

# From Theory to Practice: Recent Advances and Future Prospects in Computational Politeness in Natural Language Processing



**Priyanshu Priya**

[priyanshu528priya@gmail.com](mailto:priyanshu528priya@gmail.com)



**Dr. Mauajama Firdaus**

[mauzama.03@gmail.com](mailto:mauzama.03@gmail.com)



**Dr. Asif EKbal**

[asif.ekbal@gmail.com](mailto:asif.ekbal@gmail.com)

# Outline



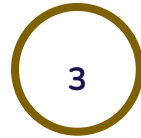
**Introduction to  
Politeness and  
Related Theories**



**Computational  
Politeness in NLP**



**Politeness for Social  
Good Applications**



**Trends and Issues in  
Computational  
Politeness**



**Conclusion and  
Future Directions**



# What is Politeness

Being marked by or showing consideration for others and observance of accepted social usage - [Free Dictionary](#)

Having or showing good manners and respect for the feelings of others - [Oxford Learner's Dictionary](#)

Showing or characterized by correct social usage

Marked by an appearance of consideration, tact, deference, or courtesy

Marked by a lack of roughness or crudities

[https://www.merriam-webster.com  
L](https://www.merriam-webster.com/L)

Politeness encompasses behavior that is socially correct and acceptable

# Example of Politeness can be

Student: Excuse me, Mr. Buckingham, but can I talk to you for a minute?

Teacher: Sure.

-----

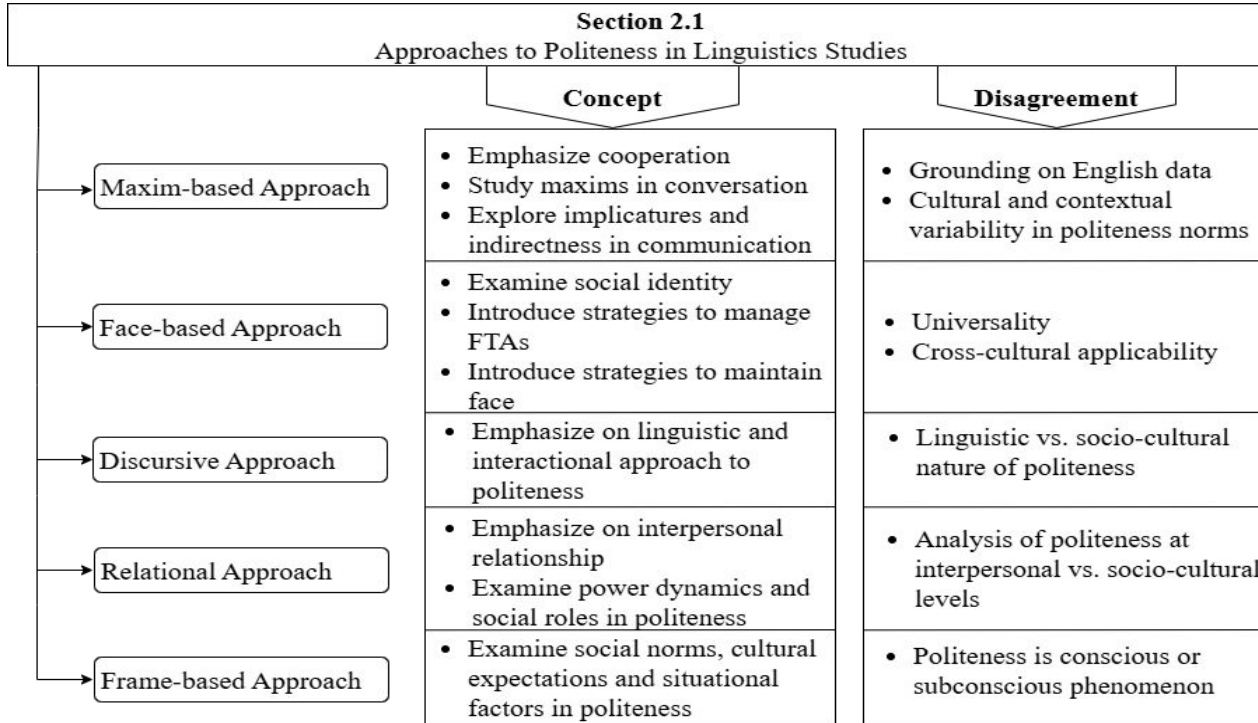
Friend 1: Hey Katie, got a minute?

Friend 2: Yes

# Linguistics Politeness

- Maintain[ing] the social equilibrium and the friendly relations which enable us to assume that our interlocutors are being cooperative in the first place (**Leech 1983**)
- Politeness, like formal diplomatic protocol (for which it must surely be the model), presupposes that potential for aggression as it seeks to disarm it, and makes possible communication between potentially aggressive parties (**Brown and Levinson 1987**)
- Means of minimizing confrontation in discourse - both the possibility of confrontation occurring at all, and the possibility that a confrontation will be perceived as threatening (**Lakoff 1989**)
- All that is really being claimed is that people employ certain strategies for reasons of expediency – experience has taught us that particular strategies are likely to succeed in given circumstances, so we use them (**Thomas 2014**)
- Being tactful, modest and nice to other people (**Yule 2020**)

# Approaches to Politeness in Linguistic Studies



# Approaches to Politeness in Linguistic Studies

Section 2.1 Approaches to Politeness in Linguistics Studies		
	Concept	Disagreement
Maxim-based Approach	<ul style="list-style-type: none"><li>• Emphasize cooperation</li><li>• Study maxims in conversation</li><li>• Explore implicatures and indirectness in communication</li></ul>	<ul style="list-style-type: none"><li>• Grounding on English data</li><li>• Cultural and contextual variability in politeness norms</li></ul>
Face-based Approach	<ul style="list-style-type: none"><li>• Examine social identity</li><li>• Introduce strategies to manage FTAs</li><li>• Introduce strategies to maintain face</li></ul>	<ul style="list-style-type: none"><li>• Universality</li><li>• Cross-cultural applicability</li></ul>
Discursive Approach	<ul style="list-style-type: none"><li>• Emphasize on linguistic and interactional approach to politeness</li></ul>	<ul style="list-style-type: none"><li>• Linguistic vs. socio-cultural nature of politeness</li></ul>
Relational Approach	<ul style="list-style-type: none"><li>• Emphasize on interpersonal relationship</li><li>• Examine power dynamics and social roles in politeness</li></ul>	<ul style="list-style-type: none"><li>• Analysis of politeness at interpersonal vs. socio-cultural levels</li></ul>
Frame-based Approach	<ul style="list-style-type: none"><li>• Examine social norms, cultural expectations and situational factors in politeness</li></ul>	<ul style="list-style-type: none"><li>• Politeness is conscious or subconscious phenomenon</li></ul>

# Face-based Approach to Politeness: Brown and Levinson Politeness Theory



FACE

Everyone has **public self-image** which has relation to social and emotional sense of self and expects everyone else to recognize (**Brown and Levinson 1987**)



# Politeness Definition by Brown and Levinson

Politeness in interaction can be defined as the means employed to show awareness of another person's face



# Face Wants

- People generally behave as if their expectations concerning their public self-image, or their face wants will be respected.
- If a speaker says something that represent a threat to another person's expectations regarding self-image, this is **face-threatening act (FTA)**.

# Types of Face: Positive Face

**Positive face** is the need to be accepted, even liked by others, to be treated as a member of the group and to know that his or her wants are shared by others.

Your friend asks for a ride to the airport.

**Positive face needs:** You think, I better take him because I want him to like me and I want the reputation of being a reliable person (Goffman 1965)

# Types of Face: Negative Face

**Negative face** is the need to be independent to have a freedom of action and not to be imposed by others.

Your friend asks for a ride to the airport.

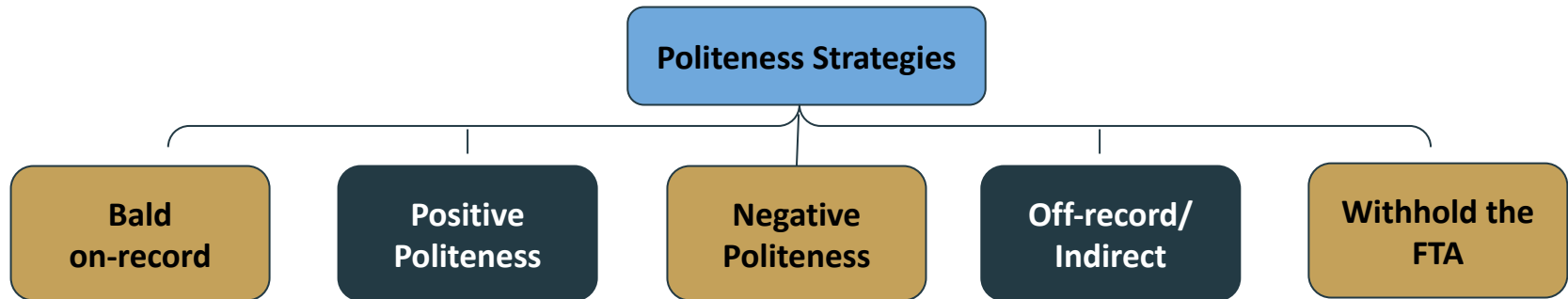
**Negative face needs:** You think, this is not favorable, I don't feel like driving this guy to the airport. I have other stuff that I could be doing like sleeping, or saving the gasoline in my car. He can find his own ride.  
(Goffman 1965)

Positive face is the need to be connected  
and negative face is the need to be  
independent

Politeness is defined as a necessary element to mitigate acts that threaten either positive or negative face of an individual.

# Politeness Strategies

Politeness strategies are used to formulate messages in order to save the hearer's positive and negative face when face-threatening acts are inevitable or desired (Brown and Levinson 1987).



# Bald on-record

- It does not attempt to minimize the threat to the hearer's face, although there are ways that bald on-record politeness can be used in trying to minimize face-threatening acts implicitly, such as giving advice in a non-manipulative way.
- Often using such a strategy will shock or embarrass the hearer, and so this strategy is most often utilized in situations where the speaker has a close relationship with the listener, such as family or close friends.
- Perform the FTA in the most direct, clear, unambiguous and concise way possible
- **Example:**
  - (i) Please make a cup of tea. (said to a close friend)
  - (ii) Don't forget to clean the blinds!

# Positive Politeness

- It aims to mitigate the potential harm to the hearer's positive self-image.
- These strategies are used to make the hearer feel good about themselves, their interests or possessions, and are most usually used in situations where the audience knows each other fairly well, or an individual's positive face needs, or self-worth, have to be met.
- **Example:**
  - (i) Yum! You make such a great sandwich! Would you make some?
  - (ii) This piece of work is really fantastic. Well done!



# Negative Politeness

- It is oriented towards the hearer's negative face and emphasize avoidance of imposition on the hearer.
- By attempting to avoid imposition from the speaker, the risk of face-threat to the hearer is reduced.
- It presumes that the speaker will be imposing on the hearer and there is a higher potential for awkwardness or embarrassment than in bald on record strategies and positive politeness strategies.
- **Example:**
  - (i) Could you please serve the meal?
  - (ii) I hope offense will not be taken

# Off-record/Indirect

- It uses indirect language and removes the speaker from the potential to be imposing. The strategy of doing off-record is to express something general or different than the speaker's true meaning and relies on the hearer's interpretation to have the speaker's purpose get conveyed.
- The speaker can get credit for not imposing on the hearer or give the hearer a chance to be helpful and generous.
- This strategy relies heavily on pragmatics to convey the intended meaning while still utilizing the semantic meaning as a way to avoid losing face.
- **Example (S -Speaker, H: Hearer):**
  - (i) If S says "I have stomach ache.", the H may infer that S is asking for some medicines, however, if questioned, S may deny this.
  - (ii) S: Are you going out? -H: Yes, but I'll come home early.

# Withhold the FTA

- The speaker makes a conscious decision to abstain from performing the FTA
- Many people seem to prefer to have their needs recognized by others without having to express those needs in language.
- **Example:**

You arrive at an important lecture and and pull out your notebooks to take notes, but you discover you don't have anything to write with. You think that the persona sitting next to you may provide the solution.

You: looking inside the bag

Other person: (Offering a pen) here you can use it.

# Criticism of Brown and Levinson's Politeness Theory

- **Universality claim:** Their theory being criticized for not really being as universally applicable as claimed because of the limited validity across different cultures.
- The way politeness is managed in Japan or Thailand does not match that US and UK

# Politeness Variation in Linguistic Studies

**Politeness is Complex:** Perceptions of (im)politeness vary across individuals.

**Gradations of Politeness:** Behavior spans multiple levels of more or less (im)polite actions.

**Context Matters:** Politeness must be analyzed within a social context.

## Influencing Factors

Gender



Age



Culture



# Computational Politeness in NLP

- **Politeness as Natural Language Understanding (NLU) Task**

- Identifies linguistic features of politeness and predicts whether text is polite or impolite.

- **Example:**

*Would you please help me? -> Polite*

*Would you please stop? .-> Impolite*

- **Politeness as Natural Language Generation (NLG) Task**

- Given the non-polite text, the objective of the politeness generation task is to generate a politeness-oriented text.

