

# Controllable and Natural Response Generation for Virtual Assistants

**Andrew, Nick, Susana**  
**Mentor: Sina**

# Motivation

- Virtual Assistants have become good enough at responding to queries to have achieved fairly mainstream use (Alexa, Siri, Google Assistant, etc.)
- Despite this, none of these assistants could pose convincingly as a human
- Responses, while accurate, cannot be considered natural

## Google Assistant gifted a new, more natural-sounding voice

By Olivia Tambini September 18, 2019

Your Google Home speaker might sound a little different

Google says that this technology makes them "sound natural, with great pitch and pacing".

TECH | AMAZON | ARTIFICIAL INTELLIGENCE

## Amazon's Alexa gets a new longform speaking style

With more natural-sounding pauses, the style is intended for longer-form content like podcasts

By Kim Lyons | Apr 16, 2020, 5:45pm EDT

Use New Alexa Emotions and Speaking Styles to Create a More Natural and Intuitive Voice Experience

Catherine Gao Nov 26, 2019

Share: [f](#) [in](#) [t](#)

Game Skills Content Skills Design News



We're excited to introduce two new Alexa capabilities that will help create a more natural and intuitive voice experience for your customers. Starting today, you can enable Alexa to respond with either a happy/excited or a disappointed/empathetic tone in the US. Emotional responses are particularly relevant to skills in the gaming and sports categories. Additionally, you can have Alexa respond in a speaking style that is more suited for a specific type of content, starting with news and music. Speaking styles are curated text-to-speech voices designed to create a more delightful customer experience for specific content. For example, the news speaking style makes Alexa's voice sound similar to what you hear from TV news anchors and radio hosts. To learn more, check out our technical documentation for emotions [here](#) and speaking styles [here](#).

# Related Work on Controllable Text Generation

Plug and Play Language Models (Dathathri et al, 2020)

- Text generation with GPT-2
- Control topic and sentiment
- Challenge: there is a trade-off between controlling attributes and grammaticality

# Related Work on Controllable Text Generation

Plug and Play Language Models (Dathathri et al, 2020)

- Text generation with GPT-2
- Control topic and sentiment
- Challenge: there is a trade-off between controlling attributes and grammaticality

What Makes a Good Conversation (See et al, 2019)

- Conditional training and weighted decoding
- Control repetition, specificity, response relatedness, and question asking

**They are NOT conditioned on semantics.**

# Related Work on Virtual Assistant Response Generation

Many papers have come out in the past few weeks!

- Few-Shot Natural Language Generation by Rewriting Templates (Kale et al, 2020)
- Multi-Domain Dialogue Acts and Response Co-Generation (Wang et al, 2020)
- Few-shot Natural Language Generation for Task-Oriented Dialog (Peng et al, 2020)

**They all use MultiWOZ, SGD-NLG or other large annotated datasets.**

**No Controllability**

# What We Want to Do

What do we mean by naturalness?

- What Makes a Good Conversation? Balance
- Correct grammar, clean sounding speech, tonality, variance in speech
- More natural verb patterns

Controllability

- “Paraphrasing” with control on sentiment

At the end of the day, naturalness, sentiment, etc. are what humans believe they are. So we will do human evaluation.

# Synthetic Almond - High Level Overview

USER

AGENT

# Synthetic Almond - High Level Overview

USER

User Utterance

hello ! what are the food places

AGENT



# Synthetic Almond - High Level Overview

USER

AGENT

User Utterance

hello ! what are the food places

User State

```
$dialogue  
@org.thingpedia.dialogue.transaction.execute;  
now => @uk.ac.cam.multiwoz.Restaurant.Restaurant() => notify;
```

