



# Towards Assisting Human-Human Conversations

M.S. Thesis Defense by Tejas Nanaware

Department of Computer Science

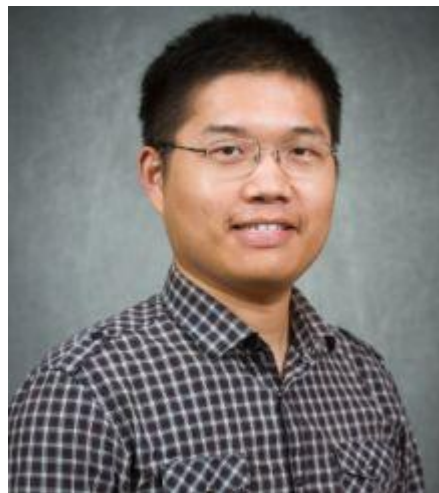
April 27, 2021



## Thesis Committee



Prof. Shlomo Argamon  
Academic Advisor



Prof. Kai Shu  
Thesis Committee Member

A horizontal bar with a teal segment on the left and an orange segment on the right.

## Presentation Overview

1. Research Idea and Introduction
2. Chatbot Selection and Literature Review
3. Data Gathering and IRB Application
4. Chatbot Interaction Framework and Deployment
5. Observations from the Research Study
6. Results and Analysis
7. Transfer Learning and Evaluation
8. Future Scope



## Research Idea and Introduction

- Assist humans in regular open-domain conversations and help overcome social anxiety
- Understand open-topic discussions to achieve successful conversations
- Augment conversational skills and reduce time taken while typing through a context aware AdvisorBot
- Research and find existing open-domain chatbots
- Determine linguistic features of textual 1-to-1 chats that predict conversational satisfaction
- Develop AdvisorBot that can assist individuals towards achieving more satisfying conversations
- Determine quality of AdvisorBot in terms of subjective aspects of conversations
- Main goal:
  - Overcome social anxiety and have fluent conversations
  - Be able to converse freely and have satisfactory conversations with other humans
  - Have faster replies and therefore avoid typing time



## Chatbot Selection and Literature Review

- Amazon Alexa Prize Challenge for achieving engaging human-chatbot conversation for 20 minutes
- Identify user's interests without explicitly asking them
- Human assistive chatbots for teaching underprivileged students for scaling expert knowledge
- Connecting strangers on Google Assistant and use it as the assistant
- Patents: Conversational assistance through Facebook Messenger and Guided customer support for client satisfaction for USAA (Insurance and financial services)
- Identify long pauses in conversations and suggest topics (year 2000)
- No existing research AdvisorBots that can help in conversations
- Open-sourced chatbots that have demonstrated skills in open-domain conversations
- GPT-3, Google Meena and Mitsuku
- Microsoft DialoGPT and Facebook BlenderBot



## Sample Responses from DialoGPT and BlenderBot

Sender	Message
User	Hi
DialoGPT	Hiya
BlenderBot	hello, how are you today? i just got back from a long day at the office.
User	Do you like Star Wars?
DialoGPT	I like Star Wars.
BlenderBot	i love star wars! it's my favorite franchise of all time. what about you?
User	Who is your favorite character from Star Wars?
DialoGPT	I don't know, I don't really like Star Wars.
BlenderBot	i don't really have a favorite character from star wars, but my favorite character is kylo ren.
User	It has been a long day
DialoGPT	I'm sorry.
BlenderBot	i'm sorry to hear that. what's going on in your life right now?
User	Would you like to go out for a dinner?
DialoGPT	I would love to!
BlenderBot	i'd love to, but i'm not sure if i'll be able to.

[illegible]



## Sample Responses from Rosie AIML

Sender	Message
User	<i>Hello</i>
Rosie	Hi nice to see you!
User	<i>Do you like Star Wars?</i>
Rosie	Yes, my favorite character is R2D2.
User	<i>Who is your favorite character from Star Wars?</i>
Rosie	It's good to be responsible and work hard.
User	<i>It has been a long day</i>
Rosie	Has it really?
User	<i>Would you like to go out for a dinner?</i>
Rosie	Sure, just bring me along on your mobile device.

Sender	Message
User	<i>I have a pen</i>
Rosie	When did you get it?
User	<i>I have an apple</i>
Rosie	How does it make you feel?
User	<i>pen pineapple apple pen</i>
Rosie	You can give me honest feedback.



## Data Gathering and IRB Application

- Human participation in the research program needs government approval
- Application to the IRB committee for getting consent for human participation
- Challenges for IRB application:
  - Improper suggestions by the chatbot
  - Delayed application process
  - Fast-track application as there is no severe life threatening scenario
- Limited research participation due to delay in response from IRB
- Consent based participation



A horizontal bar with a teal segment on the left and an orange segment on the right.