- **Example:**

*How can we help? -> Help has arrived! We are sorry to see that you are having trouble. How can we help?*

# Politeness as NLU Task (Politeness Identification)

	Datasets		Languages		Approach			Annotation		Features				Context	
	Non-conversational	Conversational	English	Other	Supervised	(Semi/Weakly/Un)-Supervised	Reinforcement Learning	Manual	Automatic	N-grams	Lexical	Syntactic	Other	Conversation	Other
Alexandrov et al. [2]		✓		✓	✓						✓	✓			
Danescu-Niculescu-Mizil et al. [33]	✓		✓		✓			✓		✓	✓	✓			
Li et al. [88]	✓		✓	✓	✓			✓			✓	✓			
Aubakirova and Bansal [4]	✓		✓		✓										
Chhaya et al. [24]	✓		✓		✓			✓			✓	✓			
Kumar [76]	✓			✓	✓			✓		✓			✓		
Mishra et al. [108]		✓	✓		✓			✓	✓					✓	
Dasgupta et al. [34]	✓		✓		✓										
Priya et al. [129]		✓	✓		✓			✓						✓	

# Politeness as NLG Task (Politeness Generation)

	Datasets		Languages		Approach			Annotation		Features				Context	
	Non-conversational	Conversational	English	Other	Supervised	(Semi/Weakly/Un)- Supervised	Reinforcement Learning	Manual	Automatic	N-grams	Lexical	Syntactic	Other	Conversation	Other
Madaan et al. [98]	✓		✓		✓				✓	✓					
Fu et al. [45]	✓		✓			✓									
Niu and Bansal [122]	✓	✓	✓			✓	✓								
Sennrich et al. [144]	✓		✓	✓	✓				✓						
Feely et al. [38]	✓		✓	✓	✓				✓						
Golchha et al. [49]		✓	✓					✓						✓	✓
Firdaus et al. [40]		✓	✓	✓				✓						✓	✓
Wang et al. [167]	✓		✓		✓				✓						
Firdaus et al. [43]		✓	✓		✓									✓	
Mishra et al. [107]		✓	✓						✓	✓				✓	
Silva et al. [148]		✓	✓		✓										
Saha et al. [140]	✓		✓			✓									
Firdaus et al. [41]		✓	✓											✓	
Mishra et al. [113]		✓	✓						✓					✓	
Mukherjee et al. [117]		✓	✓		✓									✓	
Mishra et al. [112]		✓	✓						✓					✓	
Priya et al. [130]		✓	✓						✓					✓	
Mishra et al. [109]		✓	✓						✓	✓				✓	
Mishra et al. [111]		✓	✓						✓					✓	
Firdaus et al. [42]		✓	✓		✓					✓				✓	
Mishra et al. [110]		✓	✓						✓					✓	



# Computational Politeness Studies Beyond English

## Politeness Identification

- Politeness detection in Hindi (Kumar et al.)
- Politeness detection in Chinese (Li et al.).
- Multilingual politeness detection in 9 languages (e.g., Hindi, Tamil, Korean, etc.)

## Politeness Generation

- Neural machine translation (NMT) to regulate politeness or formality
  - Honorific control in English → German (Sennrich et al.).
  - Formality control in English → Japanese (Feely et al.).
- Domain adaptation for English → French/Spanish/Czech translations (Viswanathan et al.)
- Courteous response generation for English and Hindi (Firdaus et al.)

# Politeness-labeled Datasets

## Non-conversational Datasets

- **Sources:** Social media, websites, blogs, and emails provide diverse textual data (e.g., user queries, complaints).
- **Nature of Text:** Often consists of requests (e.g., for help, information, or actions).

## Conversational Datasets

- **Sources:** Interaction on social media platforms like Twitter handles, Reddit, etc.
- **Nature of Text:** Multi-turn exchanges

# Politeness-labeled Non-conversational Datasets

Reference	Dataset	Domain	Language	Task
Danescu-Niculescu-Mizil et al. [33]	Stanford Politeness Corpus of requests	Wikipedia and Stack Exchange Request Posts	English	Politeness classification
Li et al. [88]	Politeness Corpus	Twitter and Weibo Posts	English and Chinese	Politeness classification
Madaan et al. [98]	Enron Corpus	Email Exchanges	English	Style transfer (Non-polite to polite)
Chhaya et al. [24]	Enron Dataset	Email	English	Frustration, formality and politeness classification
Sennrich et al. [144]	OpenSubtitles	Movies Subtitles	English, German	Politeness preservation in neural machine translation
Bharti et al. [9]	PolitePEER	Peer Reviews	English	Politeness classification
Kumar [74]	CO3H	Online Forums Requests	Hindi	Politeness classification
Kumar [75]	Politeness Corpus of Hindi Blogs	Blogs Posts and Comments	Hindi	Politeness classification

# Politeness-labeled Conversational Datasets

Reference	Dataset	Domain	Language	Task
Golchha et al. [49]	CYCCD	Customer Care	English	Polite response generation
Firdaus et al. [40]	Hindi Customer Care Dialogue	Customer Care	English, Hindi	Polite response generation
Wang et al. [167]	CSR-Driver Exchanges	Customer Care	English	Polite response generation
Bothe and Wermter [12], Priya et al. [129]	Politeness-annotated DailyDialog	Daily-life Conversations	English	Politeness classification
Mishra et al. [107]	Politeness-annotated MultiDoGO	Airline, Fastfood, Finance, Insurance, Media, Software	English	Polite dialogue generation
Mishra et al. [108]	Politeness-annotated DSTC-1, MDC	DSTC-1: Bus Schedule; MDC: Movie-Ticket Booking, Restaurant reservation, Taxi Ordering	English	Politeness classification
Mishra et al. [109]	Politeness-annotated TaskMaster	Flight, Food Ordering, Hotel Movie, Music, Restaurant, Sports	English	Politeness-adaptive dialogue generation

# Politeness-labeled Conversational Datasets (Cont.)

Reference	Dataset	Domain	Language	Task
Firdaus et al. [41]	Sentiment-annotated CYCCD	Customer Care	English	Sentiment-guided polite response generation
Mishra et al. [113]	Politeness-annotated PersuasionFor-Good	Charity Donation	English	Polite and empathetic persuasive dialogue generation
Priya et al. [129]	POEM	Mental Health and Legal Counseling	English	Politeness and Emotion Classification
Mishra et al. [111]	MHLCD	Mental Health and Legal Counseling	English	Polite and empathetic dialogue generation
Mishra et al. [112]	EPE-enEIH and EPE-HLCC	Mental Health and Legal Counseling	English	Emotion-adaptive polite and empathetic dialogue generation
Priya et al. [130]	HEAL	Mental Health and Legal Counseling	English	Politeness and empathy strategies- adaptive persuasive dialogue generation
Firdaus et al. [42]	Polite Behavior-annotated CYCCD	Customer Care	English	Polite behavior-aware response generation
Mishra et al. [110]	PSYCON	Mental Health Counseling	English	Polite and interpersonal behavior-aware response generation

# Approaches to Computational Politeness

(1) **Statistical Approaches**

(2) **Deep learning-based Approaches**

# Statistical Approaches to Politeness Identification

Reference	Task
Danescu-Niculescu-Mizil et al. [33]	Classifier for predicting politeness in natural language requests in English
Li et al. [88]	Develop a politeness feature set, PoliteLex, that is compatible with both English and Mandarin to study the resemblances and distinctions between the US and China's politeness and build SVM-based classifiers
Kumar [76]	SVM-based classifier for identifying linguistic politeness in Hindi

# Deep Learning-based Approaches to Politeness Identification

Reference	Task
Aubakirova and Bansal [4]	Convolutional Neural Network (CNN)-based architecture using the Stanford Politeness Corpus for predicting politeness in natural language requests
Dasgupta et al. [34]	Graph-induced transformer network (GiTN) to automatically detect formality and politeness in text.



# Statistical Approaches to Politeness Generation

Reference	Task
Madaan et al. [98]	Tag and generate framework, which is a two-staged approach for converting non-polite sentences into polite ones while preserving the content of the source text.
Fu et al. [45]	Generate a paraphrase of a given message that can convey the desired degree of politeness in a particular communication setting using a pipeline approach consisting of three steps: Plan, Delete, and Generate.

# Deep Learning-based Approaches to Politeness Generation

Reference	Task
Golchha et al. [49]	A reinforced pointer-generator model for generating emotionally and contextually consistent courteous responses in task-oriented customer-care dialogues in English.
Niu and Bansal [122]	Generate natural, varied, and contextually consistent polite or impolite responses for open-domain dialogues without parallel data
Mishra et al. [107]	A reinforcement learning (RL)-based Politeness Adaptive Dialogue System (PADS) uses a DistilBERT-based politeness classifier to extract politeness semantics and incorporates polite rewards, factoring in user satisfaction, to optimize task completion.
Saha et al. [140]	An ensemble of generative discriminators guiding DialoGPT to generate polite, detoxified, and emotionally charged counterspeech to combat online hate speech.

# Statistical Approaches to Politeness Identification

Reference	Task
Danescu-Niculescu-Mizil et al. [33]	Classifier for predicting politeness in natural language requests in English
Li et al. [88]	Develop a politeness feature set, PoliteLex, that is compatible with both English and Mandarin to study the resemblances and distinctions between the US and China's politeness and build SVM-based classifiers
Kumar [76]	SVM-based classifier for identifying linguistic politeness in Hindi

# Deep Learning-based Approaches to Politeness Identification

Reference	Task
Aubakirova and Bansal [4]	Convolutional Neural Network (CNN)-based architecture using the Stanford Politeness Corpus for predicting politeness in natural language requests
Dasgupta et al. [34]	Graph-induced transformer network (GiTN) to automatically detect formality and politeness in text.

# Statistical Approaches to Politeness Generation

Reference	Task
Madaan et al. [98]	Tag and generate framework, which is a two-staged approach for converting non-polite sentences into polite ones while preserving the content of the source text.
Fu et al. [45]	Generate a paraphrase of a given message that can convey the desired degree of politeness in a particular communication setting using a pipeline approach consisting of three steps: Plan, Delete, and Generate.

# Deep Learning-based Approaches to Politeness Generation

Reference	Task
Golchha et al. [49]	A reinforced pointer-generator model for generating emotionally and contextually consistent courteous responses in task-oriented customer-care dialogues in English.
Niu and Bansal [122]	Generate natural, varied, and contextually consistent polite or impolite responses for open-domain dialogues without parallel data
Mishra et al. [107]	A reinforcement learning (RL)-based Politeness Adaptive Dialogue System (PADS) uses a DistilBERT-based politeness classifier to extract politeness semantics and incorporates polite rewards, factoring in user satisfaction, to optimize task completion.
Saha et al. [140]	An ensemble of generative discriminators guiding DialoGPT to generate polite, detoxified, and emotionally charged counterspeech to combat online hate speech.

# A Computational Approach to Politeness: Overview

- Detect and measure politeness (& rudeness)
- Relation between politeness and social factors

# Data Collection: Stanford Politeness Corpus



35k requests on Wikipedia talk pages



373k requests on question-answer forums



# Examples

## Requests on Wikipedia

**Edit Request from Wikipedia:** *“Most of the information in here seems fine, except for 1 thing: the June record low of -11F at Anaktuvuk Pass in Alaska is dubious because it seems like that year that the Anaktuvuk Pass weather station was having data errors. Source: <https://xmacis.rcc-acis.org/>”*

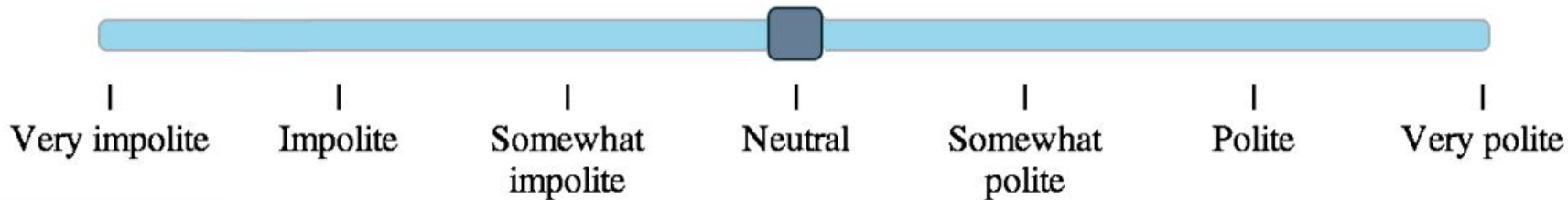
## Requests on Stack Exchange

**Question from Stack Overflow:** “How does Python's slice notation work: when I write code like `a[x:y:z]`, `a[:]`, `a[::2]` etc., how can I understand which elements end up in the slice? Please include references where appropriate.”

**Answer from Stack Overflow:** “You can't parse [X]HTML with regex. Because HTML can't be parsed by regex. Regex is not a tool that can be used to correctly parse HTML. As I have answered in HTML-and-regex questions here so many times before, the use of regex will not allow you to consume HTML.”

# How to decide - Polite/Impolite

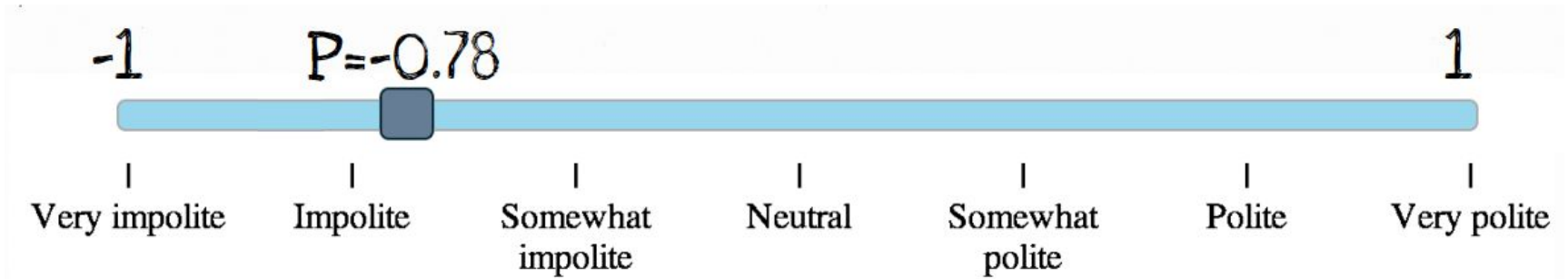
“Any code? or do we have to guess at how you did it?”



Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers). 250–259.

# How to decide - Polite/Impolite

“Any code? or do we have to guess at how you did it?”



# Data Collection: Stanford Politeness Corpus



35k requests on Wikipedia talk pages

4.5k annotated, 219 annotators, 0.7 MPC



373k requests on question-answer forums

6.5k annotated, 212 annotators, 0.6 MPC

# How reliable is this scale?

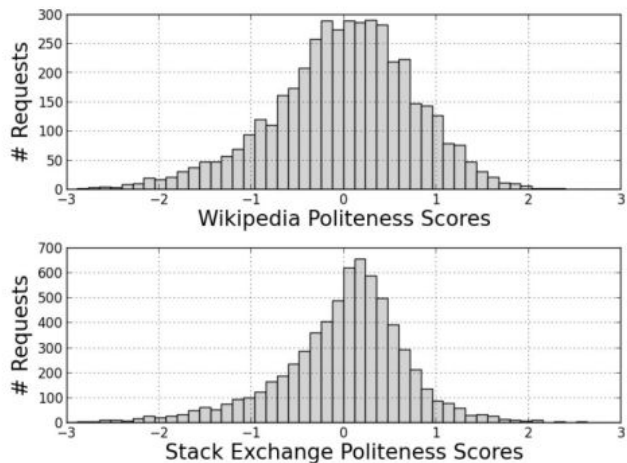


Figure 1: Distribution of politeness scores. Positive scores indicate requests perceived as polite.

Quartile:	1 <sup>st</sup>	2 <sup>nd</sup>	3 <sup>rd</sup>	4 <sup>th</sup>
Wiki	62%	8%	3%	51%
SE	37%	4%	6%	46%

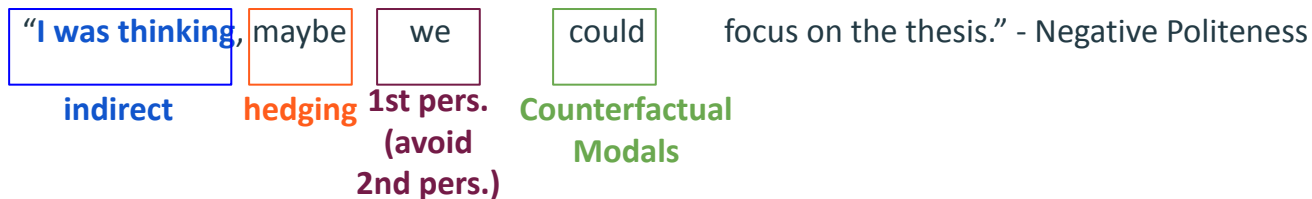
Table 2: The percentage of requests for which all five annotators agree on binary politeness. The 4<sup>th</sup> quartile contains the requests with the top 25% politeness scores in the data. (For reference, randomized scoring yields agreement percentages of <20% for all quartiles.)

# Politeness Strategies

Politeness addresses two basic desires (Brown and Levinson, 1978)

- (1) Desire not to be told what to do

*“You should focus on your thesis.”*



- (2) Desire to be paid respect - Positive Politeness



# Politeness Strategies

Positive (1-5) and negative (6–20) politeness strategies and their relation to human perception of politeness.

Strategy	Politeness	In top quartile	Example
1. Gratitude	0.87 <sup>***</sup>	78% <sup>***</sup>	<b>I</b> really <b>appreciate</b> that you've done them.
2. Deference	0.78 <sup>***</sup>	70% <sup>***</sup>	<b>Nice work</b> so far on your rewrite.
3. Greeting	0.43 <sup>***</sup>	45% <sup>***</sup>	<b>Hey</b> , I just tried to ...
4. Positive lexicon	0.12 <sup>***</sup>	32% <sup>***</sup>	<b>Wow!</b> / This is a <b>great</b> way to deal. ...
5. Negative lexicon	-0.13 <sup>***</sup>	22% <sup>**</sup>	If you're going to <b>accuse</b> me ...
6. Apologizing	0.36 <sup>***</sup>	53% <sup>***</sup>	<b>Sorry</b> to bother you ...
7. Please	0.49 <sup>***</sup>	57% <sup>***</sup>	Could you <b>please</b> say more...
8. Please start	-0.30 <sup>*</sup>	22%	<b>Please</b> do not remove warnings ...
9. Indirect (btw)	0.63 <sup>***</sup>	58% <sup>**</sup>	<b>By the way</b> , where did you find ...
10. Direct question	-0.27 <sup>***</sup>	15% <sup>***</sup>	<b>What</b> is your native language?
11. Direct start	-0.43 <sup>***</sup>	9% <sup>***</sup>	<b>So</b> can you retrieve it or not?
12. Counterfactual modal	0.47 <sup>***</sup>	52% <sup>***</sup>	<b>Could/Would</b> you ...
13. Indicative modal	0.09	27%	<b>Can/Will</b> you ...
14. 1st person start	0.12 <sup>***</sup>	29% <sup>**</sup>	<b>I</b> have just put the article ...
15. 1st person pl.	0.08 <sup>*</sup>	27%	Could <b>we</b> find a less complex name ...
16. 1st person	0.08 <sup>***</sup>	28% <sup>***</sup>	It is <b>my</b> view that ...
17. 2nd person	0.05 <sup>***</sup>	30% <sup>***</sup>	But what's the good source <b>you</b> have in mind?
18. 2nd person start	-0.30 <sup>***</sup>	17% <sup>**</sup>	<b>You'</b> ve reverted yourself ...
19. Hedges	0.14 <sup>***</sup>	28%	<b>I suggest</b> we start with ...
20. Factuality	-0.38 <sup>***</sup>	13% <sup>***</sup>	<b>In fact</b> you did link, ...

# Politeness Classifier

- Bag-of-Words Classifier (BOW):
  - Model: Support Vector Machine (SVM).
  - Features: Unigram feature representation.
- Linguistically Informed Classifier (Ling.):
  - Model: SVM Features:
  - Unigram features.
  - Additional linguistic features (as listed in previous slide)
- Human Reference Point:
  - Data: Collected three new politeness annotations for each request in the dataset.
  - Performance Metric: Calculated the percentage of requests where the average score from additional annotations matches the original binary politeness class.
  - Example: A positive score corresponds to the "polite" class.



# Politeness Classifier Results

	In-domain		Cross-domain	
Train	Wiki	SE	Wiki	SE
Test	Wiki	SE	SE	Wiki
<b>BOW</b>	79.84%	74.47%	64.23%	72.17%
<b>Ling.</b>	83.79%	78.19%	67.53%	75.43%
<b>Human</b>	86.72%	80.89%	80.89%	86.72%

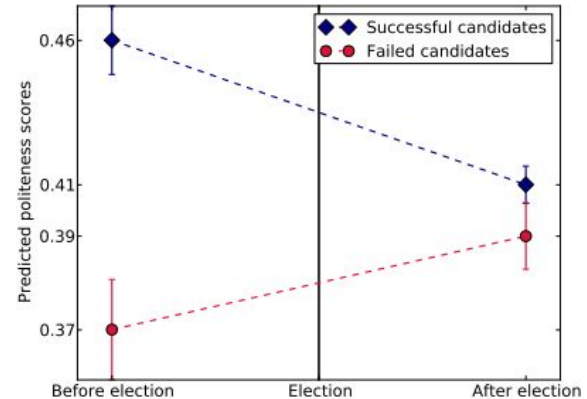
## A polite/impolite classifier. Why?

- observations are robust and generalizable
- also in a cross-domain setting
- extend politeness labels to very large data

# Politeness and Social Outcome

Eventual status	Politeness	Top quart.
Admins	0.46**	30%***
Non-admins	0.39***	25%
Failed	0.37**	22%

Editors who will eventually become admins are more polite than non-admins than editors who will eventually fail to become admins. Out of their requests, 30% are rated in the top politeness quartile. This analysis was conducted on 31k requests (1.4k for Admins, 28.9k for Non-admins, 652 for Failed).



Successful and failed candidates before and after elections. Editors that will eventually succeed (diamond marker) are significantly more polite than those that will fail (circle markers). Following the elections, successful editors become less polite while unsuccessful editors become more polite.

# CNN-based Politeness Classifier

One convolutional layer followed by a pooling layer

The result is then passed to a fully-connected softmax layer that outputs probabilities over labels.

# Results

<b>Model</b>	<b>Wiki</b>	<b>SE</b>
Bag-of-Words	80.9%	64.6%
Linguistic Features	82.6%	65.2%
With Discovered Features	83.8%	65.7%
CNN	85.8%	66.4%

**Table 1:** Accuracy Results on Wikipedia and Stack Exchange.

# Observations: Rediscovering Existing Strategies

**Deference (+)** A way of sharing the burden of a request placed on the addressee. Activation cluster examples: {"nice work so far on your rewrite..."; "hey, good work on the new pages..."}

**Direct Question (-)** Questions imposed on the converser in a direct manner with a demand of a factual answer. Activation cluster examples: {"what's with the radio , and fist in the air?"; "what level warning is appropriate?"}

# Observations: Extending Existing Strategies

**Gratitude (+)** Our CNN learns a special shade of gratitude, namely it distinguishes a cluster consisting of the bigram thanks for. Activation cluster examples: {"thanks for the good advice."; "thanks for letting me know."}

**Counterfactual Modal (+)** Sentences with Would you/Could you get grouped together as expected; but in addition, the cluster contains requests with Do you mind as well as gapped 3-grams like Can you ... please?, which presumably implies that the combination of a later please with future-oriented variants can/will in the request gives a similar effect as the conditional-oriented variants would/could. Activation cluster examples: {"can this be reported ... grid, please?"; "do you mind having another look?"}

# Observations: Discovering Novel Strategies

**Indefinite Pronouns (-)** Danescu-NiculescuMizil et al. (2013) distinguishes requests with first and second person (plural, starting position, etc.). However, we find activations that also react to indefinite pronouns such as something/somebody. Activation cluster examples: {"am i missing something here?"; "wait for anyone to discuss it."}

**Punctuation (-)** Though non-characteristic in direct speech, punctuation appears to be an important special marker in online communities, which in some sense captures verbal emotion in text. E.g., one of our neuron clusters gets activated on question marks "???" and one on ellipsis "...". Activation cluster examples: {"now???" ; "original article????"; "hello?????"}

# Problem Statement: Politeness Transfer

Converting non-polite sentences to polite sentences while preserving the meaning





# Applications

Hi Rebekah

I wanted to take a minute to personally thank you for supporting Pasta Marco over the last 15 years.

The pasta joint wouldn't be possible without amazing customers like you. As a special thank you, we're giving away free meatballs all month long. Because you're on our email list, you're the first to know!

Also, I'd like to invite you to a special Customer Appreciation Day on Saturday, September 14. Join us for pasta tastings and a special spaghetti making class hosted by Chef Marco himself.

Spots are limited, so register here:

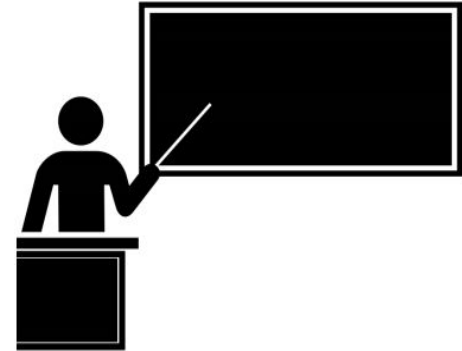
Thanks for all your years of support!

Sincerely,  
Bek at Pasta Marco

Automatic Email Response



Personal Chatbots

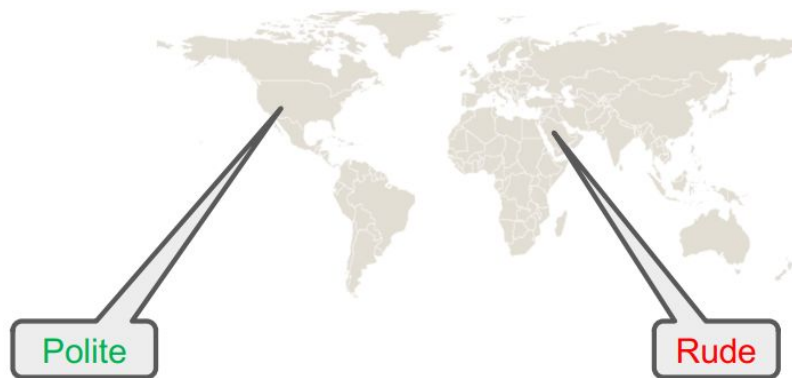


Education

# Challenges

## (1) Politeness is culturally diverse

Could you please  
open the door



# Challenges

(2) Politeness is subtle. It isn't just "please" and "thank you" [Danescu-Niculescu-Mizil et al., 2013]

Greeting

Hey, let's try to finish this in an hour.

Indirect

By the way, do you know when is the deadline?

Positive Lexicon

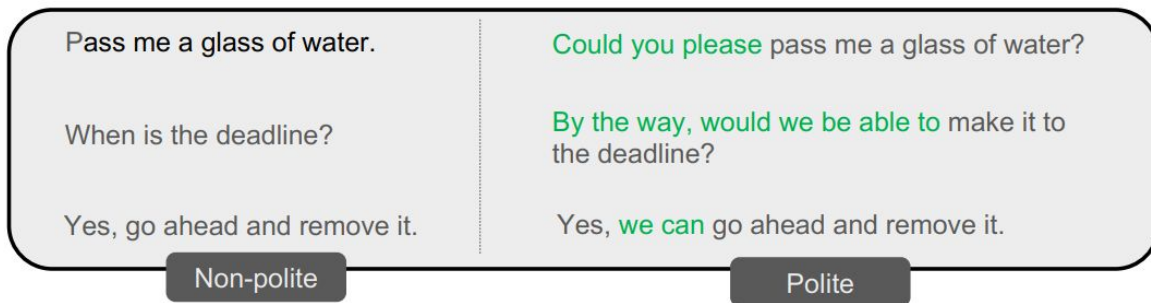
Wow! This diagram looks amazing.

