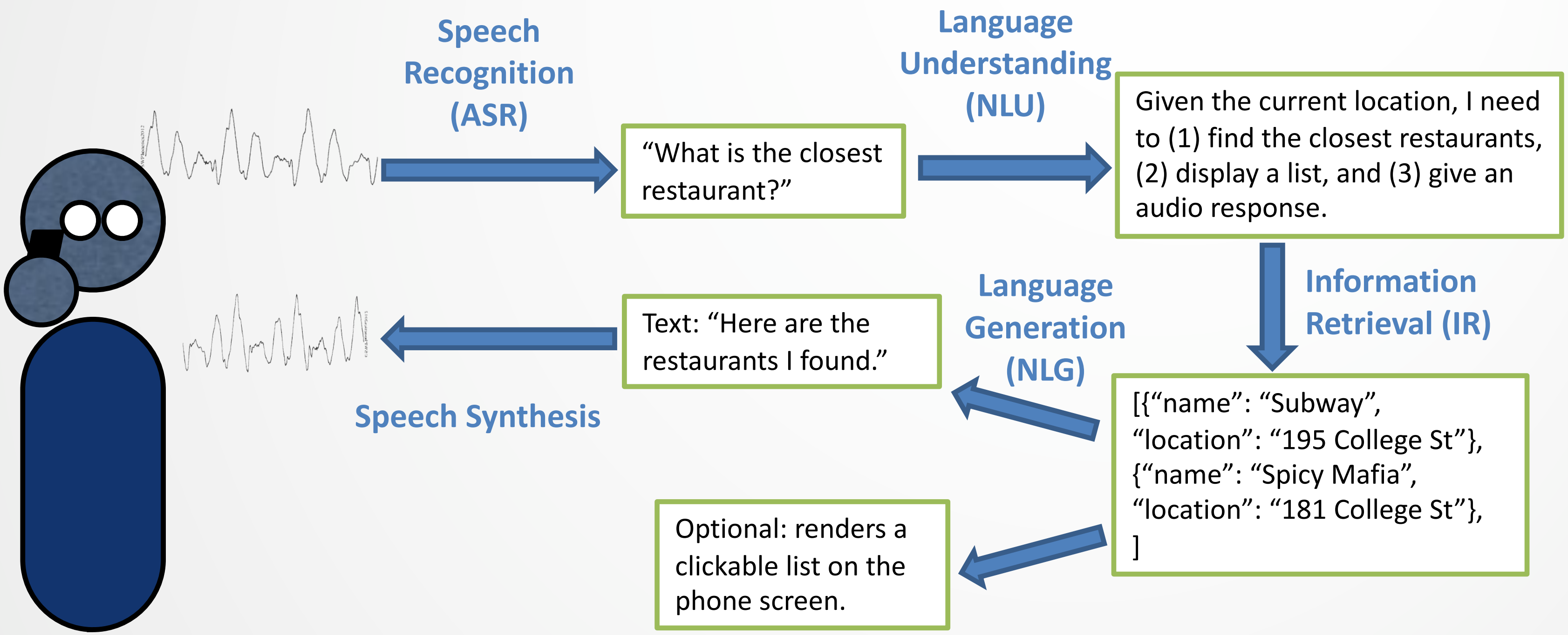
Personal assistants



Web apps vs. Dialogue Agents

	Web apps	Dialogue agents
Situation	Browsing, rarely a specific goal	Searching, with specific goal
Display	Structured	Semi-structured
Interface	Click/touch	Language
Easiness to learn	Some trial & error	No need to learn
Flexibility	Low, usually deterministic	High (one day)

Building blocks of a dialogue agent



- This illustrates one "turn" of dialogue – in multi-turn dialogues, we also need **Dialogue State Tracking (DST).**
- **ASR, NLU, IR,** and **synthesis** are all important components that we've already discussed. In this module, we will go over the remaining two components: **NLG, DST**.

Overview

Each building block is a relatively well-defined NLP task:

- Task setting.
- Approaches to address this task.
- Recent developments & debates about the task.

NATURAL LANGUAGE GENERATION

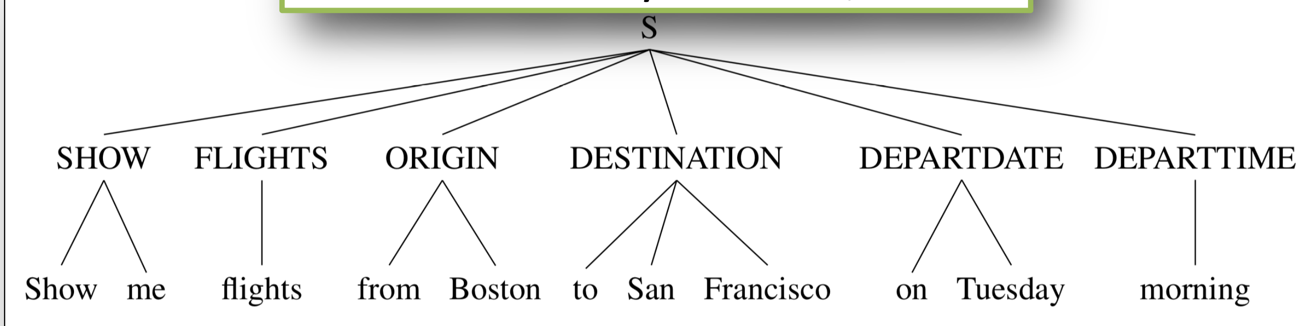
Generate coherent responses in human language

UNIVERSITY OF TORONTO

Fill the Slots

SHOW	$\rightarrow$	show me | i want | can i see|...
DEPART_TIME_RANGE	$\rightarrow$	(after|around|before) HOUR | morning | afternoon | evening
HOUR	$\rightarrow$	one|two|three|four...|twelve (AMPM)
FLIGHTS	$\rightarrow$	(a) flight | flights
AMPM	$\rightarrow$	am | pm
ORIGIN	$\rightarrow$	from CITY
DESTINATION	$\rightarrow$	to CITY
CITY	$\rightarrow$	Boston | San Francisco | Denver | Washington

That's not very scalable, is it?

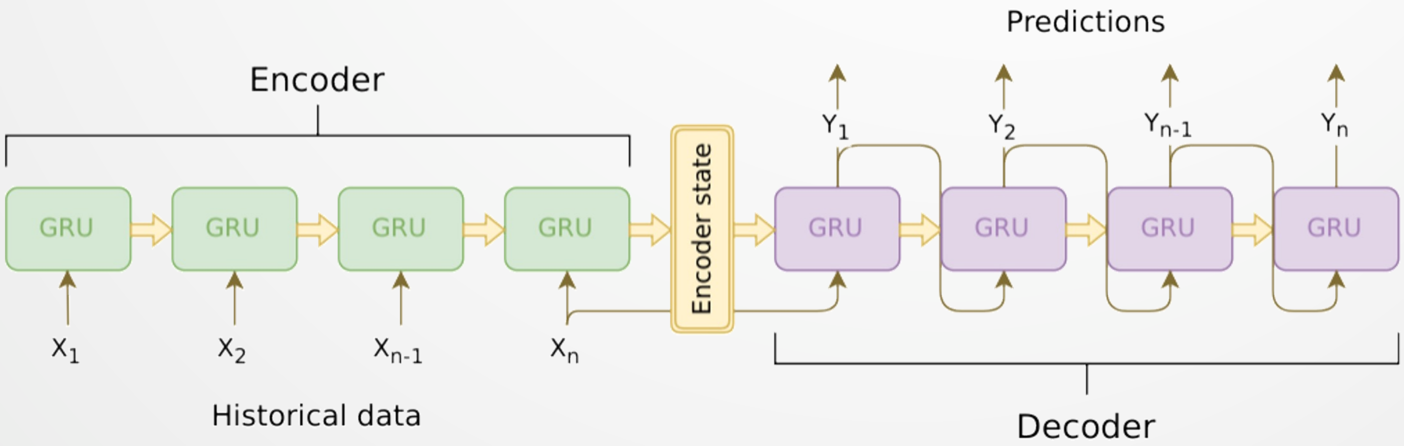


Speech and Language Processing. Daniel Jurafsky & James H. Martin. Copyright 2017. All rights reserved. Draft of August 7, 2017.