# Synthetic Almond - High Level Overview

USER

AGENT

User Utterance

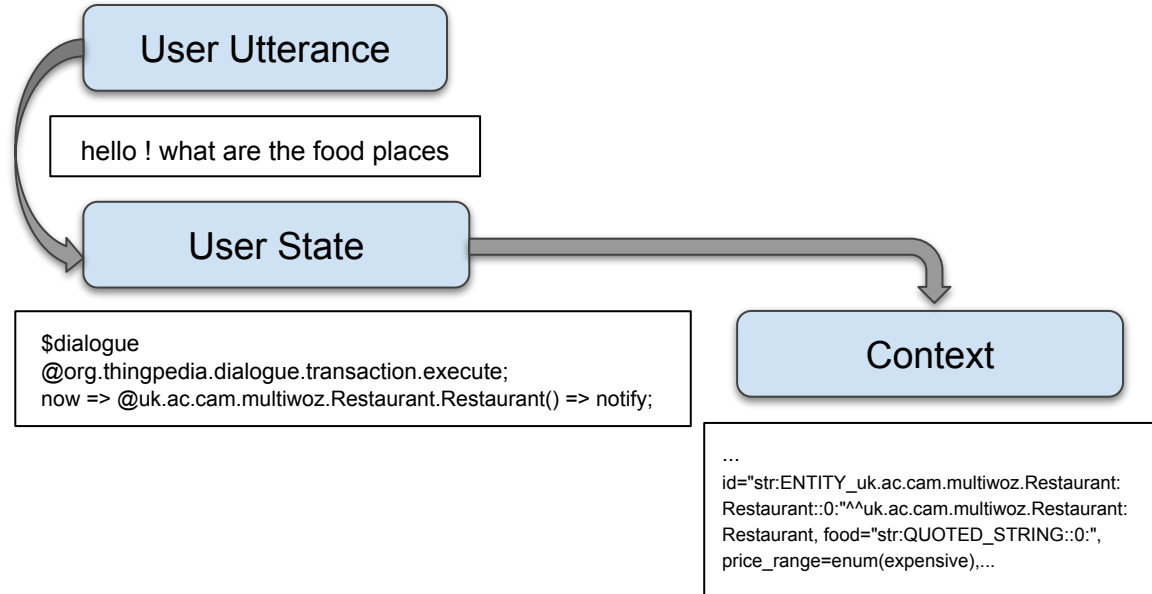
hello ! what are the food places

User State

```
$dialogue
@org.thingpedia.dialogue.transaction.execute;
now => @uk.ac.cam.multiwoz.Restaurant.Restaurant() => notify;
```

Context

```
...
id="str:ENTITY_uk.ac.cam.multiwoz.Restaurant:
Restaurant::0:"^uk.ac.cam.multiwoz.Restaurant:
Restaurant, food="str:QUOTED_STRING::0:",
price_range=enum(expensive),...
```



# Synthetic Almond - High Level Overview

USER

User Utterance

hello ! what are the food places

User State

```
$dialogue
@org.thingpedia.dialogue.transaction.execute;
now => @uk.ac.cam.multiwoz.Restaurant.Restaurant() => notify;
```

Context

```
...
id="str:ENTITY_uk.ac.cam.multiwoz.Restaurant:
Restaurant::0:"^uk.ac.cam.multiwoz.Restaurant:
Restaurant, food="str:QUOTED_STRING::0:",
price_range=enum(expensive),...
```

AGENT

Agent State

```
$dialogue
@org.thingpedia.dialogue.transaction.sys_search_question(food);
now =>
@uk.ac.cam.multiwoz.Restaurant.make_reservation(restaurant=$
?, book_time=$?, book_day=$?, book_people=$?);
```

# Synthetic Almond - High Level Overview

USER

User Utterance

hello ! what are the food places

User State

```
$dialogue
@org.thingpedia.dialogue.transaction.execute;
now => @uk.ac.cam.multiwoz.Restaurant.Restaurant() => notify;
```

Context

```
...
id="str:ENTITY_uk.ac.cam.multiwoz.Restaurant:
Restaurant::0:"^uk.ac.cam.multiwoz.Restaurant:
Restaurant, food="str:QUOTED_STRING::0:",
price_range=enum(expensive),...
```

AGENT

Agent State

```
$dialogue
@org.thingpedia.dialogue.transaction.sys_search_question(food);
now =>
@uk.ac.cam.multiwoz.Restaurant.make_reservation(restaurant=$
?, book_time=$?, book_day=$?, book_people=$?);
```

Agent Utterance

i have found several restaurants matching your request . what cuisine do you want ?