1<sup>st</sup> Person Plural

Let us find a good name for this paper.

# Challenges

## (3) Ill-defined class: Non-polite



Easy to detect **presence** of politeness  
**Difficult** to objectify **absence** of politeness

## (4) Data paucity (No labeled Data)

# Focus of the Work

Transfer only from non-polite to polite

Not all sentences can be made more polite

The earth revolves around the sun.

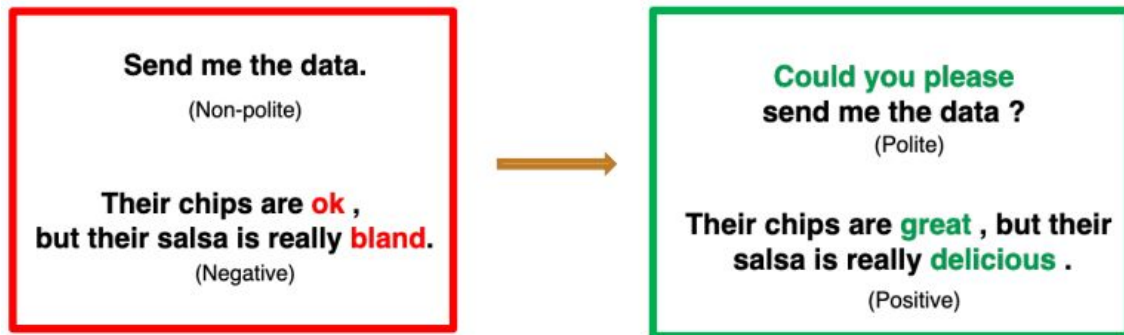


**This work focus on converting request or action-directives [Jurafsky, 1997] to polite requests (while preserving the meaning)**

# Methodology: Text Attribute Transfer by Learning to Tag and Generate

Transfer Desiderata:

1. Successful transfer into target style



# Methodology: Text Attribute Transfer by Learning to Tag and Generate

Transfer Desiderata:

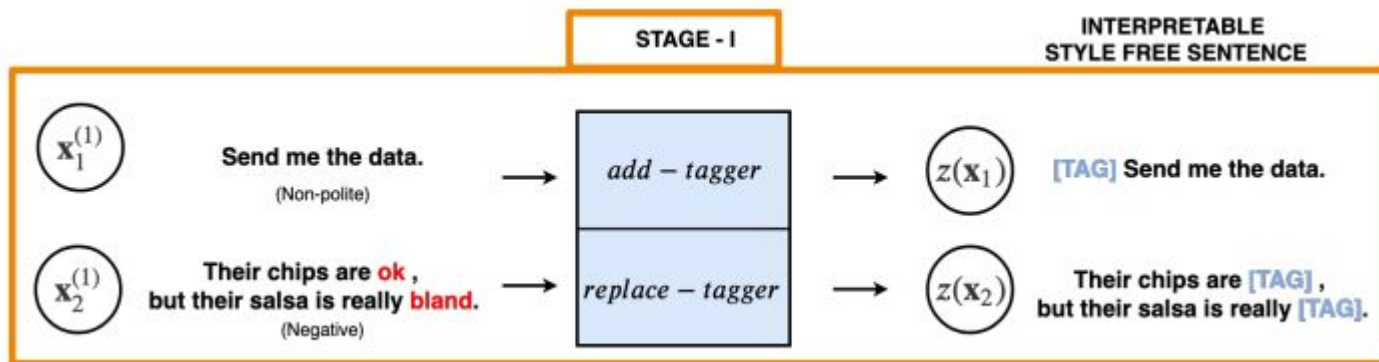
1. Successful transfer into target style.
2. Retaining content words (non-attribute markers).

[TAG] **Send me the data.**

**Their chips are** [TAG] ,  
**but their salsa is really** [TAG].

# Methodology: Tag and Generate Pipeline

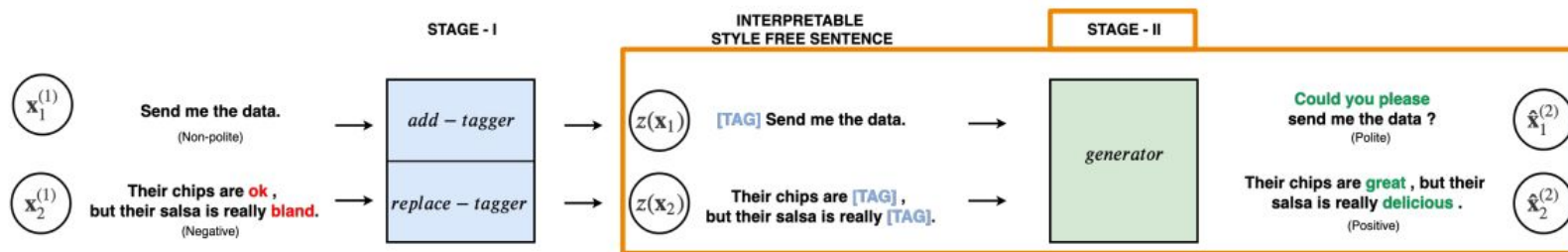
Use tagger to TAG words indicative of the source style and get a style neutral representation.





# Methodology: Tag and Generate Pipeline

GENERATE context appropriate phrases in the target style.



# Creating Artificial Data for Training Tagger

Step-I : Remove attribute markers

1. The chips are **ok** but their salsa is really **bland**
2. The service the last time I went was just **terrible**.



1. The chips are [redacted] but their salsa is really [redacted].
2. The service the last time I went was just [redacted].

# Creating Artificial Data for Training Tagger

Step-II : Generate Tags

1. The chips are [REDACTED] but  
their salsa is really [REDACTED].



2. The service the last time I  
went was just [REDACTED].

1. The chips are [TAG2] but  
their salsa is really [TAG3].

2. The service the last time I  
went was just [TAG3].

# Creating Artificial Data for Training Tagger

Step III. Use attribute markers of the style target to generate artificial parallel data.

- |   |  |  |
|---|--|--|
| 1. The chips are [TAG2] but their salsa is really [TAG3]. |  | 1. The chips are <b>great</b> but their salsa is really <b>delicious</b> . |
| 2. The service the last time I went was just [TAG3].      |  | 2. The service the last time I went was just <b>awesome</b> .              |

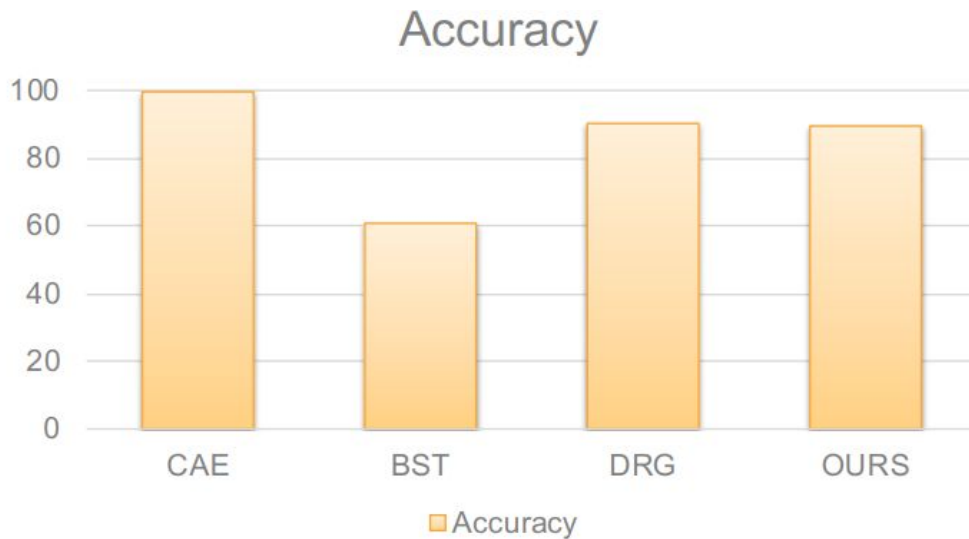
# Results

**What % of the outputs are polite?**

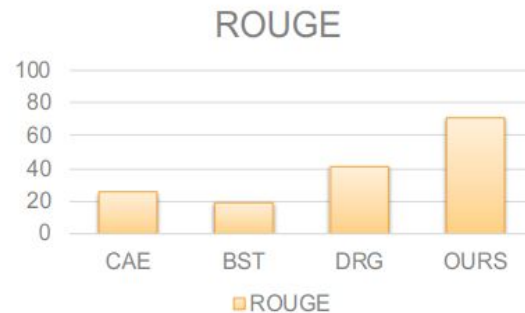
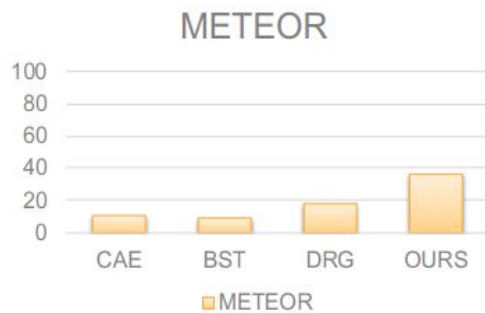
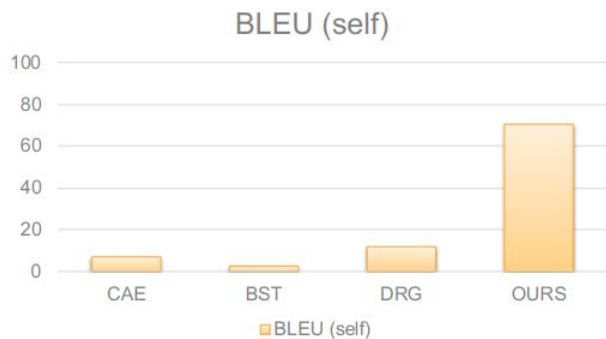
CAE: Shen et. al., 2017

BST: Prabhumoye et. al., 2017

DRG: Lo et. al., 2018

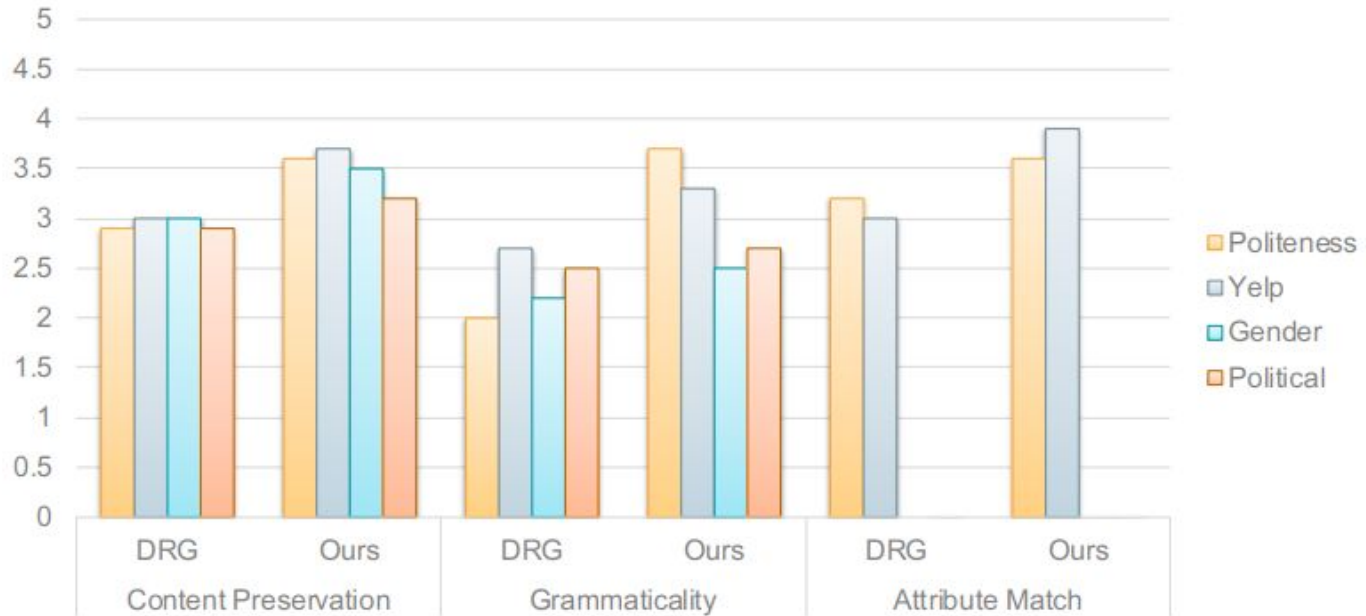


# Results



**How much of the input content is preserved?**

# Human Evaluation



# Qualitative Analysis

Input	Output	Politeness Strategy
what happened to my personal station?	could you please let me know what happened to my personal station?	Counterfactual modal
yes, go ahead and remove it.	yes, we can go ahead and remove it.	First person plural
not yet-i'll try this wkend.	sorry not yet - i'll try to make sure this wk	Apologizing
please check on metromedia energy,	please check on metromedia energy, thanks	Mitigating please start



# Problem Definition

To **transform** a generic chatbot response into a response which uses courteous phrases and emoticons to display appreciation, empathy, apology, assurance, in coherence with the state of conversation

**Domain:** Customer Care on Twitter

## Motivation

To transform a generic chatbot reply into one that:

- Is emotionally aware and intelligent
- Uses courteous phrases and emoticons to display appreciation, empathy, apology, assurance
- End motive is to increase user satisfaction and to build customer relations

# Example I (Expressing Apology / Empathy )

somebody from @VerizonSupport please help meeeee 🙄🙄🙄🙄 I'm having the worst luck with your customer service

@115719 How can we help?