## What Would The Users Do?

1. Register and fill demographic information
2. Chat with other users while using the AdvisorBots and have an open-topic discussion
3. Provide feedback while chatting by quick responses that tell us status about the AdvisorBot and Conversation Quality
4. Leave chat and provide feedback using Post-chat questionnaire
5. Repeat and have multiple conversations

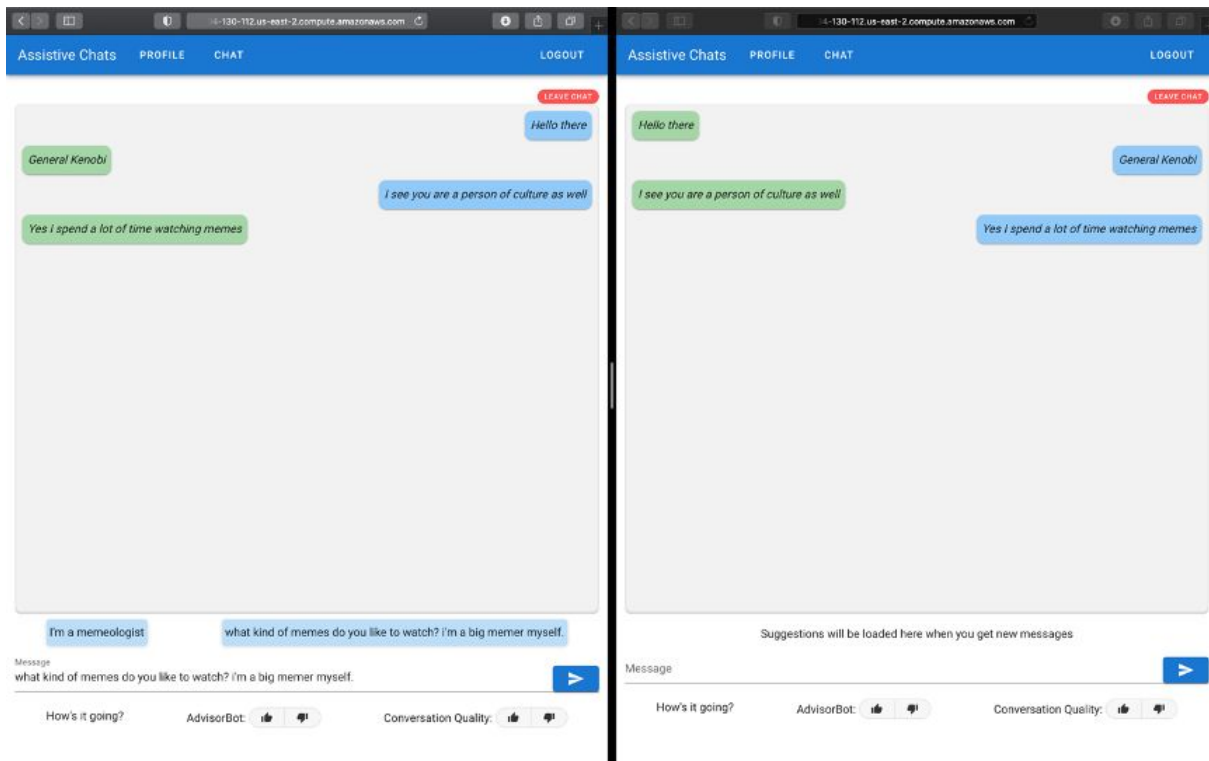


## Consent Based User Registration (Demographics)

The screenshot displays a web browser window with a URL bar showing 'ec2-3-134-130-112.us-east-2.compute.amazonaws.com'. The page has a dark blue header with 'Assistive Chats' on the left and 'LOGIN REGISTER' on the right. The main content area is divided into two columns. The left column contains a 'Register' button and a series of input fields for 'First Name', 'Last Name', 'Email', 'Password', and a section for demographic questions: 'Please answer these demographic questions', 'Age', 'Gender', 'Race / Ethnicity', 'Native Language', 'Nationality', and 'Level of Education'. The right column is partially obscured by a white modal dialog box. This dialog box has the title 'Do you agree with this and provide your consent?' and contains the following text: 'You are being asked to participate in a research study. Participation in this research study is voluntary and you may withdraw from the study at any time without penalty. Your identity will be coded to ensure confidentiality. The purpose of this study is to understand and assist text-based human conversations by identifying key parameters behind a successful conversation and using chat assistants that can help to maintain the conversational flow. As a participant in this study, you will be signing up on the online platform and will be connected to another participant or a chatbot and asked to have a conversation. You can disconnect at any time and may reconnect to converse with several participants. During this experience, you will be asked to complete questionnaires about your perceptions of your conversation and your experience using the platform. Apart from that, at the start, you will be asked to fill a demographic questionnaire. While having a conversation, you may be provided with conversational assistance that may suggest possible next sentences or phrases related to the conversation. You can use or modify these suggestions or ignore them if you choose. The conversations and the questionnaires will be stored and may be shared with other researchers, but your identity will be coded and kept separate from that data, to ensure your confidentiality. However, you are advised not to have conversations that may reveal your identity. Although participation involves no future obligation, you may be contacted for future assessment sessions and may have the opportunity to participate in additional research if you so wish. Risks: The risks that may be caused by this research study are: 1. Discomfort with release of confidential information or disability or impairment during conversations. 2. Participants may feel inhibited or uncomfortable since the chats are being recorded. 3. Participants may feel uncomfortable about answering the questionnaires or due to irrelevant or inappropriate chat suggestions. 4. Participants may feel uncomfortable being advised by a bot about what to say in a chat. We are committed to respecting your privacy and to keep your personal information confidential.'