UNIVERSITY OF TORONTO

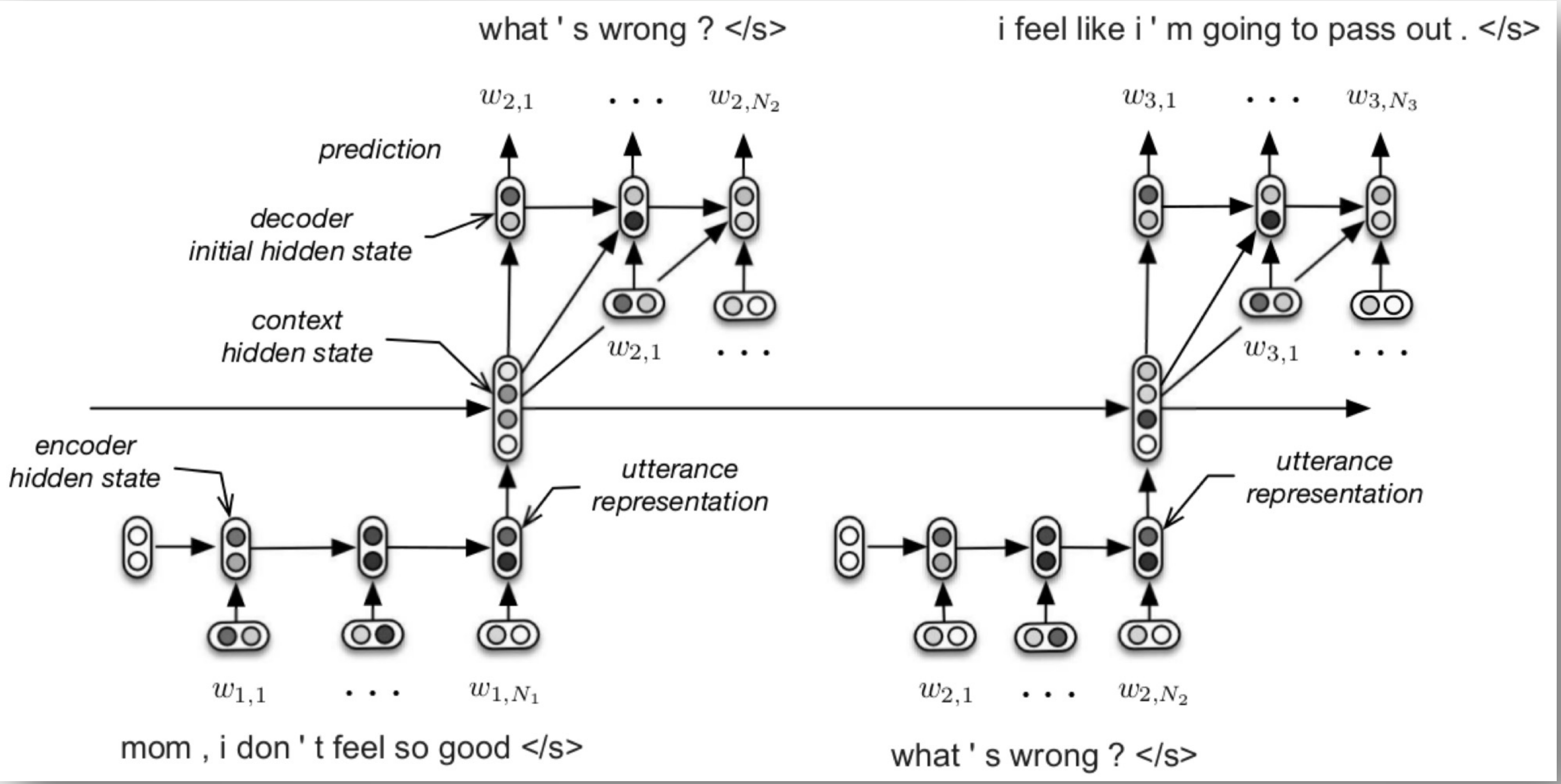
Sequence-to-sequence

- Generating a response can be considered as "translation".
- Sequence-to-sequence methods in translation may apply as well.
  - E.g., including beam search



UNIVERSITY OF TORONTO

End-to-end ~~translation~~ dialogue systems



Serban I V., Sordoni A, Bengio Y, et al. (2015) Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models.

Extensions exist that add **variational encoding** or **diversity-promoting objective functions** to avoid Siri-like repetitiveness repetitiveness.

UNIVERSITY OF TORONTO

End-to-end dialogue systems

- **Claim**: "we view our model as a **cognitive system**, which has to carry out natural language **understanding, reasoning, decision** making, (*sic*) and natural language generation".
- **Objective**: Perplexity (where U is an utterance)…

$$\exp\left(-\frac{1}{N_w}\sum_{n=1}^{N}\log P_\theta(U_1^n, U_2^n, U_3^n)\right)$$

Serban I V., Sordoni A, Bengio Y, et al. (2015) Building End-To-End Dialogue Systems Using Generative Hierarchical Neural Network Models.

- **Overhype** *vb.* make exaggerated claims about (a product, idea, or event) ; publicize or promote excessively

Language degeneration problem

- NLG models like to repeat themselves.
- Here's an example from GPT-2:

```
prompt = 'I sometimes get bored.'
input_ids = tokenizer.encode(prompt, return_tensors='pt')
greedy_output = model.generate(input_ids, max_length=50)  # Greedy search
tokenizer.decode(greedy_output[0], skip_special_tokens=True)
```

```
'I sometimes get bored. I'm not sure if I'm going to be able to do it. I'm not sure if I'm going to be able to do it. I'm not sure if I'm going to be able to do it.'
```

```
beam_output = model.generate(
    input_ids, max_length=50, num_beams=5, early_stopping=True)  # Beam search
tokenizer.decode(beam_output[0], skip_special_tokens=True)
```

```
'I sometimes get bored. I don't know what to do. I don't know what to do. I don't kn
ow what to do. I don't know what to do. I don't know what to do. I don't know what'
```

Avoid repeating n-grams

- This is hacky, but it works!

```
beam_output = model.generate(
    input_ids, max_length=50, num_beams=5,
    no_repeat_ngram_size=2, early_stopping=True)  # Beam search + no repeat
tokenizer.decode(beam_output[0], skip_special_tokens=True)
```

```
'I sometimes get bored. I don\'t know what to do about it.\n\n"I\'m not sure if I\'m
going to go back to school or not, but I think it\'s time for me to get back on my f
eet."\n'
```

```
beam_output = model.generate(
    input_ids, max_length=50, num_beams=5,
    no_repeat_ngram_size=3, early_stopping=True)  # Beam search + no repeat
tokenizer.decode(beam_output[0], skip_special_tokens=True)
```

```
'I sometimes get bored. I don\'t know what to do with myself.\n\n"I don\'t want to g
o to the gym. I want to do something else."\n\nHe added: "I\'m not going to do anyth
ing else.'
```

Why do these models repeat?

- An intuition: NLG models have limited memories.
  - What happened, e.g., >3 steps away is forgotten.
  - The last 3 steps only provide a finite number of "patterns".
  - Therefore, once entering a cycle in NLG, it is hard to get out.
- How to address this problem?
  - Use larger hidden vectors, attention connections, specially-designed network structures, etc.
  - Introduce some randomness.