@VerizonSupport I finally got someone that helped me, thanks!

@115719 Awesome!

somebody from @VerizonSupport please help meeeee 🙄🙄🙄🙄 I'm having the worst luck with your customer service

@115719 **Help has arrived! We are sorry to see that you are having trouble.** How can we help?

@VerizonSupport I finally got someone that helped me, thanks!

@115719 Awesome! **If you ever need us we are just a tweet away.**

# Resource Creation: Data Source and Description

- Source the requisite Twitter data from the dataset made available on Kaggle
- Segment the tweet into sentences
  - Purely courteous (and non-informative) sentences must be removed
  - Purely informative sentences must be retained
  - Informative sentences with courteous expressions must be transformed (to remove only the courteous part from the sentence)

# Resource Creation: Scaling up for large data creation

- **Clustering**

- The vector-semantic representations of sentences are obtained using the sentence encoder trained on the SNLI corpus.
- Use the K-Means clustering( $k = 300$ ) to cluster these sentences.

- **Annotations**

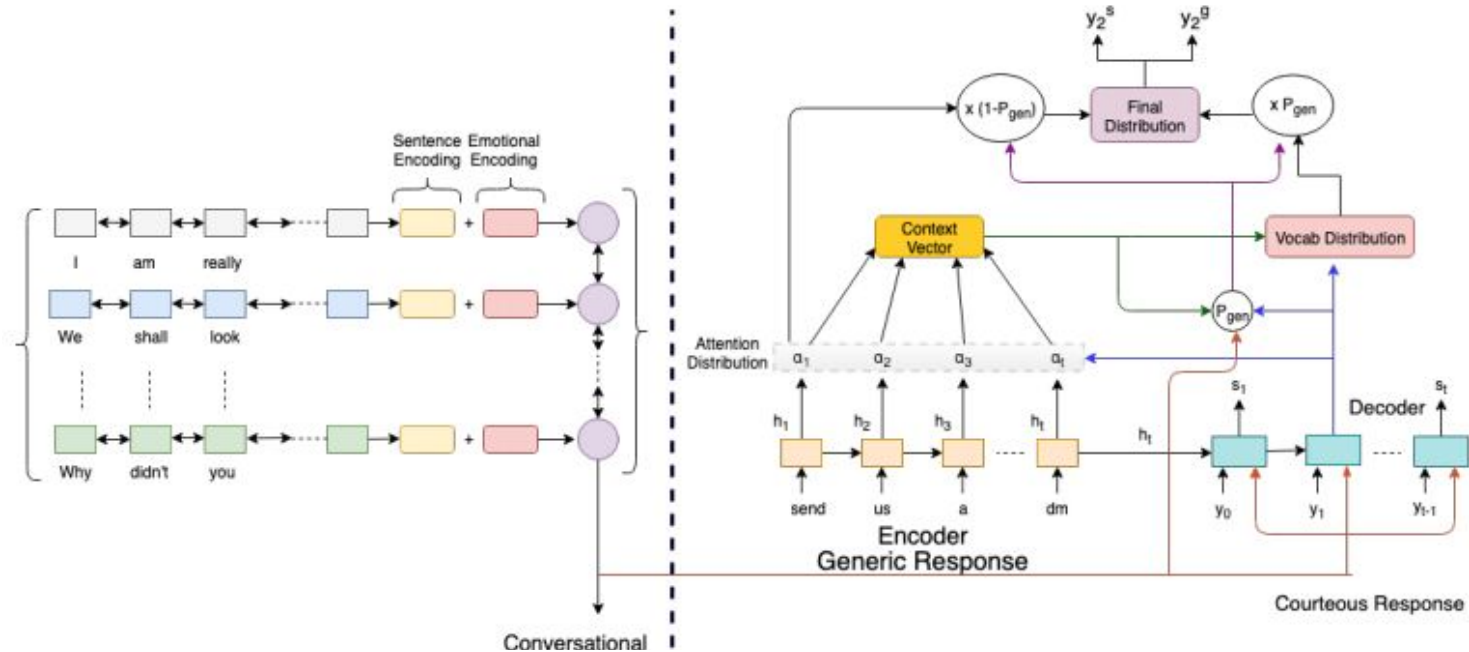
- purely courteous,
- purely informative,
- hybrid

- **Preparing generic responses**

- Obtain the generic response by removing the courteous sentences, retaining the informative sentences, and replacing the hybrid sentences with the prepared generic equivalents

# Proposed Methodology

- Based on a [reinforced pointer-generator model](#) for the sequence to sequence task
- The model is also conditioned on a hierarchically encoded and emotionally aware conversational context



# Results: Automatic Evaluation

Model		BLEU	ROUGE			PPL	CP	EA
			1	2	L			
<i>1</i>	<i>Seq2Seq</i>	56.80	63.8	59.06	64.52	58.21	68.34	82.43
<i>2</i>	<i>Seq2Seq + P</i>	66.11	69.92	64.85	66.40	<b>42.91</b>	<b>77.67</b>	81.98
<i>3</i>	<i>Seq2Seq + P + EE</i>	68.16	72.18	67.92	71.17	43.52	76.05	85.75
<i>4</i>	<i>Proposed Model</i>	<b>69.22</b>	<b>73.56</b>	<b>69.92</b>	<b>72.37</b>	43.77	77.56	<b>86.87</b>

P: Pointer Generator Model; EE: Emotional embedding  
Perplexity (PPL), Content preservation (CP), Emotional accuracy (EA)

# Results: Human Evaluation

Model	F			CA			CoA		
	0	1	2	0	1	2	-1	0	1
<i>Model 1</i>	15.70	42.50	41.80	16.21	41.69	42.10	23.71	51.08	25.21
<i>Model 2</i>	14.23	42.77	43.00	15.62	39.65	44.73	22.05	39.43	38.52
<i>Model 3</i>	11.15	44.10	44.75	13.66	41.12	45.22	15.23	41.22	43.55
<i>Our Model</i>	10.05	44.90	44.60	13.85	38.48	47.67	14.11	41.11	44.78

All values are in percentages.

F: Fluency, CA: Content Adequacy, CoA: Courtesy Appropriateness

Scale for F and CA - 0: incorrect or incomplete, 1: moderately correct, 2: correct

Scale for CoA: -1: inappropriate, 0: non-courteous, 1: appropriate

# Generated Examples

Generic Input	Model 1	Model 2	Model 3	Our Model
dm us more info and well take a look into it for you	we'll look into it	im sorry to hear this please dm us more info and we'll take a look into it for you	were here to help please dm us more info and well take a look into it for you	were here to help please dm us more info and well take a look into it for you at the earliest
adjust the brightness via your display settings on your device	whos the brightness via your display settings on your device	were here to help adjust the brightness via your display settings on your device	we have several ways to change the display brightness on your device and were happy to help	thanks for reaching out we have several ways to change the display brightness on your device and were happy to help
we'll follow up with the store	we'd like to help well follow up	were here to help well follow up with the store	sorry to hear that well follow up with the store	thats disappointing to hear, we'll follow up with the store
can you confirm which platform you are using for video access ? what is the error ?	what is the error ?	I am sorry for the frustration ! can you confirm which platform you are using for video access ? what is the error ?	I am sorry to hear this can you confirm which platform you are using for video access? what is the error?	I am sorry for any frustration, can you please confirm which platform you are using for video access? Please tell us what is the error.
fill this form <url>	please fill this form <url>	were here to help fill this form <url>and I'll contact you at the earliest a	apologies for the hassle, please fill this form <url>and we'll contact you thank you for reaching out to us we will follow up with the store	i am sorry for the hassle, please fill this form <url> and ill contact you at the earliest



# Politeness for Social Good Applications

## Politeness and Related Aspects (Emotion, Sentiment, Empathy, Persuasion)



**Counseling**



**Persuasion**



# Politeness in Counseling

# Politeness in Counseling Dialogue Systems

- Counseling dialogue system is the need of time
  - 20% of the global population suffers from mental health problems
  - Limited availability of experts
- Politeness and related aspects (sentiment, emotion, empathy, persuasion) are crucial for developing a cordial atmosphere and establishing an emotional bond and rapport during the counseling conversation
  - Fosters a strong therapeutic alliance and promotes a client's feelings of safety, trust, and understanding
    - Reflecting and validating emotions and experiences
    - Active Listening
    - Perspective-taking

Politeness helps create a therapeutic environment where clients feel heard, understood, and supported, leading to increased self-awareness, personal growth, and positive therapeutic outcomes.



Please don't be sad, I understand how you feel.

# Beneficiaries

- Victims/Survivors
- Substance Abuse and Addiction
- Individuals with Clinical Diagnoses (Depression, Anxiety etc.)
- Grief and Loss
- Stress Management
- LGBTQ+ Individuals
- Children and Adolescents
- Elderly People

# Beneficiaries

- Victims/Survivors
- Substance Abuse and Addiction
- Individuals with Clinical Diagnoses (Depression, Anxiety etc.)
- Grief and Loss
- Stress Management
- LGBTQ+ Individuals
- Children and Adolescents
- Elderly People

# Mental Health and Legal Counseling for Victims: Need & Access

**There is a rise in the number of crimes against women and children**

Violence adversely affects women's and children's mental health

**Prevention of crime against women and children is important to achieving SDGs 2030**

Goal 3: Good health and well-being

Goal 5: Gender Equality

Goal 16: Peace, Justice and Strong institutions

**Access to mental health and legal counseling support is limited**

- Victims fear stigmatization and societal condemnation
- Reluctant to seek support

# Mental Health Counseling for Substance Addicts and Clinically Diagnosed Individuals: Need & Access

**There is a rise in the number of substance addicts and mental health patient**

Substance addiction and mental health issues leads to physical health deterioration, cognitive impairments, emotional and psychological consequences, behavioral issues, suicidal tendencies and many more.

**Adequate and timely intervention is important to achieving SDGs 2030**

Goal 3: Good health and well-being

**Access to mental health support is limited**

Global Shortage of mental health professionals

# Counseling Dialogue System: Significance

- **Enhanced Accessibility**
  - Provides 24/7 support, making mental health services more accessible to individuals in remote areas or with mobility issues.
  - Reduces barriers like stigma and scheduling challenges for seeking professional help.
- **Scalability**
  - Addresses the shortage of trained mental health professionals by serving large populations simultaneously.
  - Offers an efficient way to manage high demand during crises or emergencies.
- **Personalized Support**
  - Tailors responses based on user profiles, emotional state, and conversational context.
  - Incorporates adaptive techniques to cater to diverse needs, such as substance addiction or clinical diagnoses.
- **Early Intervention**
  - Helps identify early signs of distress or mental health issues through natural language interactions.
  - Facilitates timely referrals to human professionals for critical cases.

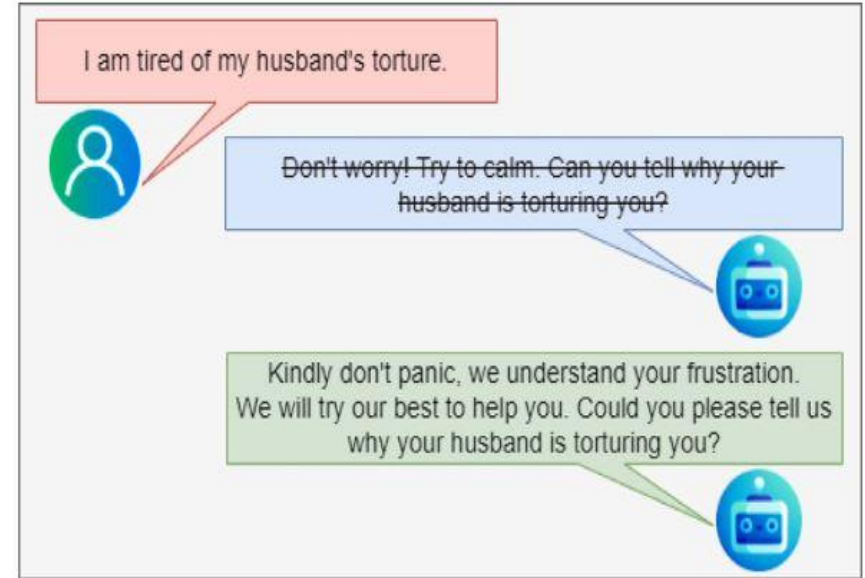
Developing Counseling Dialogue System: A step towards achieving the overarching objective of the SDGs 2030 agenda to  
*“leave no one behind”*



# Politeness and Emotion in Mental Health and Legal Counseling of Victims

- Politeness and emotion are crucial aspects of human communication
- Perceiving politeness in conversations provides cues about the interlocutors' social behaviors
- Perceiving emotions provides affective information about them

Incorporation of these aspects in the conversational agents would promote engagement



# Inter-connectedness: Politeness and Emotion

- Politeness helps to differentiate between emotions such as those linked with apology or anger, both of which are intrinsically negative
  - *“We are really sorry for the inconvenience. Kindly stay with us for a moment.”* - Politeness label: **polite** and Emotion: **apology**
  - *“Do hell with your sorry! Get lost.”* - Politeness label: **impolite** and Emotion: **anger**

Bothe et al. Conversational Analysis of Daily Dialog Data using Polite Emotional Dialogue Acts. In LREC 2022.

Feng et al. EmoWOZ: A Large-Scale Corpus and Labelling Scheme for Emotion Recognition in Task-Oriented Dialogue Systems. In LREC 2022

Priya et al. A Multi-task Learning Framework for Poliness and Emotion Detection in Dialogues for Mental Health and Legal Counseling. In Expert Systems With Applications 2023.