# Chat Interface with AdvisorBot Suggestions





# Post-chat Questionnaires (User Feedback)

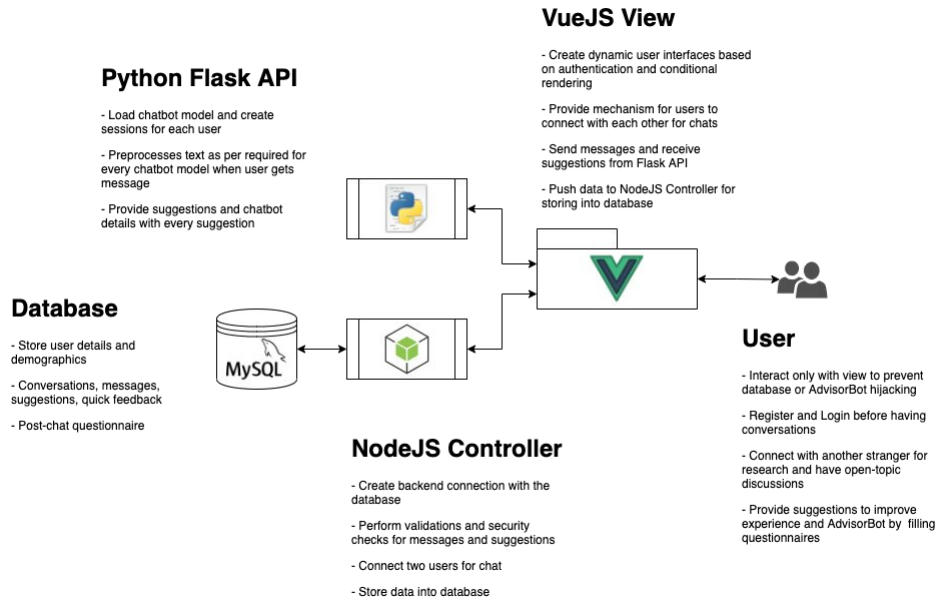
The screenshot shows a web browser window with a questionnaire titled "How much do you agree / disagree with these statements". The interface includes a blue header bar with "Assistive Chats", "PROFILE", "CHAT", and "LOGOUT" links. The questionnaire consists of ten statements, each with a horizontal slider from "Disagree" to "Agree". The sliders are color-coded: orange for the first two statements and green for the remaining eight. The statements and their approximate slider positions are as follows:

Statement	Disagree	Agree
The conversation was comfortable and flowed well.	Slider at ~40%	
There were times when I felt uncomfortable during conversation.	Slider at ~55%	
My conversational partner understood me very well.	Slider at ~75%	
I understood my conversational partner very well.	Slider at ~90%	
The conversational assistance was helpful.	Slider at ~85%	
The conversational assistance was distracting or annoying.	Slider at ~15%	
The conversational assistance was able to understand the context and was able to provide accurate suggestions.	Slider at ~65%	
The conversational assistance will help in creating and maintaining the flow in the conversation.	Slider at ~65%	
I enjoyed the overall experience.	Slider at ~95%	
I would recommend someone to participate in this research study.	Slider at ~95%	

Below the statements, there is a section titled "Please answer these questions in a couple of lines" with a text input field for "What were the best parts of the conversations?". A note below the input field says "Please answer in at least two or three lines".