Let's get random by sampling

- Randomly sample token according to the distribution of tokens.

$$x_t \sim P(x_t = w | x_{1..(t-1)})$$

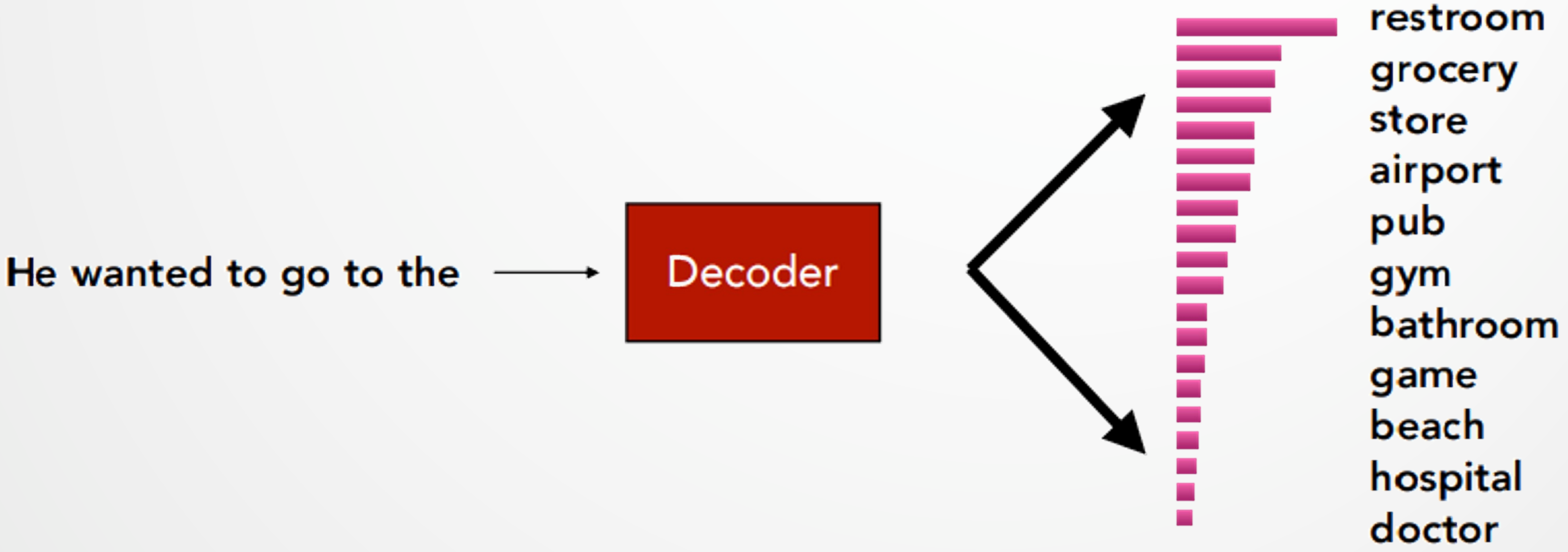


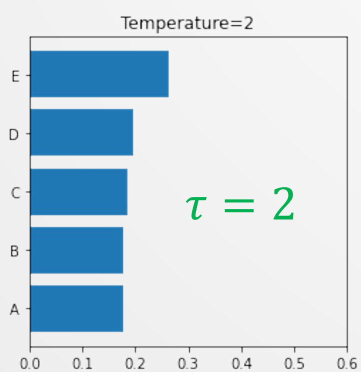
Image source: Antoine Bosselut's tutorial at ACL 2020 "Decoding from Neural Text Generation Models" Link

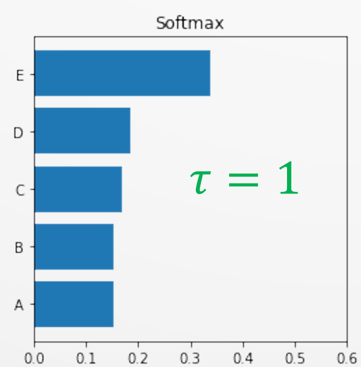
Scale the temperature

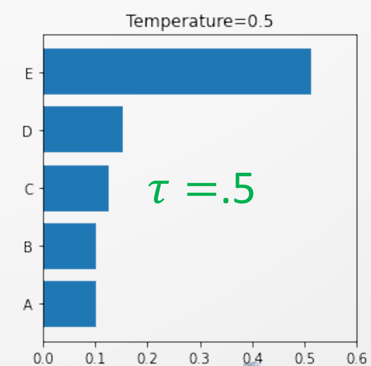
- The distribution might get too random…
- A solution: tune the temperature in softmax.

$$P\left(x_t = w \middle| x_{1..(t-1)}\right) = \frac{e^{o_n}}{\sum_{m=1}^{M} e^{o_m}}$$

$$P\left(x_t = w \middle| x_{1..(t-1)}\right) = \frac{e^{o_n/\tau}}{\sum_{m=1}^{M} e^{o_m/\tau}}$$







Sample from only top k

- We don't need *all* tokens in the vocabulary in this step.
- Those with small probabilities should have no chance at all; only consider top k candidates.

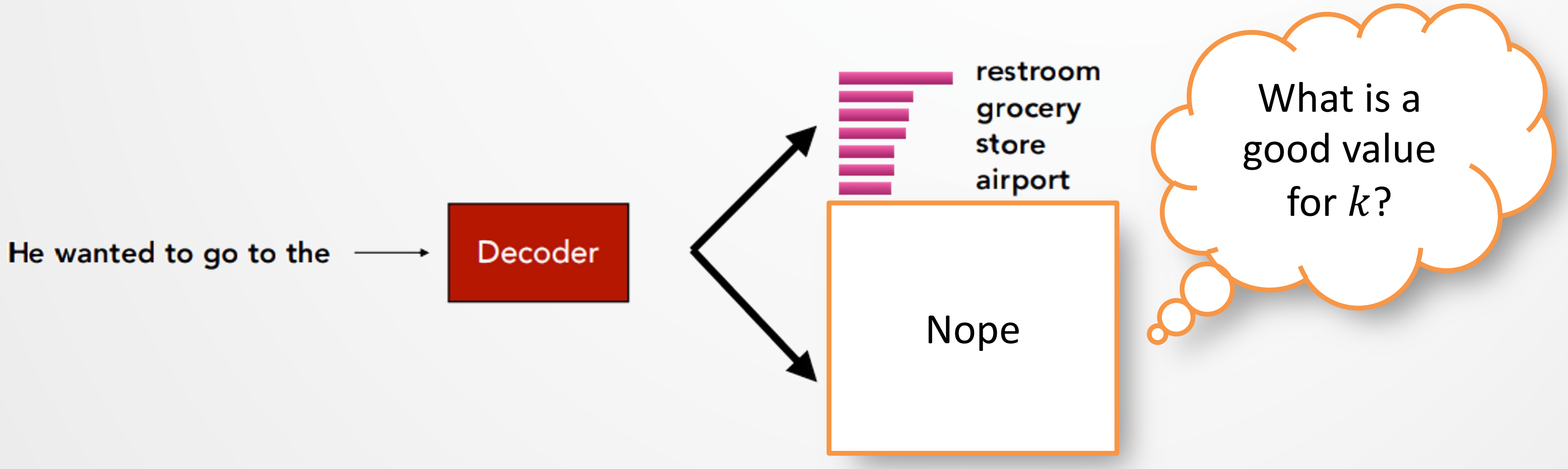
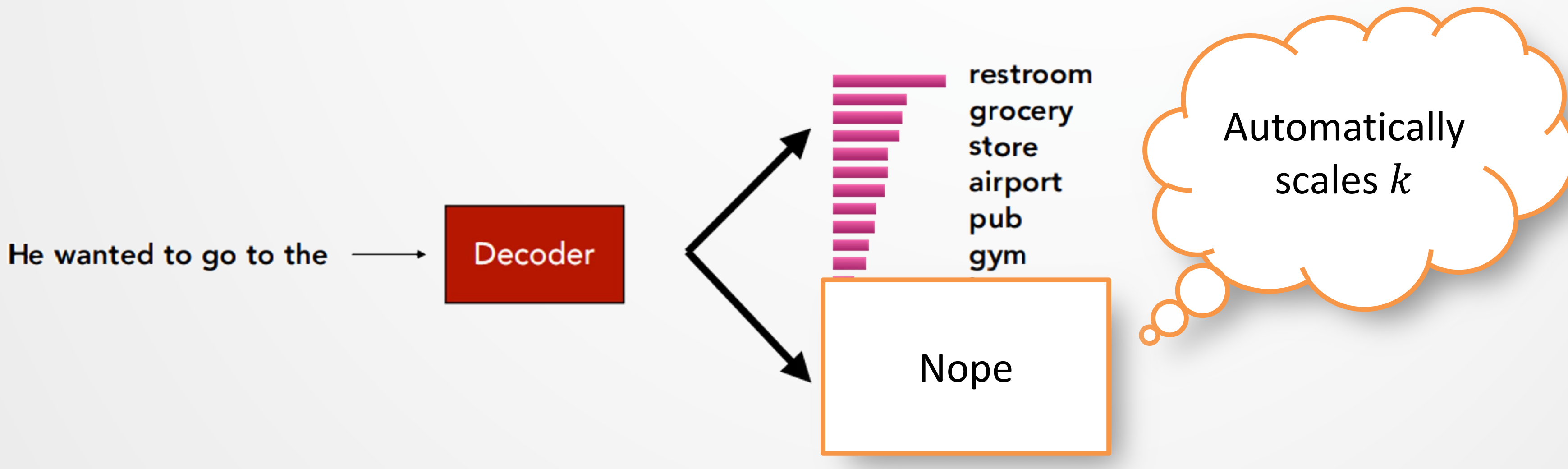


Image source: Antoine Bosselut's tutorial at ACL 2020 "Decoding from Neural Text Generation Models" [Link](#)

Top-p ("nucleus") sampling

- Sample from subset of vocabulary ("nucleus"), where probability mass is concentrated
- Sample from those candidates with $p > p_*$
  - where p_* is a hyper-parameter.