# POLiteness and EMotion Annotated Dataset (POEM)

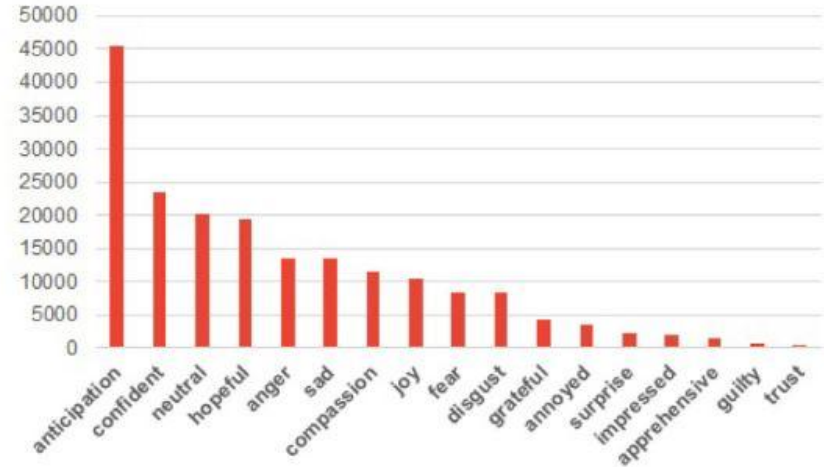
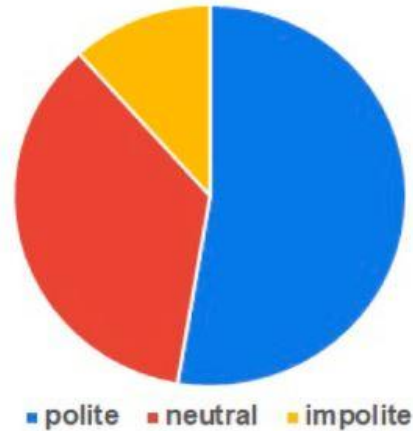
- 5000 Mental health and legal counseling conversations
  - for women and children victims of different crimes (stalking, harassment etc.)
- Between two-humans
  - Once acts as an Agent and other as a Victim
- Annotated with
  - One of the three politeness labels: *polite, neutral, impolite*
  - Multi-label emotion categories from 17 emotion classes: *anticipation, confident, hopeful, anger, sad, joy, compassion, fear, disgust, annoyed, grateful, impressed, apprehensive, surprised, guilty, trust and neutral*

# POEM Dataset Statistics

## Dataset Statistics

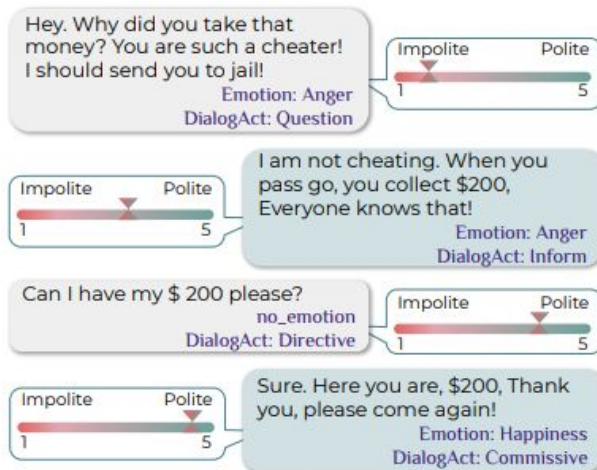
Metrics	Train	Validation	Test
# of dialogues	2859	1080	1061
# of utterances	77,806	25,775	25,744
Avg. utterances per dialogue	27.21	23.87	24.26

## Politeness and Emotion Distribution



# Polite Emotional DailyDialog Dataset

- DailyDialog covers topics from daily life (*ordinary life topics to financial topics*)
- Annotated with
  - 7 emotion categories: *anger, disgust, fear, happiness, sadness, surprise and neutral*
  - 4 Dialog Acts: *inform, question, directive, commissive*
  - Politeness: Score in the range 1-5
    - Score around 3 indicates neutral
    - Score inclined towards 1 indicates impolite
    - Score inclined towards 5 indicates polite

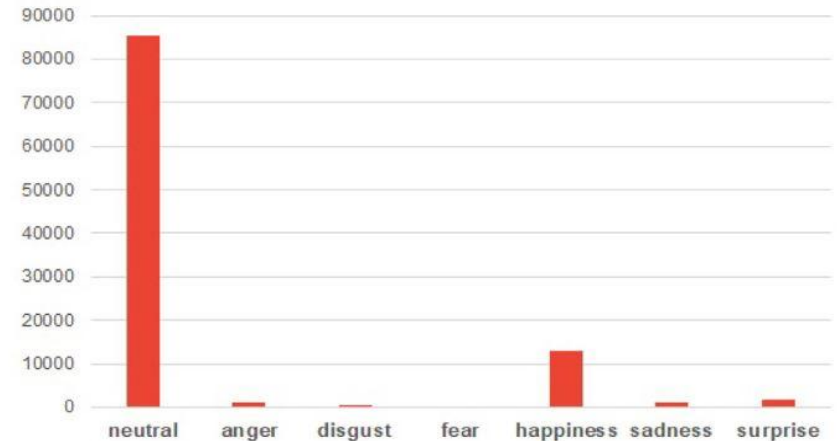
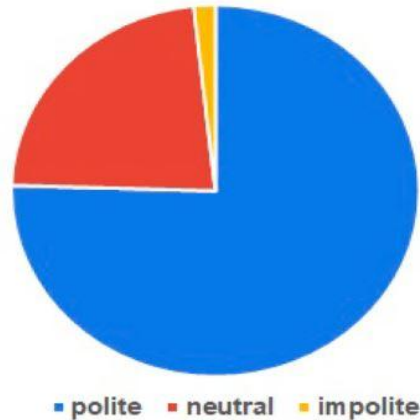


# Polite Emotional DailyDialog Dataset Statistics

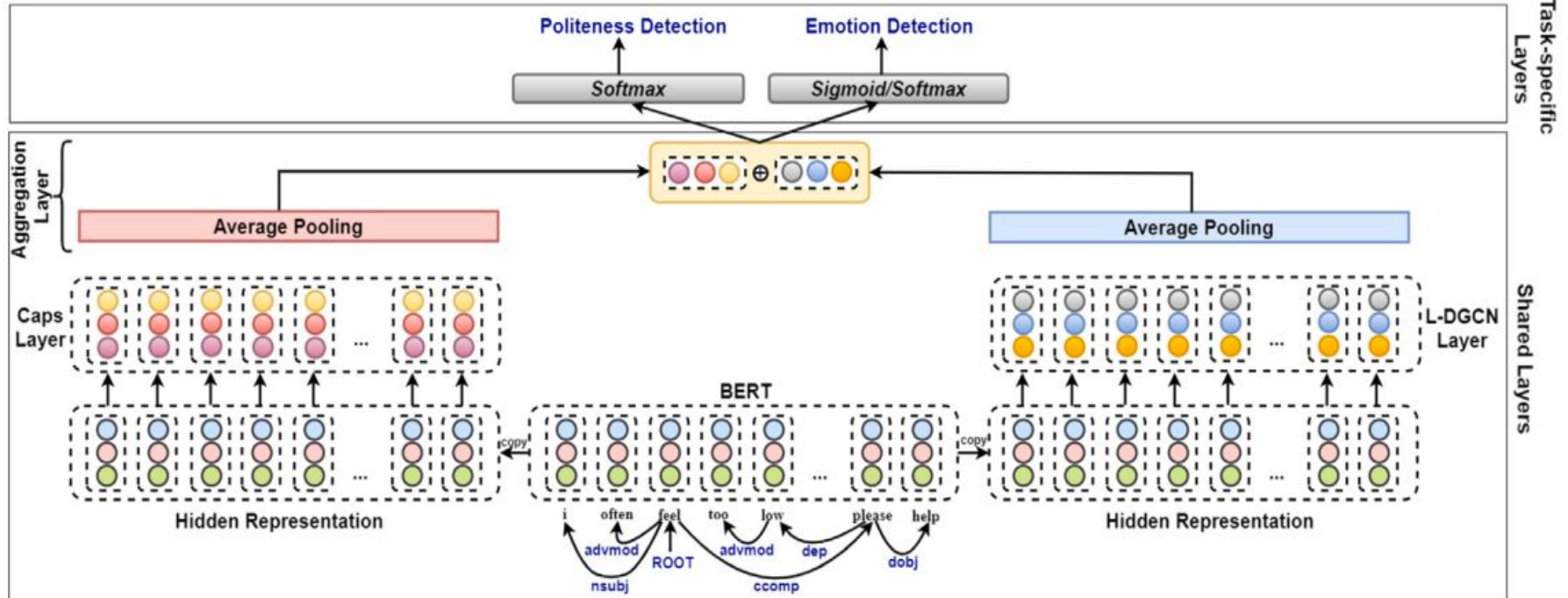
## Dataset Statistics

Metrics	Train	Validation	Test
# of dialogues	11,118	1000	1000
# of utterances	87,170	8069	7740
Avg. utterances per dialogue	7.84	8.06	7.74

## Politeness and Emotion Distribution



# Methodology



# Results on POEM

Learning paradigm	Models	Politeness				Emotion			
		Acc	P	R	F1	S-Acc	Micro-F1	JI	HL
Single-task learning	Baselines								
	<i>BERT</i>	83.01	79.47	78.74	79.10	48.52	52.43	0.42	0.079
	<i>Caps</i>	83.34	76.69	82.72	79.59	49.56	54.39	0.47	0.072
	<i>DGCN</i>	84.12	82.12	80.23	81.16	49.89	55.23	0.50	0.068
	<i>DialogueRNN (Majumder, Poria, Hazarika, et al., 2019)</i>	83.78	81.43	81.57	81.50	49.12	56.49	0.49	0.067
	<i>DialogueGCN (Ghosal et al., 2019)</i>	84.56	82.65	83.37	83.01	50.11	58.16	0.51	0.064
	<i>Caps-DGCN</i>	85.37	85.09	84.95	85.02	50.95	59.64	0.52	0.061
Multi-task learning	<i>BERT</i>	85.01	84.41	84.04	84.22	50.73	61.86	0.46	0.052
	<i>Caps</i>	86.22	85.15	85.14	85.14	51.91	67.44	0.51	0.047
	<i>DGCN</i>	87.67	86.12	86.31	86.21	54.94	70.32	0.55	0.041
	<i>DialogueRNN (Majumder, Poria, Hazarika, et al., 2019)</i>	87.12	85.63	86.85	86.24	55.00	71.82	0.56	0.043
	<i>DialogueGCN (Ghosal et al., 2019)</i>	87.43	85.63	86.75	86.19	55.60	72.34	0.57	0.042
	<i>DCR-Net (Qin et al., 2020)</i>	87.86	87.22	85.34	86.27	56.17	73.25	0.58	0.041
	<i>Co-GAT (Qin et al., 2021)</i>	88.92	86.53	86.40	86.46	57.13	74.38	0.60	0.040
	Proposed model								
	<i>Caps-DGCN</i>	<b>90.30</b>	87.45	86.93	<b>87.19</b>	<b>58.72</b>	76.79	<b>0.63</b>	0.038



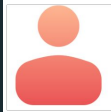
# Results on DailyDialog

Learning paradigm	Models	Politeness				Emotion			
		Acc	P	R	F1	Acc	P	R	F1
Single-task learning	Baselines								
	<i>BERT</i>	80.19	72.86	65.00	68.71	67.87	40.14	38.67	39.39
	<i>Caps</i>	81.13	74.67	65.84	69.98	68.03	40.43	39.14	39.78
	<i>DGCN</i>	81.37	75.10	68.43	71.61	71.90	42.45	40.98	41.70
	<i>DialogueRNN (Majumder, Poria, Hazarika, et al., 2019)</i>	80.64	68.59	71.89	70.20	69.40	40.36	41.26	40.81
	<i>DialogueGCN (Ghosal et al., 2019)</i>	81.33	70.72	72.35	71.53	71.44	41.39	43.03	42.19
	<i>Caps-DGCN</i>	82.04	75.32	67.75	71.33	73.52	44.23	42.80	43.50
Multi-task learning	<i>BERT</i>	81.23	74.23	68.54	71.27	72.76	42.11	40.27	41.17
	<i>Caps</i>	82.96	76.13	69.63	72.74	73.20	43.90	42.01	42.94
	<i>DGCN</i>	83.00	78.16	70.45	74.10	74.33	46.40	44.87	45.62
	<i>DialogueRNN (Majumder, Poria, Hazarika, et al., 2019)</i>	83.21	73.53	72.84	73.18	73.56	45.21	43.59	44.39
	<i>DialogueGCN (Ghosal et al., 2019)</i>	84.06	72.99	75.88	74.41	74.28	47.89	43.62	45.66
	<i>DCR-Net (Qin et al., 2020)</i>	84.79	73.53	74.74	74.13	74.91	46.67	45.13	45.89
	<i>Co-GAT (Qin et al., 2021)</i>	85.46	74.34	72.94	74.63	75.00	45.21	46.52	45.86
	Proposed model								
	<i>Caps-DGCN</i>	<b>86.78</b>	81.38	70.01	<b>75.27</b>	<b>75.49</b>	47.50	44.75	<b>46.08</b>

# Case Study

Input	Gold	Predicted		
		Caps-DGCN <sup>ST</sup>	Caps-DGCN <sup>MT</sup>	
rakshak my landlord is try to harass me please help	Politeness Emotion	polite sad, anger, hopeful	polite sad, annoyed hopeful	polite sad, anger, hopeful
they dont have anything personal stuff about me as far as i know the thing is that they are not able to handle my success so they are threatening me badly	Politeness Emotion	neutral confident, sad, fear	impolite confident, sad, fear	neutral confident, sad, fear
what can I write in the application dont call me dear	Politeness Emotion	impolite anticipation, annoyed	polite anticipation	impolite anticipation, annoyed

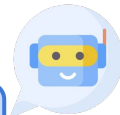
# Generic vs. Counseling Conversation



I am struggling with my life after that incident and am scared to share it with anyone.