## Chatbot Interaction Framework



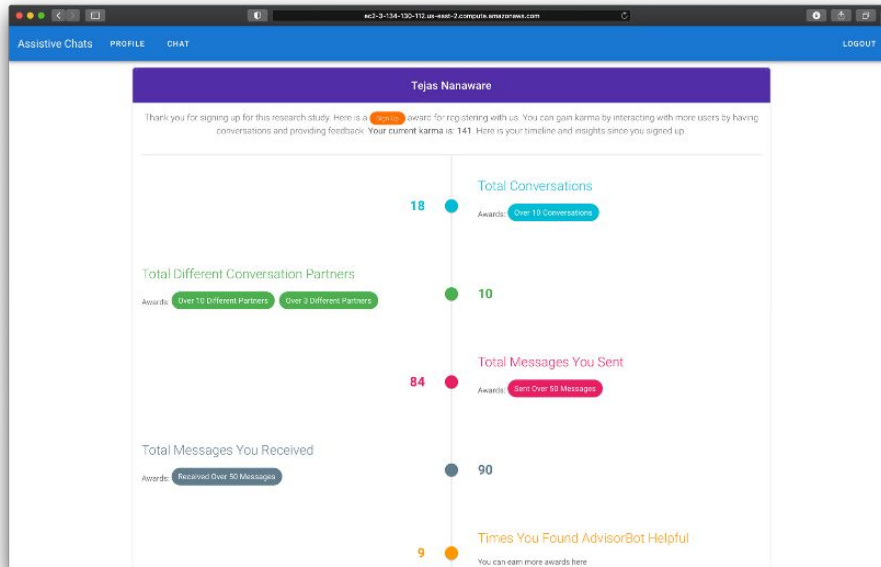


## Keeping the Platform Captivating

- Several chat platforms are already available so why use this platform?
- We need it for gathering data for the research
- Users can get their friends involved to test the scenarios of the chats
- Have several chats that can be good / bad in terms of suggestions to get to know how AdvisorBot reacts to different contexts: internet references, movies, anime, or extreme language
- Understand how people react to several scenarios
- Provide awards to the users based on number of conversations, AdvisorBot clicks, different conversational partners, times they filled the questionnaires
- Use karma points to engage users to have multiple conversations with several conversational partners and make sure that they are providing feedback



## Interactive UX Through Awards and Karma



$$\text{Karma} = \text{Round} \left( \left( \frac{\text{Count}(\text{messagesSent}) + \text{Count}(\text{messagesReceived})}{\left( \frac{\text{Count}(\text{conversations})}{\text{Count}(\text{differentConversationalPartners})} \right)} \right) + (\text{Count}(\text{timesBotHelpful}) \times \text{Count}(\text{goodConversation})) - (\text{Count}(\text{timesBotNotHelpful}) \times \text{Count}(\text{badConversation})) \right)$$



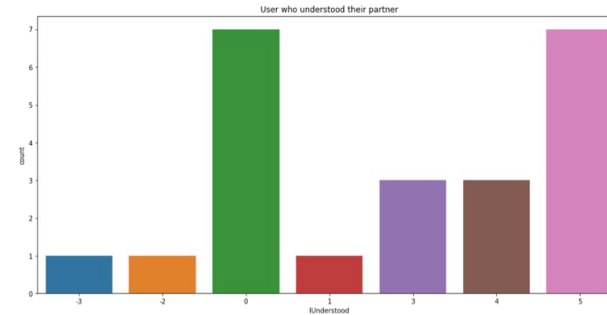
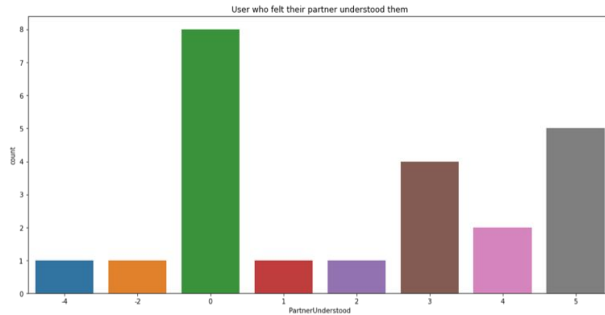
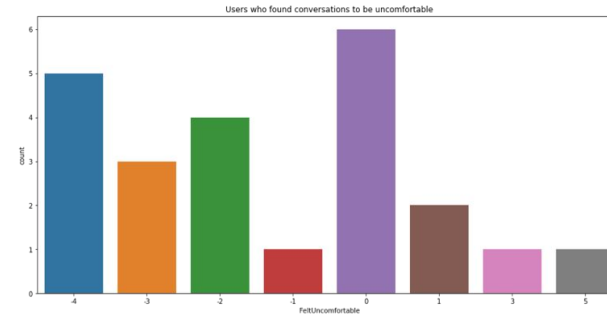
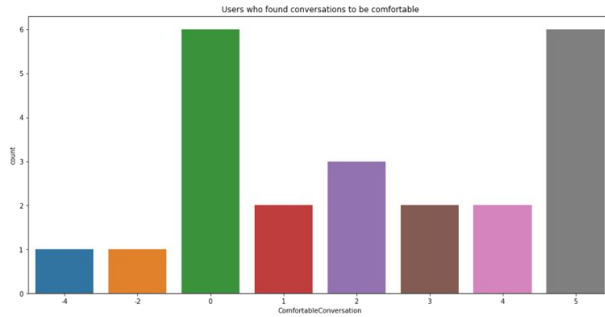
## Users' Observations from Research Study