# Natural Language Generation

USER

User Utterance

hello ! what are the food places

User State

```
$dialogue
@org.thingpedia.dialogue.transaction.execute;
now => @uk.ac.cam.multiwoz.Restaurant.Restaurant() => notify;
```

Context

```
...
id="str:ENTITY_uk.ac.cam.multiwoz.Restaurant:
Restaurant::0:"^uk.ac.cam.multiwoz.Restaurant:
Restaurant, food="str:QUOTED_STRING::0:",
price_range=enum(expensive),...
```

AGENT

Agent State

```
$dialogue
@org.thingpedia.dialogue.transaction.sys_search_question(food);
now =>
@uk.ac.cam.multiwoz.Restaurant.make_reservation(restaurant=$
?, book_time=$?, book_day=? book_people=?);
```

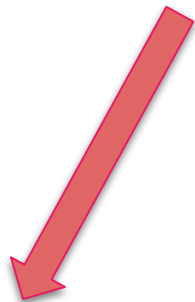
Agent Utterance

i have found several restaurants matching your request . what cuisine do you want ?

# Main Goal

AGENT

Agent Utterance



Naturalized Agent  
Utterance

I'm seeing many restaurants. What type  
of food would you like to eat today?

# Main Goal

AGENT

Agent Utterance

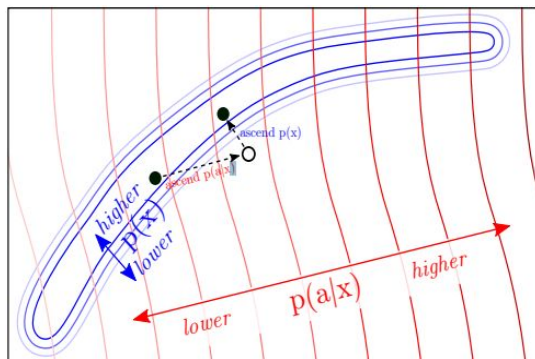
Naturalized Agent Utterance

Naturalized Agent Utterance With Sentiment

I'm seeing many restaurants. What type of food would you like to eat today?

I'm seeing some great restaurants! Can I ask which type of food you are looking for?

# Main Approach : Building on GPT-2



GPT-2

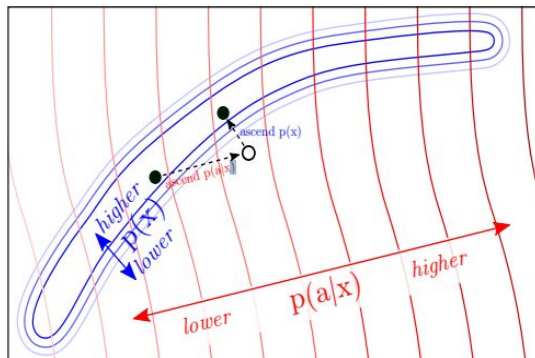
i see <name\_1> and <name\_2>. both are restaurant in the east part of town. <s> |

Agent Utterance (A)

Naturalized Agent Utterance



# Main Approach : Building on GPT-2



found

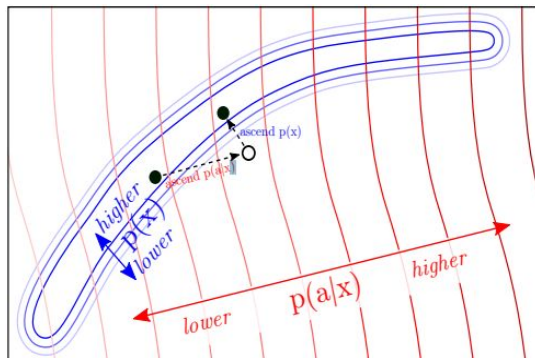
GPT-2

i see <name\_1> and <name\_2>. both are restaurant in the east part of town. <s> I

Agent Utterance (A)

Naturalized Agent Utterance

# Main Approach : Building on GPT-2



great

GPT-2

i see <name\_1> and <name\_2>. both are restaurant in the east part of town. <s> I found

Agent Utterance (A)

Naturalized Agent Utterance (With Sentiment)

# Proposed Demo

Interactive agent that we can talk to



Gio (User)

I want an Italian restaurant nearby.