Holtzman, Ari, et al. "The curious case of neural text degeneration." *ICLR 2020*

Generation vs copying

- Sometimes, we want to **quote** from the input as part of a response.
- Sometimes, we want to **synthesize** the response.
- Pointer-Generator: let seq2seq models **learn to choose** from the two modes!
  - Compute a probability p_{gen} based on the decoder state and decoder inputs. Overall:

$$P(w) = p_{gen} P_{generate}(w) + \left(1 - p_{gen}\right) P_{copy}(w)$$

- $P_{generate}$ follows seq2seq computation.
- $P_{copy} \approx$ frequency of w in source document.

See, Abigail, Peter J. Liu, and Christopher D. Manning. "Get to the point: Summarization with pointer-generator networks." ACL 2017.

Evaluation criteria

- Some metrics to quantify the quality of generated sentences:
- If you have a target sentence:
  - N-gram overlap: **BLEU**, **ROUGE**, **METEOR**, …
  - Distance based: **Levenshtein**, ...
- To measure the diversity:
  - **Self-BLEU**: repetitiveness with oneself.
  - **Type-token ratio** (TTR): vocabulary richness.
- There are many other evaluation criteria!

Celikyilmaz, Asli, Elizabeth Clark, and Jianfeng Gao. "Evaluation of text generation: A survey." *arXiv preprint arXiv:2006.14799* (2020).

Type-Token Ratio (TTR)

- $TTR = \frac{N.\,unique\ tokens}{N.tokens}$
- More repetition -> lower TTR
- TTR measures the lexical richness.

what are thoughts well what are thoughts is a good question

Number of types (unique words) = 8
Number of total words = 11
Type-token ratio = 8 / 11 = 72.7%

PUTTING THE PIECES TOGETHER

Putting it together, for responding realistically and consistently

Dialogue Acts

- Everything in a discourse is a kind of **action** being performed by the speaker or writer.
- In speech, these are referred to as **speech acts**.
- In dialogue, these are referred to as **dialogue acts**.
- Dialogue is more complicated than mere speech, e.g.:
  - The hearer can **ground** on the speaker's utterances (i.e., acknowledge, and make it clear that the speaker understands).
  - There are some actions to **correct** the misunderstandings of the other speaker.

Dialogue Acts

- Here is a hypothetical conversation between a human and a smart assistant. Each utterance is labeled by a dialogue act :

“Where is the closest Subway?” [seek_information]
“The closest subway is the Queen’s Park subway station.” [information]
“No. I mean the Subway restaurant.” [correction]
“I see. Here is the information about this Subway.” [acknowledgement + information]
“Can you make an order?” [action_instruction]
“Ok.” [agree]

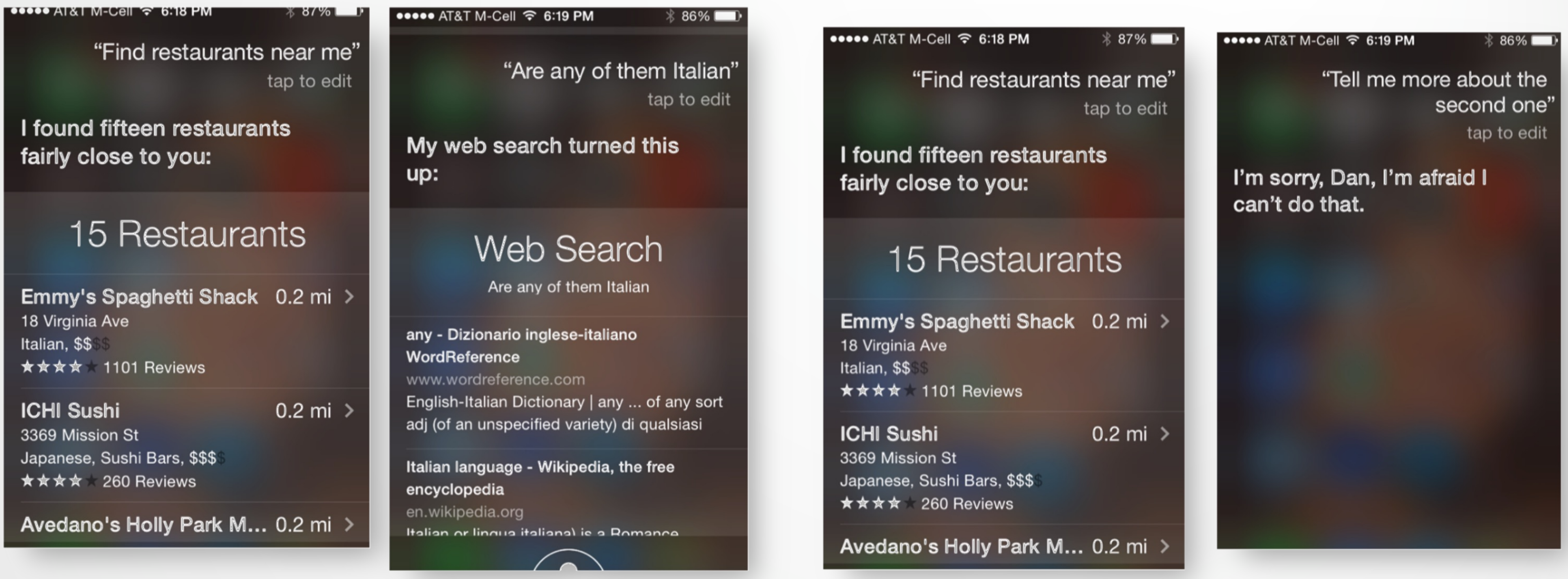
Dialogue Acts Classification

- Detecting the dialogue act is a popular NLP task.
- This is usually handled as a sequential tagging task.
- A typical dataset is **Switchboard Dialogue Act (SwDA) Corpus**.
- Here are some example annotations:

Name	Tag	Example
Statement-non-opinion	sd	Me, I'm in the legal department.
Statement-opinion	sv	I think it's great.
Agree / Accept	aa	That's exactly it.
Conventional-closing	fc	Well, it's been nice talking to you.