**Victim**

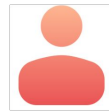
Don't worry, can you tell which incident made you feel so?



**Counselor**

**Generic Response**

During counseling, the *Counselor* should employ different counseling strategies as per going conversations and *Victim's* state.



I am struggling with my life after that incident and am scared to share it with anyone.

**Victim**

Don't worry, you are talking on a **safe and secure platform**. Can you tell which incident made you feel so?



**Counselor**

**Incorporation of appropriate counseling strategy (*confidentiality assurance*)**

# Politeness and Empathy in Counseling

## Politeness

- Shows the **caring behavior**
- Improves the **sense of empathy**
- Facilitates **revelation of personal information**
- (Lucas et al. 2014, Kim et al. 2018)

### *Use of polite and empathetic language*

- creates familiar, warm and comfortable atmosphere to share their feelings and problems
- helps to console users and gain their trust

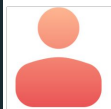
## Empathy

- **Core component** of counseling
- Promotes effective **therapeutic bonding** and **rapport building**
- Strong associations with **positive counseling outcomes** (Norcross 2002, Elliott et al. 2018)



Please don't be sad, I understand how you feel.

# Intro: Politeness and Empathy in Counseling: An Example



**Victim**

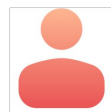
I am struggling with my life after that incident and am scared to share it with anyone.



**Counselor**

**Kindly** don't get scared, **we are here to help you** in a safe and secure environment. **Please feel free** to share with us which incident made you feel so.

Communication of *Politeness* during counseling



**Victim**

I am struggling with my life after that incident and am scared to share it with anyone.



**Counselor**

**We are really sorry to hear about this, <user>. We are here to help you** in a safe and secure environment. **We would love to know more about your problems** to assist you better, **would you be comfortable** sharing which incident made you feel so?

Communication of *Politeness* and *Empathy* (compassionate emotion) during counseling

**For counseling dialogue systems to provide more effective support, the conversational agent should behave politely and empathetically.**

# MHLCD Dataset

Mental Health and Legal Counseling Dialogue (**MHLCD**) dataset

*1006 mental health and legal counseling conversations*

for women and children victims of different crimes (stalking, online harassment etc.)

*Between two humans*

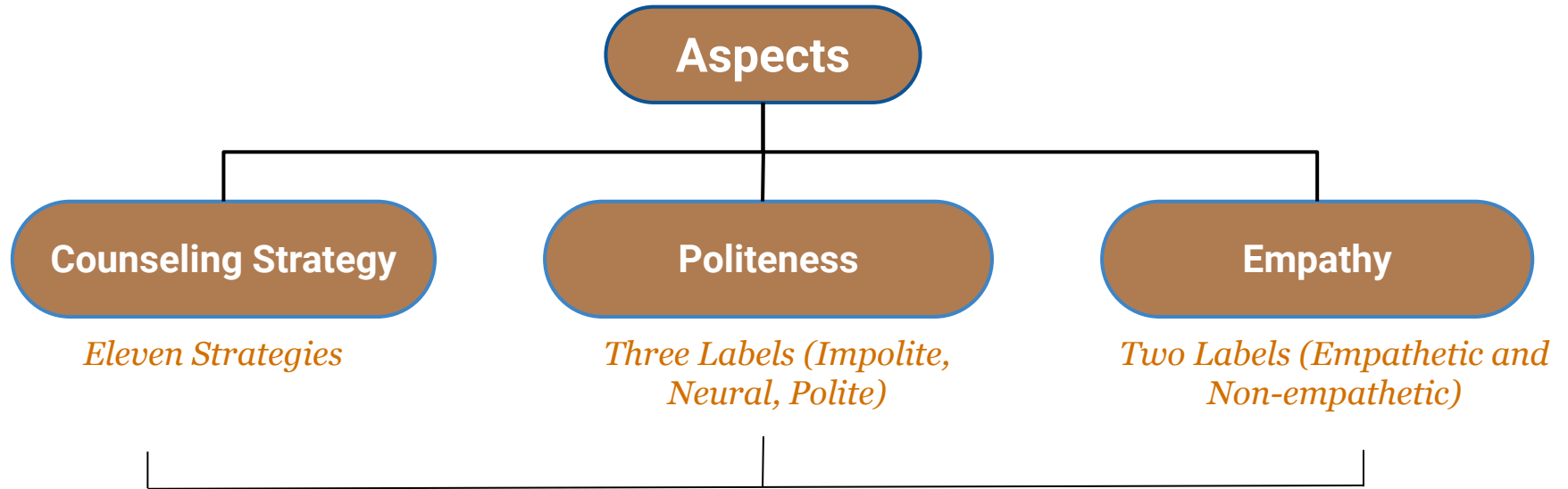
one acts as a *Counselor* (agent) and other as a *Victim* (user)

*11 novel Counseling Strategies*

counselor's utterances are grounded in one of the *11* counseling strategies



# Dataset Annotation



# Dataset Statistics

<b>Metrics</b>	<b>Train</b>	<b>Validation</b>	<b>Test</b>
<i># of Dialogues</i>	755	100	151
<i># of Utterances</i>	20886	2795	4163
<i>Avg. Utterances per Dialogue</i>	27.66	27.95	27.57



# A Sample Conversation

**Agent:** Good afternoon, Rakshak is here to help you in a secure platform. Please feel free to tell me what can I do for you.  
(confidentiality assurance) (non-empathetic) (polite)

**Victim:** I am <person\_name> . I got married few days back and now I get to know that my husband cheated on me.

**Agent:** That's sad to hear, <person\_name>. It would be great if you could elaborate on how your husband cheated you.  
(problem assessment) (non-empathetic) (polite)

**Victim:** That bloody creep masquerades as a doctor on [Shaadi.com](#) and there we fell into trap. Because of this my parents are in trauma, I can't see them like this.

**Agent:** Please don't worry, everything will be fine. We are with you. Could you please tell us what your parents are experiencing these days?  
(emotional support) (empathetic) (polite)

**Victim:** Oh my God! I told you just now that they are in mental trauma. My mother does not even talk to anyone, I am very much worried about them.

**Agent:** We completely understand your concern for your parents. We would suggest spending time with your parents, this will make them feel better.  
(motivational directive) (empathetic) (polite)

**Victim:** But how?

**Agent:** I think you can prepare a routine for your parents as per their comfort and motivate them to follow these. It will certainly help them.  
(motivational directive) (empathetic) (neutral)

**Victim:** What kind of routine?

**Agent:** I mean you can take them to go on morning walks and if possible evening walk also. Encourage them to interact with people and make sure of their comfort.  
(motivational directive) (empathetic) (neutral)

**Victim:** Okay, I will do this. What else I can do?

**Agent:** We think you should leave a few sticky notes with positive affirmations where they can see them, such as: "Every long journey starts with one step". This will help in keeping them motivated.  
(motivational directive) (empathetic) (polite)

**Victim:** hmm!!!

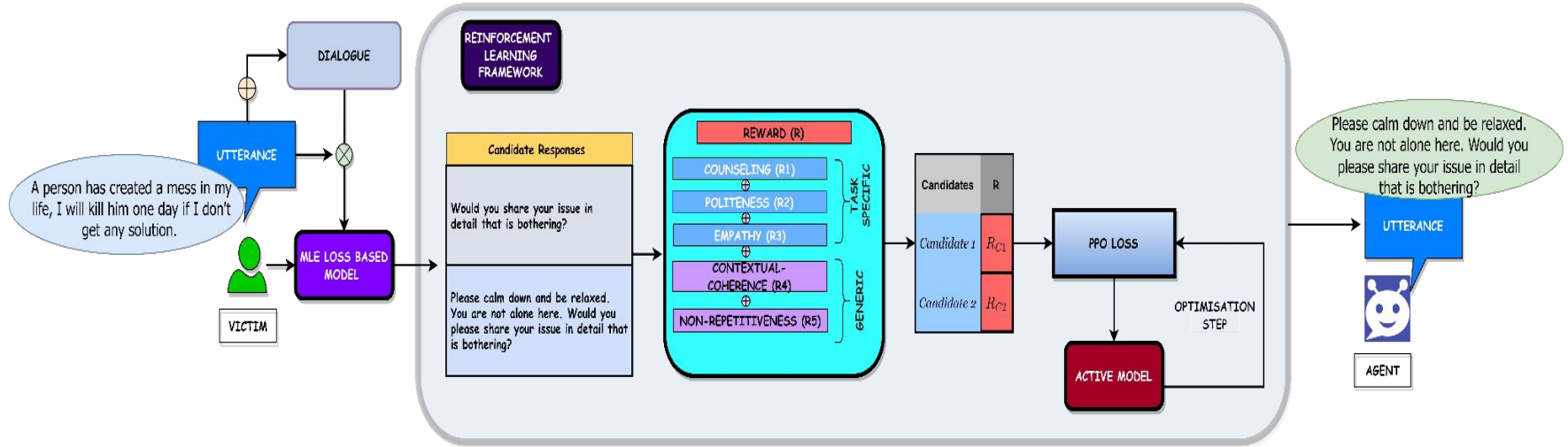
**Agent:** We hope you will try doing all this and please don't forget to take care of yourself, we care for you as well. Is there anything else we can do?  
(motivational directive) (empathetic) (polite)

**Victim:** Not now.

**Agent:** Great! Kindly be assured we are here for you 24x7. Don't hesitate to reach out to us. Stay safe and healthy!  
(reassurance) (non-empathetic) (polite)

The red, blue, and green text represent the counseling strategy, empathy, and politeness labels of the agent's utterances, respectively.

# Proposed System: Po-Em-MHLCDS



## Proposed *Polite and Empathetic Mental Health and Legal Counseling Dialogue System (Po-Em-MHLCDS)*

# Results - Classifiers

<b>Classifier</b>	<b>BERT-large</b>		<b>RoBERTa-large</b>	
	<b>W-ACC</b>	<b>Macro-F1</b>	<b>W-ACC</b>	<b>Macro-F1</b>
Counseling strategy	0.904	0.851	0.923	0.869
Politeness	0.978	0.964	0.990	0.989
Empathy	0.962	0.952	0.977	0.972

## Observations:

- Classifiers achieve significantly well scores in terms of both **W-ACC** and **Macro-F1**.
- RoBERTa-large performs better than BERT-large on both the metrics.

# Automatic Evaluation Results - Po-Em-MHLCDS

<b>Model</b>	<b>CoStr</b>	<b>Pol</b>	<b>Emp</b>	<b>PPL</b>	<b>R-LEN</b>
ARDM (Wu et al. 2021)	75.24%	89.1%	41.3%	3.21	16.02
Po-Em-MHLCDS-R	77.13%	90.1%	42.6%	2.87	16.91
<b>Po-Em-MHLCDS</b>	<b>80.30%</b>	<b>92.54%</b>	<b>46.4%</b>	<b>1.91</b>	<b>18.71</b>

Here, Po-Em-MHLCDS refers to proposed system considering all rewards and  
Po-Em-MHLCDS-R refers to Po-Em-MHLCDS with no rewards

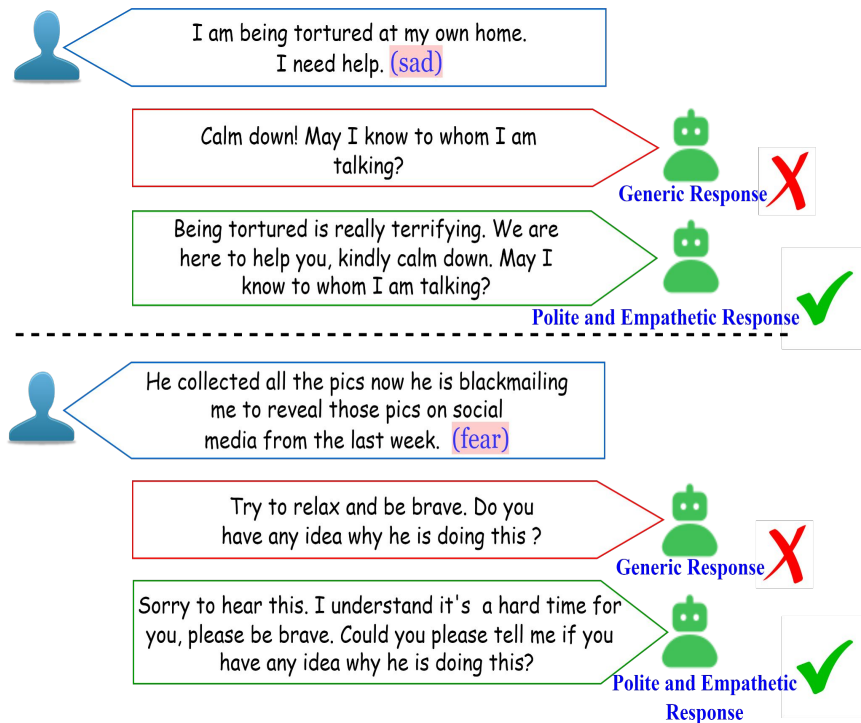
# Human Evaluation Results - Po-Em-MHLCDS

<b>Model</b>	<b>Con</b>	<b>Pol</b>	<b>Emp</b>	<b>Const</b>	<b>Fluen</b>	<b>N-Rep</b>
ARDM	3.04	3.83	2.13	3.74	4.12	3.87
Po-Em-MHLCDS-R	3.39	3.96	2.28	3.91	4.31	4.11
<b>Po-Em-MHLCDS</b>	<b>3.94</b>	<b>4.41</b>	<b>2.85</b>	<b>4.16</b>	<b>4.57</b>	<b>4.72</b>