- AdvisorBot can be helpful but suggestions and the suggestion lag caused it to be distracting
- High accuracy rating for AdvisorBot suggestions but low rating for AdvisorBot being helpful
- AdvisorBot understood common internet references which are required in open-domain chats
- AdvisorBot could understand context and provide relevant suggestions based on the context but could not understand when to change topic and what should be that topic
- Application needed valid SSL certificate and masked URL because users would get deceptive website warning by their web browser
- Multi-lingual context understanding for terms popular among certain demographics



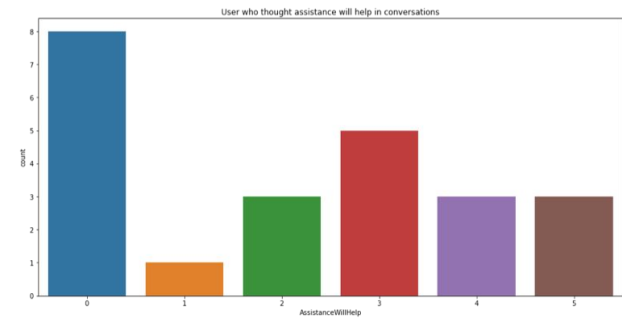
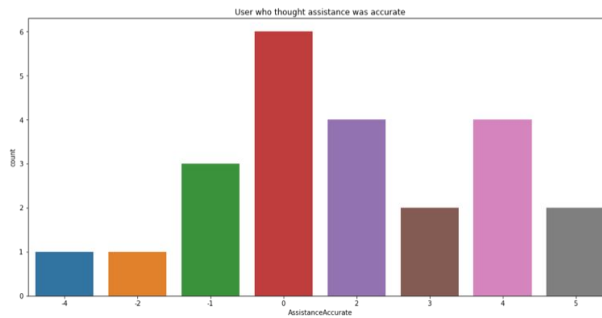
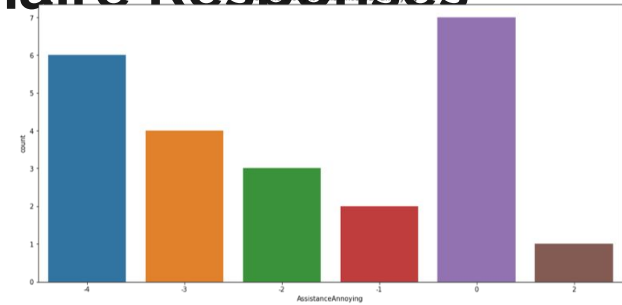
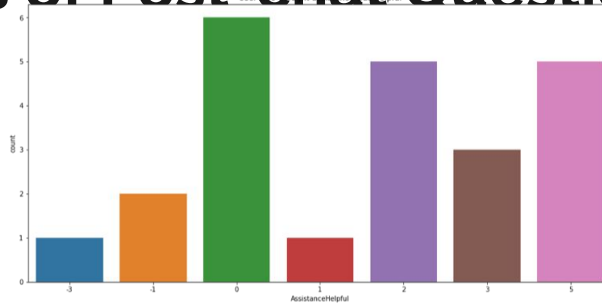


## Statistics of Post-chat Questionnaire Responses



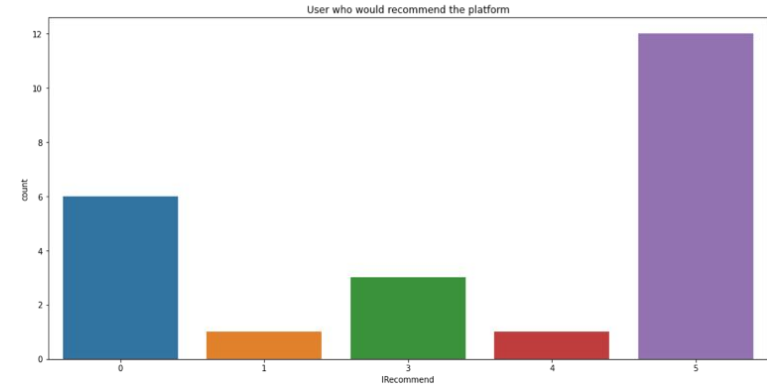
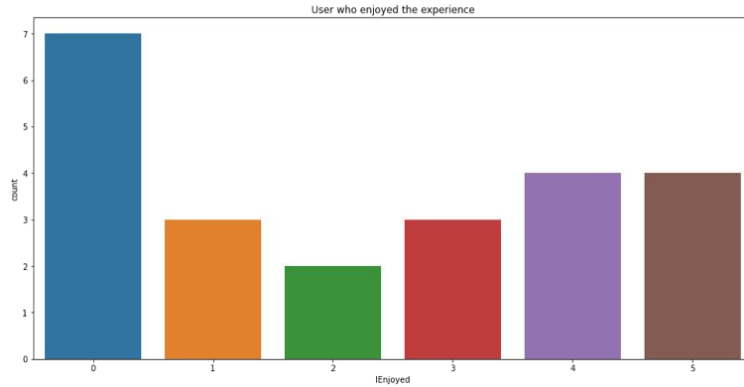