The Switchboard Dialog Act Corpus website: http://compprag.christopherpotts.net/swda.html

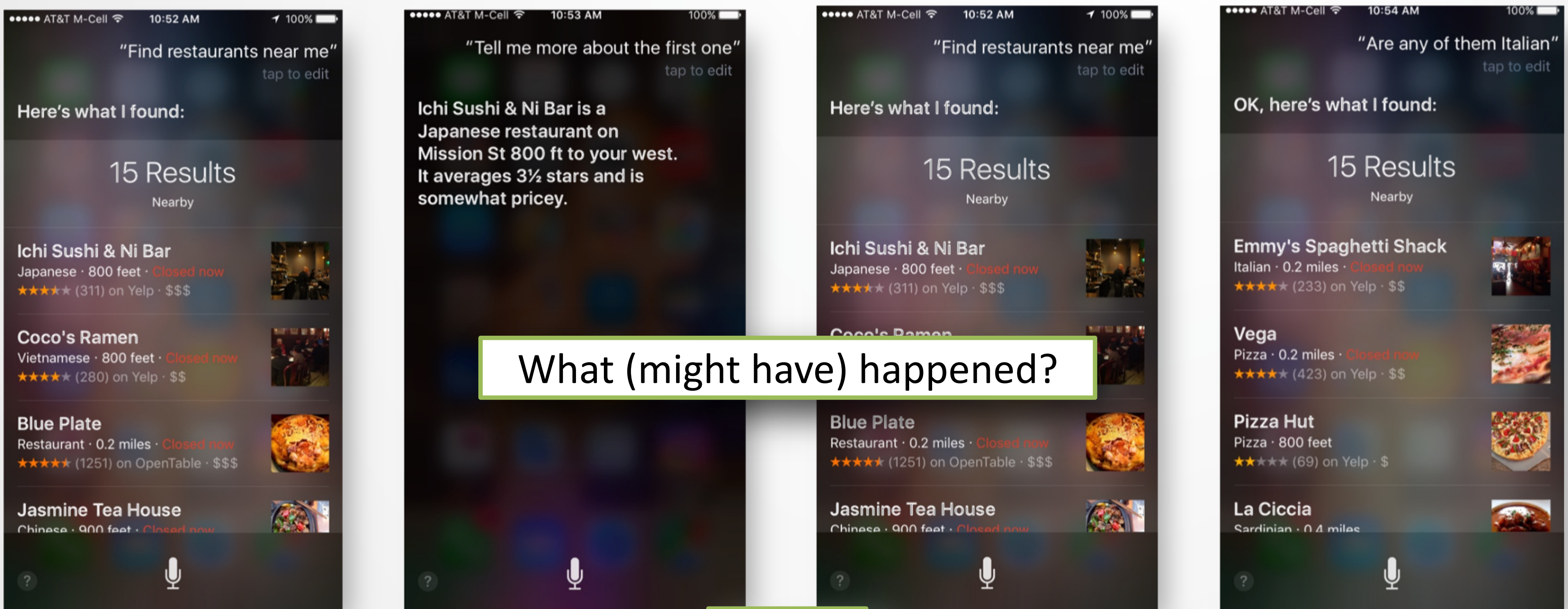
Let me Bing that for you



(a) **2014** (b)

Speech and Language Processing. Daniel Jurafsky & James H. Martin. Copyright 2017. All rights reserved. Draft of August 7, 2017.

Let me actually answer that for you



What (might have) happened?

(a) **2017** (b)

Speech and Language Processing. Daniel Jurafsky & James H. Martin. Copyright 2017. All rights reserved. Draft of August 7, 2017.

Chatbots should track the states

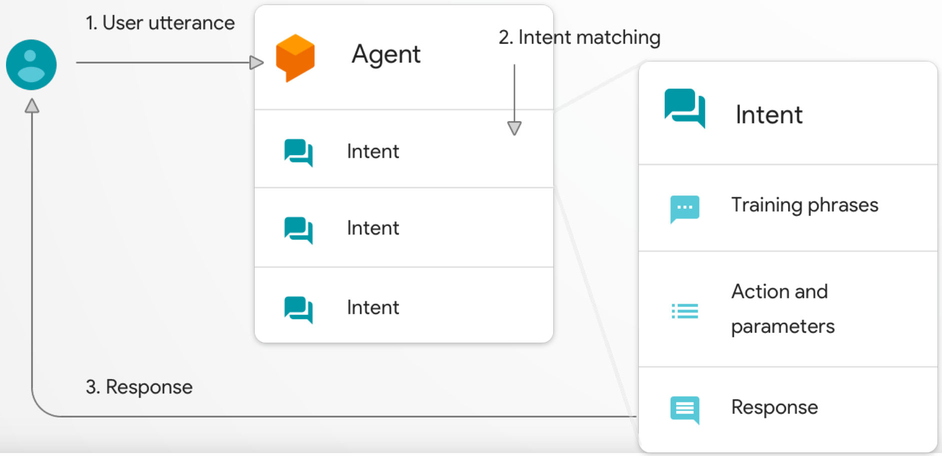
- When interacting with chatbots, there can be multiple turns.
- Dialogue responses should consider **both the context and the inquiry**.

"Where is the closest Subway?"
"The closest subway is the Queen's Park subway station."
"No. I mean the Subway restaurant."
"I see. Here is the information about this Subway."
"Can you make an order?"
(A) "Where do you want to make this order?"
(B) "Ok. What would you like to have?"

Is option A or B better?

States of (dis-)belief

- Map utterances to **dialogue acts** and **beliefs** about the world.
  - Maintain (*and update*!)* those beliefs. * Humans can barely do this.



https://dialogflow.com/docs/intro

act type	inform[*] / request[*] / select[123] / recommend/[123] / not found[123] request booking info[123] / offer booking[1235] / inform booked[1235] / decline booking[1235] welcome[*] /greet[*] / bye[*] / reqmore[*]
slots	address[*] / postcode[*] / phone[*] / name[1234] / no of choices[1235] / area[123] / pricerange[123] / type[123] / internet[2] / parking[2] / stars[2] / open hours[3] / departure[45] destination[45] / leave after[45] / arrive by[45] / no of people[1235] / reference no.[1235] / trainID[5] / ticket price[5] / travel time[5] / department[7] / day[1235] / no of days[123]

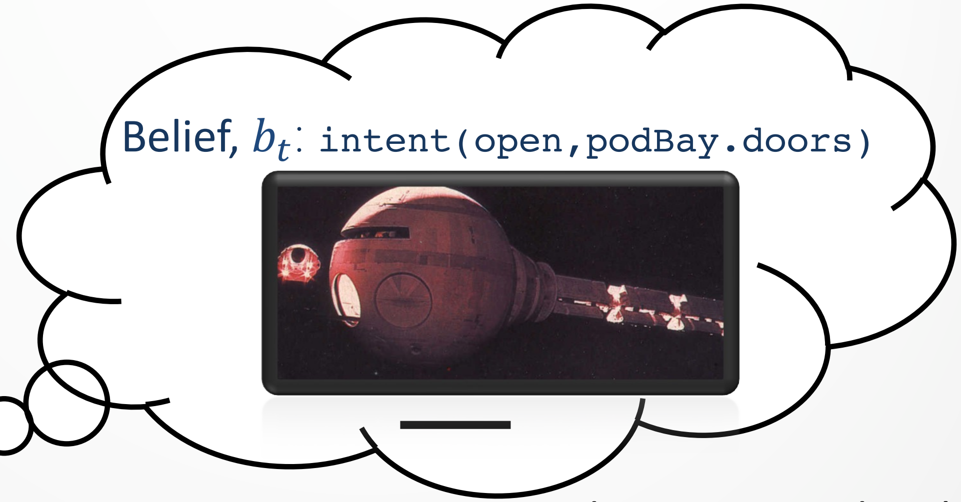
Mrkšić N, Séaghdha DÓ, Wen T-H, et al. (2016) Neural Belief Tracker: Data-Driven Dialogue State Tracking. http://arxiv.org/abs/1606.03777

	Core dialog acts
Info-request	Speaker wants information from addressee
Action-request	Speaker wants addressee to perform an action
Yes-answer	Affirmative answer
No-answer	Negative answer
Answer	Other kinds of answer
Offer	Speaker offers or commits to perform an action
ReportOnAction	Speaker notifies an action is being/has been performed
Inform	Speaker provides addressee with information not explicitly required (via an Info-request)
	Conventional dialog acts
Greet	Conversation opening
Quit	Conversation closing
Apology	Apology
Thank	Thanking (and down-playing)
	Feedback/turn management dialog acts
Clarif-request	Speaker asks addressee for confirmation/repetition of previous utterance for clarification.
Ack	Speaker expresses agreement with previous utterance, or provides feedback to signal understanding of what the addressee said
Filler	Utterance whose main goal is to manage conversational time (i.e. dpeaker taking time while keeping the turn)
	Non-interpretable/non-classifiable dialog acts
Other	Default tag for non-interpretable and non-classifiable utterances

Dinarelli M, Quarteroni S, Tonelli S. (2009) Annotating spoken dialogs: from speech segments to dialog acts and frame semantics. *Proc 2nd Work Semant Represent Spok Lang* 2009;:34–41. http://dl.acm.org/citation.cfm?id=1626301

State of (dis-)belief

- Use **reinforcement learning** to make these explicit.