I see Terún.  
It is a moderate Italian restaurant in the area.  
Would you like me to book the restaurant for you?



Almond (VA)



Gio (User)

No, give me something more luxurious.

I found this restaurant:  
name=Vina Enoteca,  
cuisine=Italian,  
price=expensive,  
location=Palo Alto



Almond (VA)

Gio didn't like my first suggestion. I need to reflect this in my **sentiment**

This is the second time, I should be **concise**

# Proposed Demo

Interactive agent that we can talk to

Must:

- Responses are natural
- User can select sentiments by checking a box



Gio (User)

I want an Italian restaurant nearby.



Gio (User)

No, give me something more luxurious.

I see Terún.  
It is a moderate Italian restaurant in the area.  
Would you like me to book the restaurant for you?



Almond (VA)

I found this restaurant:  
name=Vina Enoteca,  
cuisine=Italian,  
price=expensive,  
location=Palo Alto

Gio didn't like my first suggestion. I need to reflect this in my **sentiment**

This is the second time, I should be **concise**



Almond (VA)

# Proposed Demo

Interactive agent that we can talk to

Must:

- Responses are natural
- User can select sentiments by checking a box

Reach:

- User can select conciseness level



Gio (User)

I want an Italian restaurant nearby.



Gio (User)

No, give me something more luxurious.

I see Terún.  
It is a moderate Italian restaurant in the area.  
Would you like me to book the restaurant for you?



Almond (VA)

I found this restaurant:  
name=Vina Enoteca,  
cuisine=Italian,  
price=expensive,  
location=Palo Alto

Gio didn't like my first suggestion. I need to reflect this in my **sentiment**

This is the second time, I should be **concise**



Almond (VA)

# Proposed Demo

Interactive agent that we can talk to

Must:

- Responses are natural
- User can select sentiments by checking a box

Reach:

- User can select conciseness level



Gio (User)

I want an Italian restaurant nearby.



Gio (User)

No, give me something more luxurious.

I see Terún.  
It is a moderate Italian restaurant in the area.  
Would you like me to book the restaurant for you?



Almond (VA)

I found this restaurant:  
name=Vina Enoteca,  
cuisine=Italian,  
price=expensive,  
location=Palo Alto

Gio didn't like my first suggestion. I need to reflect this in my **sentiment**

This is the second time, I should be **concise**



Almond (VA)

How about Vina Enoteca? It will be an upscale dining experience!  
Should I make the reservation now?

# Schedule

| Week      | Finish  | Start  |
|-----------|---|--|
| 5/5-5/11  | Further research into related work  | Baselines: current model & feasibility check for PPLM<br>Connecting current model to PPLM  |
| 5/12-5/18 | Baselines: current model & PPLM<br>Continue working on model for natural output   | Initial round of human evaluation for naturalness of baseline  |
| 5/19-5/25 | Mini-evaluation of naturalized output <ul style="list-style-type: none"><li>- Reconsider which controllable aspects make output more engaging and “human” in our opinions</li></ul> | Sentiment control <ul style="list-style-type: none"><li>- Use plug and play model as reference</li><li>- Classifier for automatic metric</li></ul> |
| 5/26-6/1  | Sentiment control <ul style="list-style-type: none"><li>- May need to revisit previous steps if necessary</li></ul>   | If time permits, open scope from pos/neg sentiment to more generalized emotion   |
| 6/2-6/8   |   | <b>Final Evaluation</b>  |