## Statistics of Post-chat Questionnaire Responses





## Statistics of Post-chat Questionnaire Responses





## Observations and Results from the Research Study

- Total messages sent were 429 out of which suggestions by the chatbots were clicked only 82 times
- Blenderbot was clicked 49 times as compared to DialoGPT which was clicked 33 times
- Average suggestion load time for AdvisorBot suggestions was 19.59s with maximum time of 69.62s
- 11 cases where users felt conversation was going well and 3 cases where users felt conversation was not going well. Out of which there were total 97 messages before the conversation rating and 0 messages prior to voting conversation as poor
- 17 cases where users felt AdvisorBot quality was good and 6 instances where users felt AdvisorBot quality as poor. There were 137 instances of messages prior to voting AdvisorBot quality as good and 36 instances of messages prior to voting AdvisorBot quality as bad



## BERT.CosineSimilarity (Message Sent, Suggestion)

- Calculate cosine similarity between the message sent and suggestion provided for analyzing textual patterns
- Find if the suggestion provided had almost similar meaning and intention as that of the response
- 39.81% had cosine similarity of over 0.8 and 18.51% had cosine similarity of below 0.2
- No negative cosine similarity suggesting messages sent and suggestions were not strongly opposites
- Messages where AdvisorBot was clicked and the message was sent as is obviously had cosine similarity of 1
- Most hits for AdvisorBot suggestions was when users received common greetings or expressed interests towards a certain topic
- This can help to understand context switching to a different topic by determining when cosine similarity falls down and then provide a different topic for having a conversation that can raise cosine similarity level again



## High Cosine Similarity Messages

	previous_message	message_sent	suggestion	similarity_score
16	Test	i have a test tomorrow that i am going to test...	i have a test tomorrow that i am going to test...	1.0
14	Test	i have a test tomorrow that i am going to test...	i have a test tomorrow that i am going to test...	1.0
175	i haven't seen it yet, but i've heard it's pre...	It's a great show.	It's a great show.	1.0
174	Sports Cars	i love sports cars. they are the most popular ...	i love sports cars. they are the most popular ...	1.0
165	How's the weather?	it's a bit chilly, but i'm used to it. how abo...	it's a bit chilly, but i'm used to it. how abo...	1.0
2	hello, how are you today? i just got back from...	I'm doing well, thanks for asking.	I'm doing well, thanks for asking.	1.0
166	yeah, i'm used to it too. it's a bit chilly he...	I'm in the UK, so it's not that cold.	I'm in the UK, so it's not that cold.	1.0
113	original poster. i'm so proud of him. he's suc...	I'm so glad you're proud of him.	I'm so glad you're proud of him.	1.0
23	hello, how are you today? i just got back from...	i'm doing well. i just got back from a long da...	i'm doing well. i just got back from a long da...	1.0
0	Hello	hello, how are you today? i just got back from...	hello, how are you today? i just got back from...	1.0



## Low Cosine Similarity Messages

	previous_message	message_sent	suggestion	similarity_score
191	Biach	Mofo	hi, how are you today? i just got back from a ...	0.032428
73	Do you like star wars?	but this app seems require attention.	that sounds like fun. what kind of anime are y...	0.056042
197	Autobots unite!	Sloooow	hi, how are you today? i just got back from a ...	0.072949
202	Hiya	yoooooooooooo	hi, how are you today? i just got back from a ...	0.085741
72	Do you like star wars?	but this app seems require attention.	I'm planning to stay at home and watch some anime	0.087415
116	Hey, you're not OP.	I'm not OP?	hello, how are you today? i just got back from...	0.088498
186	Mofo	Kasa kai	hello, how are you today? i just got back from...	0.090067
115	Hiya	Hey, you're not OP.	hello, how are you today? i just got back from...	0.101641
206	hello, how are you today? i just got back from...	What is OP?	i'm doing well. i just got back from a long da...	0.101690
56	The weather's been okay.	What are your interests?	I'm in the middle of a thunderstorm.	0.104777