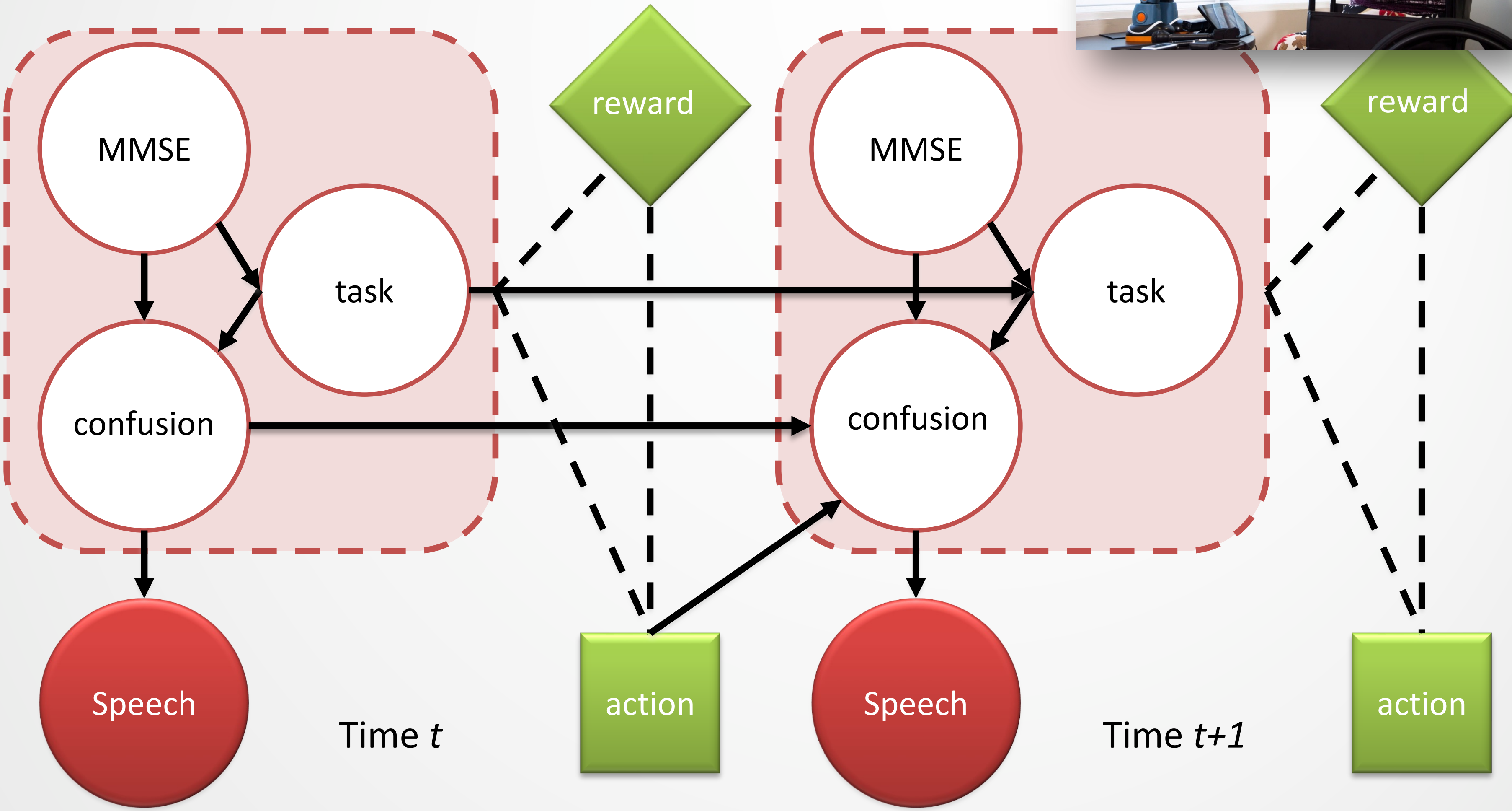


Very negative reward r_k associated with the door being open

Policy	$\pi(b) = a$
Return	$R_t = \sum_{k=t}^{T} \gamma^{(k)} r_k$
Value	$V^{\pi}(\boldsymbol{b}) = \mathrm{E}[R_t \mid b_t = \boldsymbol{b}]$
Q	$Q^{\pi}(\boldsymbol{b}, \boldsymbol{a}) = \mathrm{E}[R_t \mid b_t = \boldsymbol{b}, a_t = \boldsymbol{a}]$

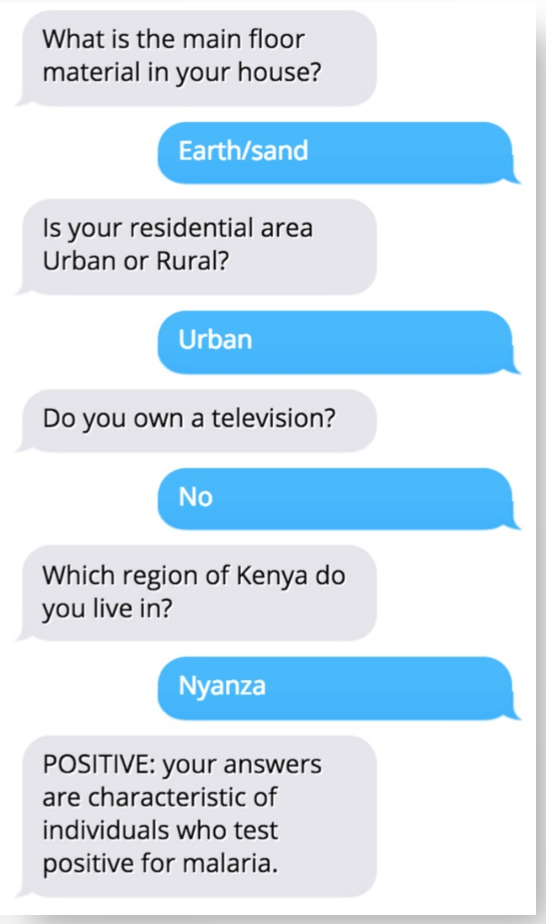
Li J, Monroe W, Ritter A, *et al.* (2017) Deep Reinforcement Learning for Dialogue Generation. doi:10.18653/v1/S17-1008

Aside – RL in dialogue



Chinaei H, Currie LC, Danks A, *et al.* (2017) Identifying and avoiding confusion in dialogue with people with Alzheimer's disease. *Computational Linguistics* **43**:377–406.

Aside – RL in dialogue

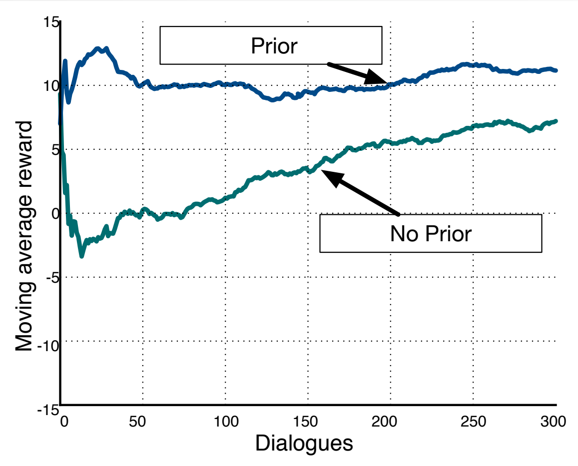


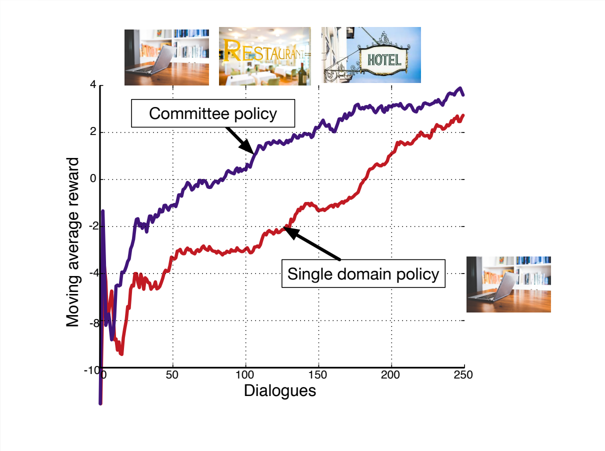
Rajpurkar *et al* (2017) Malaria Likelihood Prediction By Effectively Surveying Households Using Deep Reinforcement Learning. *ML4H*.

Aside – RL in dialogue

- Challenge 1 : data is limited in a particular domain
  Solution 1 : learn a distributed architecture with Gaussian priors

- Challenge 2 : Estimates of Q aren't shared across different domains
  Solution 2 : Use a Bayesian 'committee machine'



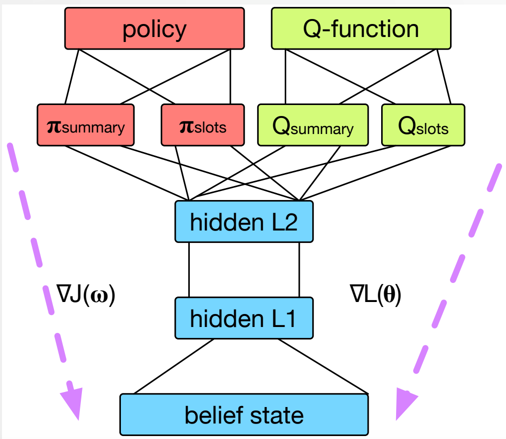


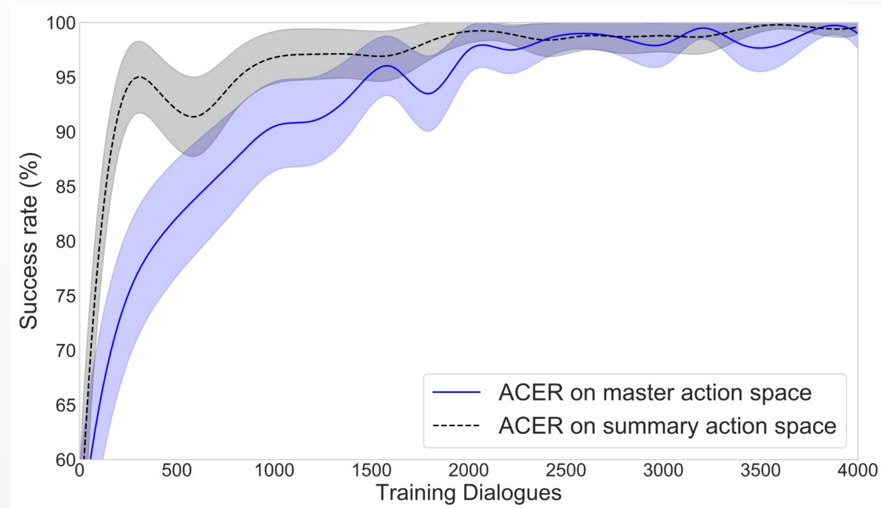
Gašić *et al* (2015) Distributed dialogue policies for multi-domain statistical dialogue management, ICASSP, https://ieeexplore.ieee.org/stamp/stamp.jsp?tp=&arnumber=7178997

Gašić *et al* (2015) Policy Committee for adaptation in multi-domain spoken dialogue systems, ASRU

Aside – RL in dialogue

- ACER learns an ‘off policy’ gradient ∇J and modified loss ∇L.
  - Avoid bias through replaying experience





The off-policy version of the Policy Gradient Theorem [30] is used to derive the gradients $\nabla_\omega J(\omega) \approx g(\omega)$:

$$g(\omega) = \sum_{b \in \mathbb{B}} d^\mu(b) \sum_{a \in \mathbb{A}} \nabla_\omega \pi(a|b) Q_\pi(b,a) \qquad (1)$$

$$\nabla L(\theta) = \nabla_\theta (Q^{ret} - Q_\theta(\mathbf{b},a))^2$$

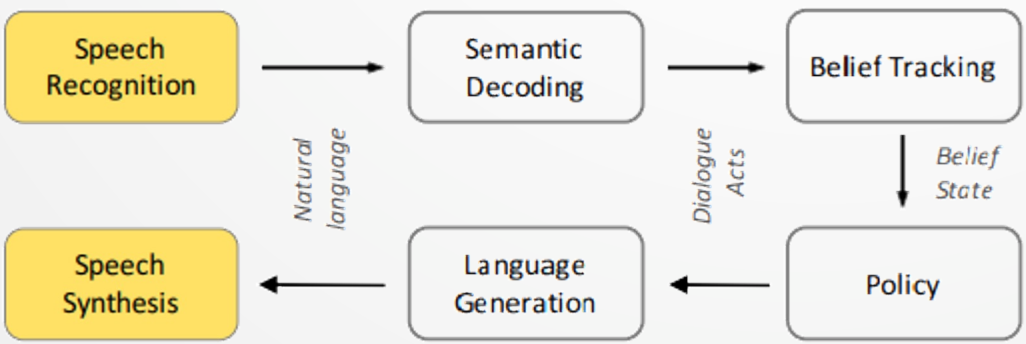
$$Q^{ret} = Q(\mathbf{b},a) + \mathbb{E}_{\mu}\left[\sum_{t \geq 0} \gamma^t \left(\prod_{s=1}^{t} \lambda \min\left(1, \rho(a_s|\mathbf{b}_s)\right)\right)(r_t + \gamma V(\mathbf{b}_{t+1}) - Q(\mathbf{b}_t, a_t))\right]$$

From Milica Gašić, Cambridge

Weisz, Budzianowski, Su, Gašić, (2018) Sample efficient deep reinforcement learning for dialogue systems with large action spaces, IEEE TASLP https://arxiv.org/pdf/1802.03753.pdf

PyDial toolkit

- *PyDial* (pydial.org) is an open-source Python toolkit for building dialogue systems. PyDial has 3 key components:
  - **Agent** module.
  - **User Simulation** module
    - used in, e.g., RL-based algorithms
  - **Evaluation** module.



Ultes, Rojas-Barahona, Su, *et al* (2017) PyDial: A Multi-domain Statistical Dialogue System Toolkit, ACL, https://www.aclweb.org/anthology/P17-4013

Corpora for dialogue

Metric	DSTC2	SFX	WOZ2.0	FRAMES	KVRET	M2M	MultiWOZ
# Dialogues	1,612	1,006	600	1,369	2,425	1,500	**8,438**
Total # turns	23,354	12,396	4,472	19,986	12,732	14,796	**115,424**
Total # tokens	199,431	108,975	50,264	251,867	102,077	121,977	**1,520,970**
Avg. turns per dialogue	14.49	12.32	7.45	**14.60**	5.25	9.86	13.68
Avg. tokens per turn	8.54	8.79	11.24	12.60	8.02	8.24	**13.18**
Total unique tokens	986	1,473	2,142	12,043	2,842	1,008	**24,071**
# Slots	8	14	4	**61**	13	14	25
# Values	212	1847	99	3871	1363	138	**4510**

Table 1: Comparison of our corpus to similar data sets. Numbers in bold indicate best value for the respective metric. The numbers are provided for the training part of data except for FRAMES data-set were such division was not defined.

- Ubuntu dialogue corpus and AMI Meeting corpus are also popular.

Budzianowski P, Wen T-H, Tseng B-H, *et al.* (2018) MultiWOZ - A Large-Scale Multi-Domain Wizard-of-Oz Dataset for Task-Oriented Dialogue Modelling http://arxiv.org/abs/1810.00278

The DSTC challenges

- DSTC challenge is held (almost) annually since 2012.
  - DSTC 1-6: "Dialogue State Tracking Challenges"
  - DSTC 7-present: "Dialogue Systems Technology Challenges"
- What "**dialogue state**" exactly means depends on the problem settings. For example:
  - In DSTC1 (dialogue about bus timetable information): The dialogue system infers the bus that the user wants to take.
  - In DSTC2 (dialogue about restaurant search): The **dialogue state** includes the slot/value attributes of the user goal, their search method, etc.