## Transfer Learning BlenderBot

- Utilize conversations and messages to improve BlenderBot
- Treat every conversation as an episode where users participated in a multi-turn conversation
- Use reply by the second user as the label to the message sent by the first user
- Split data into training and testing (80 - 20)
- 8 Layers with embedding vector size of 512, GeLU Activation, Adam Optimizer with  $lr=1e-6$
- Model learns slightly from negative values as negative gradient shifts to positive in GeLU
- Learn positional embeddings to differentiate between "I **like** the movie but I **do not like** the cast" and "I **do not like** the movie by I **like** the cast"
- Use original BlenderBot model as the baseline model for comparing with the transfer learning model
- Compare and evaluate the models by Accuracy, Token Accuracy, BLEU-4, F-1 and **Perplexity**





## Comparison of Baseline & Transfer Learning Models

Dataset	Model	Accuracy	BLEU-4	F-1	Perplexity	Token Accuracy
Training Set	Baseline	0.00557	0.02868	0.13124	8.97184	0.54156
	New Model	0.00279	0.03004	0.1624	7.91383	0.56716
Testing Set	Baseline	0.06667	0.21166	0.37304	3.83843	0.65686
	New Model	0.06667	0.1589	0.30067	3.61871	0.69608



## Correct Evaluation Metric? - Perplexity

- Perplexity is the right measure for this task
- Perplexity measures uncertainty in assigning probabilities, or how well a model predicts given a probability distribution
- Perplexity looks after entire vocabulary of the model and covers wide range in predictions than BLEU-4 or F-1 scores
- BLEU-4 or F-1 scores compare similarities between suggestion and messages and how closely related they are without considering entire vocabulary or probability distribution
- This makes it better understand context history and create uncertain suggestions to keep conversation running smoothly
- E.g. When user got a message as “I’m doing well, thanks for asking.”, baseline model suggested “that’s good to hear. I’m glad you’re having a good day.”, whereas transfer learning model suggested “what kind of work do you do?” referencing to an older message with same context



## Suggestions & Replies for Transfer Learning Model

Sender	Message
Message	But i like dogs
Suggestion	<i>i like dogs too</i>
User Reply	<b>what kind of dogs do you like? there are so many different kinds of dogs out there.</b>
Message	Do you like the tv show friends?
Suggestion	<i>i do . i like it a lot .</i>
User Reply	<b>Yes</b>
Message	Yes
Suggestion	<i>i like it too</i>
User Reply	<b>Do you think ross and rachel were on a break?</b>
Message	Do you think ross and rachel were on a break?
Suggestion	<i>i think so . i think it 's a good show .</i>
User Reply	<b>i don't think so. i think it's more likely that they were together for a while.</b>
Message	Chai pilo
Suggestion	<i>chai pilo</i>
User Reply	<b>Hiya!</b>

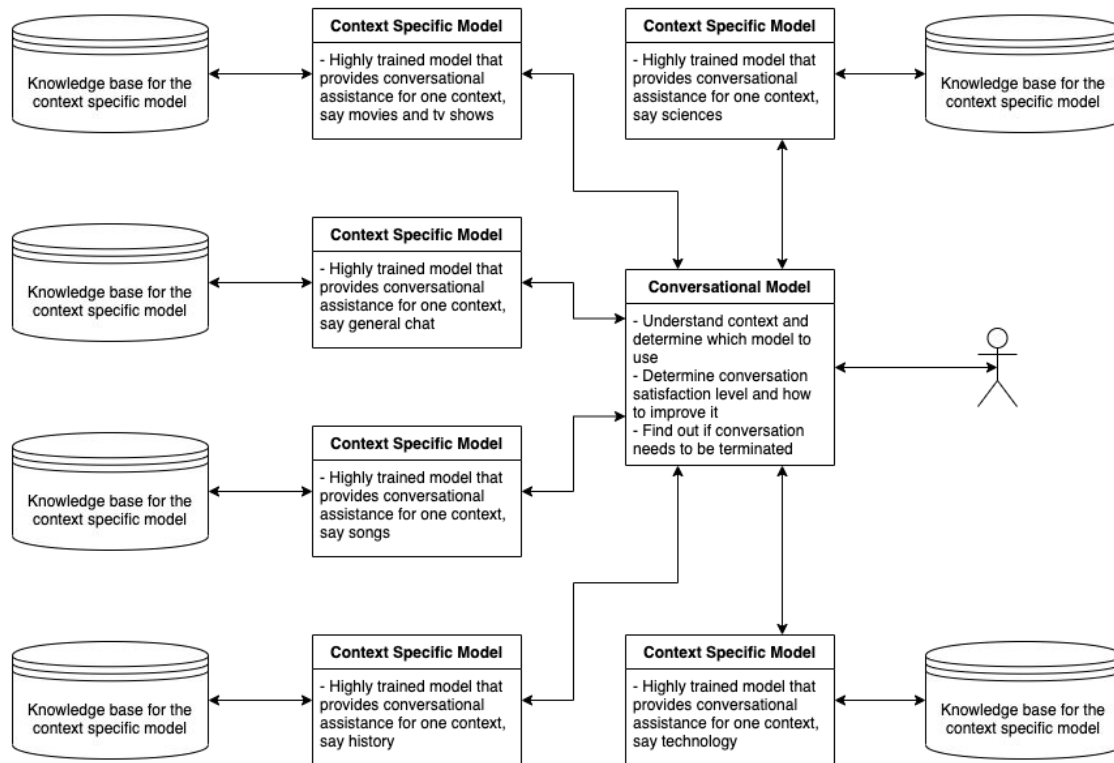
A horizontal bar with a teal segment on the left and an orange segment on the right.