Recent DSTC challenges

DSTC10 @ AAAI-22 contains some more complex challenges. Here are the 3 tasks in Track 1 (Internet meme and open-domain dialogue)

1. **Text response modeling.** Generate coherent and natural text response based on chat history containing memes and texts.
2. **Meme Retrieval.** Based on the multimodal chat history, select a suitable meme to respond.
3. **Meme Emotion Classification.** Based on the multimodal chat history, predict the user's sentiment.

Recent DSTC challenges

Here are the 2 tasks in DSTC10 Track 2:

1. **Dialogue State Tracking**. Fill each slot with the estimated string.
2. **Conversation modeling** given knowledge access:
   2.1: Binary classification. Decide if continue the dialogue or trigger the "knowledge access"
   2.2: Select from knowledge sources.
   2.3: Given the conversation context and the "knowledge snippet", generate a system response.

EVALUATING DIALOGUE AGENTS

Lessons from HitchBot?

People (sometimes) like cute things that are smaller than they are.

Participant evaluation

- Human chats with model for 6 turns and rates 8 dimensions:
  - Avoiding repetition
  - Interestingness
  - Making sense
  - Fluency
  - Listening
  - Inquisitiveness
  - Humanness
  - Engagingness

"How often did this user say something which didn't make sense?"

"How much did you enjoy talking to this user?"

See, Abigail, et al. "What makes a good conversation? how controllable attributes affect human judgments." *NAACL* (2019).

Observer evaluation

Annotators look at two conversations and decide which is better, in terms of:

- **Engaging**: who would you prefer to talk to for a long conversation?
- **Interesting**
- **Humanness**
- **Knowledgeable**: If you had to say that one speaker is more knowledgeable and one is more ignorant, who is more knowledgeable?

Li, Margaret, Jason Weston, and Stephen Roller. "Acute-eval: Improved dialogue evaluation with optimized questions and multi-turn comparisons." NeurIPS 2019 Workshop on Conversational AI

Evaluation for task-based systems

- Task-based systems are evaluated by **task success**.
- For a slot machine: we can also use **slot error rate**.

"Make an appointment with Frank at 10:30 in LM 162"

Slot	Filler
Person	Frank
Time	11:30 AM
Location	LM 162

Slot error rate: 1/3

Task success: In the end, was the correct meeting added to the calendar?

Evaluation for task-based systems

- A more fine-grained method to evaluate task-based dialogue systems is **user satisfaction survey**.

Item	Description
TTS Performance	Was the system easy to understand?
ASR Performance	Did the system understand what you said?
Task Ease	Was it easy to find the message/flight/train you wanted?
Interaction Pace	Was the pace of interaction appropriate?
User Expertise	Did you know what you could say at each point?
System response	How often was the system slow in replying you?
Expected Behavior	Did the system work the way you expected it to?
Future Use	Do you want to use the system in the future?

Walker, Marilyn, Candace Kamm, and Diane Litman. "Towards developing general models of usability with PARADISE." *Natural Language Engineering* 6.3-4 (2000): 363-377.

Automatic evaluation

Automatic evaluation methods (e.g., BLEU) are usually *not* used to evaluate chatbots.

- They correlate poorly with human judgements.
- There are multiple correct ways for a dialogue to proceed, but BLEU just "checks *the* ground truth".

One current research direction is **adversarial evaluation**.

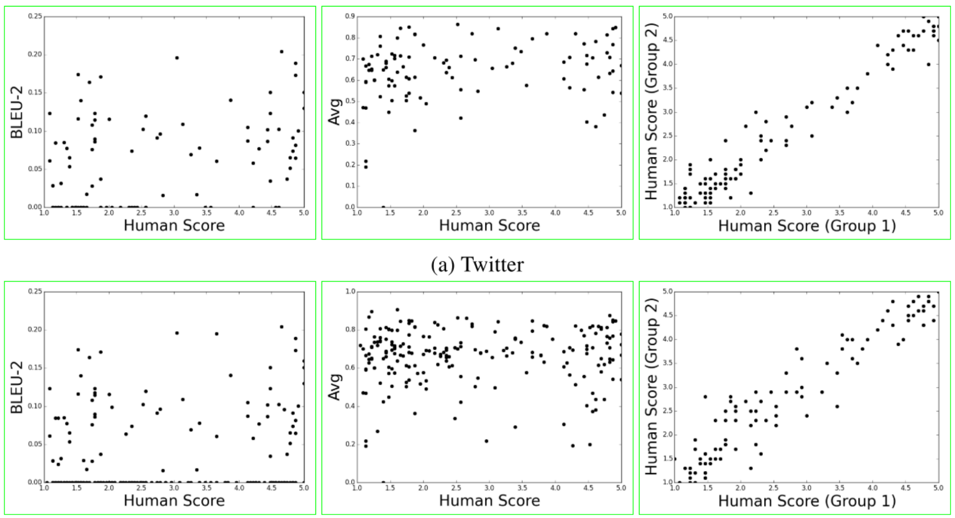
- This is inspired by the Turing Test.
- Train a "Turing-like" classifier to distinguish human responses vs. machine responses.
- Successful dialogue systems are good at fooling the evaluator.

Jurafsky and Martin, *Speech and Languages Processing*, 3rd version

Evaluating end-to-end dialogue

	Ubuntu Dialogue Corpus			Twitter Corpus		
	Embedding Averaging	Greedy Matching	Vector Extrema	Embedding Averaging	Greedy Matching	Vector Extrema
R-TFIDF	0.536 ± 0.003	0.370 ± 0.002	0.342 ± 0.002	0.483 ± 0.002	0.356 ± 0.001	0.340 ± 0.001
C-TFIDF	0.571 ± 0.003	0.373 ± 0.002	0.353 ± 0.002	0.531 ± 0.002	0.362 ± 0.001	0.353 ± 0.001
DE	$\mathbf{0.650 \pm 0.003}$	0.413 ± 0.002	0.376 ± 0.001	$\mathbf{0.597 \pm 0.002}$	0.384 ± 0.001	0.365 ± 0.001
LSTM	0.130 ± 0.003	0.097 ± 0.003	0.089 ± 0.002	0.593 ± 0.002	$\mathbf{0.439 \pm 0.002}$	$\mathbf{0.420 \pm 0.002}$
HRED	0.580 ± 0.003	$\mathbf{0.418 \pm 0.003}$	$\mathbf{0.384 \pm 0.002}$	$\mathbf{0.599 \pm 0.002}$	$\mathbf{0.439 \pm 0.002}$	$\mathbf{0.422 \pm 0.002}$

Table 2: Models evaluated using the vector-based evaluation metrics, with 95% confidence intervals.



(b) Ubuntu

Figure 1: Scatter plots showing the correlation between metrics and human judgements on the Twitter corpus (a) and Ubuntu Dialogue Corpus (b). The plots represent BLEU-2 (left), embedding average (center), and correlation between two randomly selected halves of human respondents (right).

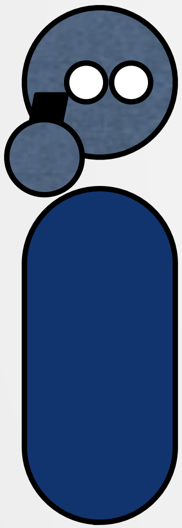
Liu C-W, Lowe R, Serban I V., *et al.* (2016) How NOT To Evaluate Your Dialogue System: An Empirical Study of Unsupervised Evaluation Metrics for Dialogue Response Generation. http://arxiv.org/abs/1603.08023

Security and privacy issues

- DNNs can generate toxic content.
  - Apparently, GPT3 etc., have ‘radicalized’ knowledge about extremist, racism, etc.
  - The knowledge might come from the training data.
- How to avoid them?
  - Filtering does not solve everything.
  - We need to be careful about the training data!
  - Domain-specific experts are needed.

Hutson, Matthew. "Robo-writers: The rise and risks of language-generating AI." *Nature* 591.7848 (2021): 22-25.

Dialogue summary



- Discourse & pragmatics provide huge research potential.
  - (i.e., many problems are unsolved)
- Developing dialogue technology requires integrating almost all NLP technologies.
- Dialogue technology also has wide application potential.

Goodbye