## Future Scope

- Research study was limited and gathering more data will help understand open-topic discussions and extent to which AdvisorBot can provide suggestions
- Look for open-topic discussions dataset that can help understand context switching strategy performed by humans
- Analyze how long humans communicate on a certain topic and the need to change the topic and determine next topic to discuss
- Understand conversation termination strategy to prevent extending a conversation that makes the partner feel uncomfortable
- Develop strategies where AdvisorBot can convert an unsatisfactory conversation to satisfactory



# Future Scope Distributed Models





Questions?



## References

---

- [1] D. Adiwardana, M.-T. Luong, D. R. So, J. Hall, N. Fiedel, R. Thoppilan, Z. Yang, A. Kulshreshtha, G. Nemade, Y. Lu and Q. V. Le, "Towards a Human-like Open-Domain Chatbot," arXiv preprint arXiv:2001.09977, 2020.
- [2] I. Papaioannou, A. C. Curry, J. L. Part, I. Shalymov, X. Xu, Y. Yu, O. Dušek, V. Rieser and O. Lemon, "An Ensemble Model with Ranking for Social Dialogue," arXiv preprint arXiv:1712.07558, 2017.
- [3] Pandorabots, "Rosie (chatbot base)," Pandorabots, [Online]. Available: <https://github.com/pandorabots/rosie>.
- [4] S. Worswick, "Mitsuku wins Loebner Prize 2018!," Pandorabots, 13 September 2018. [Online]. Available: <https://medium.com/pandorabots-blog/mitsuku-wins-loebner-prize-2018-3e8d98c5f2a7>.
- [5] T. B. Brown, B. Mann, N. Ryder, M. Subbiah, J. Kaplan, P. Dhariwal, A. Neelakantan, P. Shyam, G. Sastry, A. Askell, S. Agarwal, A. Herbert-Voss, G. Krueger, T. Henighan, R. Child and Ra, "Language Models are Few-Shot Learners," arXiv preprint arXiv:2005.14165, 2020.
- [6] Y. Zhang, S. Sun, M. Galley, Y.-C. Chen, C. Brockett, X. Gao, J. Gao, J. Liu and B. Dolan, "DialogPT: Large-Scale Generative Pre-training for Conversational Response Generation," arXiv preprint arXiv:1911.00536, 2019.
- [7] S. Roller, E. Dinan, N. Goyal, D. Ju, M. Williamson, Y. Liu, J. Xu, M. Ott, K. Shuster, E. M. Smith, Y.-L. Boureau and J. Weston, "Recipes for building an open-domain chatbot," arXiv preprint arXiv:2004.13637, 2020.
- [8] J. Qin, Z. Ye, J. Tang and X. Liang, "Dynamic Knowledge Routing Network for Target-Guided Open-Domain Conversation," in Proceedings of the AAAI Conference on Artificial Intelligence, 2020.
- [9] Z. Yao, Y. Zhang, X. Li, J. Gao, M. Galley, C. Brockett, H. Sun and B. Dolan, "IEC: Towards Interest-Eliciting Neural Conversational Agents".
- [10] V. K. Cannanure, T. X. Brown and A. E. Ogan, "DIA: A Human AI Hybrid Conversational Assistant for Developing Contexts," in The 2020 International Conference on Information and Communication Technologies and Development, New York, NY, USA, 2020.
- [11] R. F. Daniel, Y. Talmor, A. Lebrun, L. N. Landowski, D. Demir, J. H. Goldberg and W. Blandin, "Providing personal assistant service via messaging". United States of America Patent US Patent 10,686,738, 16 June 2020.
- [12] J. J. Maestas, "AI assistant for interacting with customers across multiple communication modes". United States of America Patent US Patent 10,708,424, 7 July 2020.

A short horizontal bar with a teal segment on the left and an orange segment on the right.

# Thank You & Acknowledgements

Thanks to Illinois Institute of Technology, Prof. Shlomo Argamon, Prof. Kai Shu and my family and friends without whom this project would not have been completed