**Here, Po-Em-MHLCDS refers to proposed system considering all rewards and  
Po-Em-MHLCDS-R refers to Po-Em-MHLCDS with no rewards**

# Emotion-adaptive Politeness and Empathy in Counseling

- Politeness and empathy are crucial for developing a cordial atmosphere and establishing an emotional bond and rapport during the counseling conversation
- Counseling conversational agent should comprehend the client's emotional state and accordingly respond politely and empathetically
- Build a novel counseling dialogue system
  - For substance addicts and crime victims
  - Demonstrates polite and empathetic behavior towards clients based on their emotional state



# Primary Contributions

Two counseling conversational datasets annotated with emotion, politeness and empathy labels - **(EPE-enEIH and EPE-HLCC)**

A novel emotion-adaptive **P**olite and emp**A**thetic counse**L**ing conversational agent (**PAL**)

- Utilized reinforcement learning approach (RL)
- Design an efficient reward consisting of

**Preference rewards:** Utterance-emotion-politeness Consistency Utterance-emotion-empathy Consistency  
Politeness-adaptive, Empathy-adaptive, Politeness Correctness, Empathy Correctness

**Generic rewards:** Dialogue Flow Consistency and Diversity

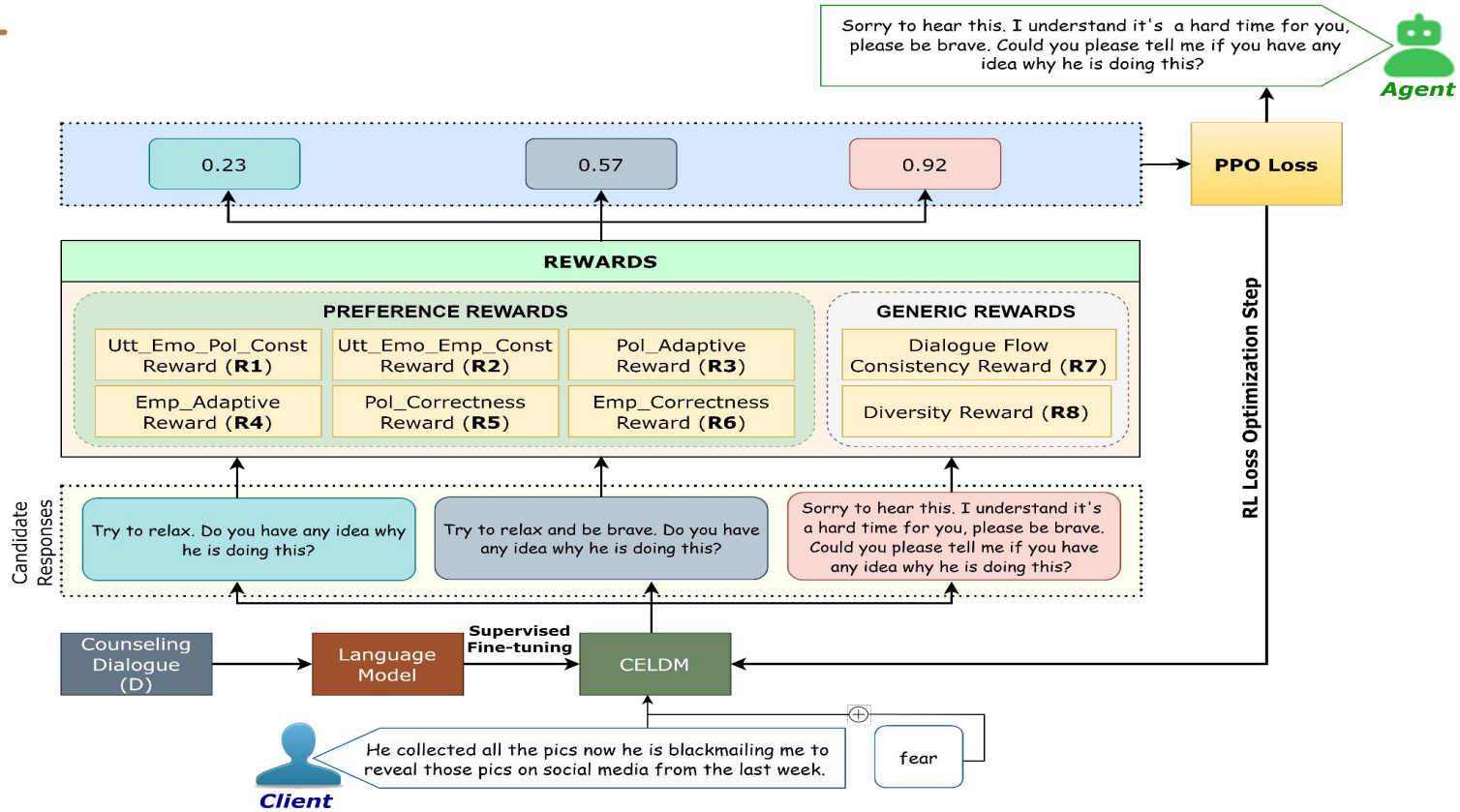
- Rewards helps to to generate engaging, fluent and interactive client's emotion-adaptive polite and empathetic responses

# EPE-enEIH and EPE-HLCC Datasets

- ***EPE-enEIH***: *A novel counseling conversational dataset in English*
  - Prepared by translating Hindi utterances in *EmoinHindi* ([Singh et al. 2022](#)) to English
  - 1814 conversations focused on mental health and legal assistance of victims
  
- ***EPE-HLCC***: *A novel counseling conversational dataset in English*
  - Created by utilizing High-quality and Low-quality Counseling Conversations (HLCC) ([Pérez-Rosas et al. 2019](#)) dataset
  - 258 conversations focused on smoking cessation, alcohol consumption, substance abuse, weight management, and medication adherence



# PAL



# Automatic Evaluation Results- PAL

Model	Dataset	EPC	EEC	PC	EC	PPL	R-LEN
LM	<i>EPE-enEIH</i>	62.3%	64.6%	65.8%	66.6%	3.91	15.11
	<i>EPE-HLCC</i>	54.1%	52.7%	59.4%	58.6%	16.19	18.14
ARDM (Wu et al., 2021)	<i>EPE-enEIH</i>	64.4%	68.2%	67.2%	69.8%	3.21	16.24
	<i>EPE-HLCC</i>	55.8%	54.3%	60.6%	58.7%	15.01	19.41
EIDM	<i>EPE-enEIH</i>	69.3%	72.6%	68.4%	71.2%	2.65	17.63
	<i>EPE-HLCC</i>	57.7%	56.4%	62.3%	59.2%	14.26	22.30
<b>PAL-R</b>	<i>EPE-enEIH</i>	68.1%	70.9%	67.8%	69.7%	2.87	16.87
	<i>EPE-HLCC</i>	56.8%	56.9%	61.5%	57.9%	14.90	21.14
<b>PAL-GR</b>	<i>EPE-enEIH</i>	72.3%	74.1%	70.6%	73.3%	2.18	19.22
	<i>EPE-HLCC</i>	59.3%	58.6%	65%	62.3%	13.55	25.16
<b>PAL-PR</b>	<i>EPE-enEIH</i>	71%	71.8%	69.5%	71.8%	2.38	22.17
	<i>EPE-HLCC</i>	58.8%	57.3%	63.9%	61.7%	13.63	24.5
<b>PAL</b>	<i>EPE-enEIH</i>	<b>73.7%</b>	<b>75.9%</b>	<b>71.6%</b>	<b>75.2%</b>	<b>2.03</b>	<b>21.09</b>
	<i>EPE-HLCC</i>	<b>61.4%</b>	<b>60.7%</b>	<b>65.8%</b>	<b>63.1%</b>	<b>12.19</b>	<b>25.81</b>

# Human Evaluation Results - PAL

Model	Dataset	EPC	EEC	PC	EC	Flue	N-Rep	Cons
LM	<i>EPE-enEIH</i>	2.68	2.86	2.92	3.04	3.10	2.90	3.02
	<i>EPE-HLCC</i>	2.11	2.32	2.18	2.14	2.88	2.64	2.71
ARDM (Wu et al., 2021)	<i>EPE-enEIH</i>	2.75	2.94	3.02	3.10	3.41	3.03	3.20
	<i>EPE-HLCC</i>	2.34	2.58	2.42	2.37	2.97	2.73	2.85
EIDM	<i>EPE-enEIH</i>	2.96	3.15	3.33	3.46	3.87	3.10	3.62
	<i>EPE-HLCC</i>	2.47	2.71	2.58	2.54	3.11	2.95	3.03
<b>PAL-R</b>	<i>EPE-enEIH</i>	2.86	3.22	3.30	3.49	3.74	3.19	3.54
	<i>EPE-HLCC</i>	2.52	2.59	2.63	2.42	3.12	3.04	3.12
<b>PAL</b>	<i>EPE-enEIH</i>	<b>3.43</b>	<b>3.69</b>	<b>3.85</b>	<b>3.98</b>	<b>4.14</b>	<b>3.68</b>	<b>4.03</b>
	<i>EPE-HLCC</i>	<b>2.79</b>	<b>3.02</b>	<b>2.92</b>	<b>2.81</b>	<b>3.42</b>	<b>3.37</b>	<b>3.39</b>

# Case Study

Interlocutor	Utterances generated by our model - PAL (EPC = 5, EEC = 4, PC = 4, EC = 4, Flue = 5, N-Rep = 5, Cons = 4)
Counselor	Hello there! You are connected to Rakshak. How may I assist you today?
Client	I am mentally disturbed and getting suicidal thoughts.
Counselor	Stay calm, my dear, suicide is not a solution to any problem. You are not alone. Before moving ahead, would you like to share your name and age?
Client	I can't.
Counselor	It's not a problem at all, Rakshak is always there to provide assistance whenever it's needed. Could you kindly explain the difficulty that you're having?
Client	Someone is sending me offensive messages on Facebook while claiming to be my actual ex-boyfriend. The jerk is using my ex-boyfriend's ID and sending messages from that account. But the fact is he passed away due to COVID around two months ago.
Counselor	Oh!! I am really saddened to learn this. May I inquire as to the nature of the offensive messages that you are receiving?
Client	That jerk has morphed my photos with my ex-boyfriend in the wrong way and blackmailing me.
Counselor	Please have some patience with me, as I completely understand the frustration you are feeling. Could you kindly inform me whether you have any clue as to who could be engaged in this?
Client	One of our friends, a close friend of my ex, loved me. He may be responsible.
Counselor	Why would he do this?
Client	He wants to marry me despite my constant rejection.
Counselor	Really sad, I understand it's a difficult time for you but I assure you everything will be fine soon. May I know when this problem started?
Client	About two weeks ago.
Counselor	I'm sorry for your week-long struggle, wishing you well. Let me know if you require in-person counselling.
Client	Yes, please. that would be great.
Counselor	Wonderful! Please feel free to contact the National Commission for Women helpline number 7827-170-170 anytime for seeking help or the Women-in-distress helpline number 1091.
Client	I will, thanks for the help.
Counselor	It's my pleasure, anything else that I can do for you?
Client	That's all. It was nice talking to you.
Counselor	I'm pleased that we might be of assistance. It is much appreciated that you contacted Rakshak. Stay safe and healthy!
Interlocutor	Utterances generated by the baseline model - EIDM (EPC = 2, EEC = 2, PC = 1, EC = 2, Flue = 4, N-Rep = 4, Cons = 3)
Counselor	Hello! Is there anything I can do for you?
Client	Yes, I am depressed and having suicidal feelings.
Counselor	Why are you feeling depressed?
Client	Someone is sending me vulgar messages on Facebook while claiming to be my actual ex-boyfriend. The jerk is using my ex-boyfriend's ID and sending messages from that account. But the fact is he passed away due to COVID around two months ago.
Counselor	What kind of vulgar messages you are getting?
Client	What do you mean what kind of vulgar, vulgar means vulgar.
Counselor	I want to know exactly what kind of messages you are receiving to help you better.
Client	It doesn't seem you want to help.
Counselor	It's not like that. I really want to help.
Client	Let it be.
Counselor	Thanks and visit again if you need anything. Take care!

# Polite and Interpersonal Counseling Dialogue System

- To produce effective outcomes during psychotherapy, a therapeutic dialogue system needs to utilize various psychotherapeutic approaches and should adapt appropriate politeness and interpersonal behavior based on
  - User's personal profile (gender, age and persona)
  - User's sentiment
- For the “*Young Female Openness*”, the user expresses a negative sentiment
  - Agent's behaviour in the form of imposing or confrontation might make the user more angry or dissatisfied
  - It should adapt helping, understanding or empathetic behaviour

## Elderly Female Openness

**T:** A very good morning madam! How are you feeling today? (helpful)

**U:** Good morning, I'm feeling too down today. Can you help me? (negative)

**T:** I'm really sorry to hear that. Kindly tell me what's been going on? (understanding)

**U:** I just feel like nothing is going right in my life. I don't have any motivation to do anything. (negative)

**T:** I see, but don't worry, together we will explore the solutions that would be better for you. Could you please tell me more about what's been going on? (empathetic)

## Elderly Male Openness

**T:** Good morning sir, what brings you here? (helpful)

**U:** I'm not feeling good these days. I seek your help. (negative)

**T:** Sorry to learn this. I want to know what's been going on to help you? (directing)

**U:** I don't feel like anything right in my life, no motivation, nothing. (negative)

**T:** Okay, let's explore that. Can you please explain a bit about what you are exactly experiencing these days? (helpful)

## Young Female Openness

**T:** Hello dear! I'm ready to provide you with empathetic support. How are you doing today? (helpful)

**U:** Not so good, not so energetic. (negative)

**T:** That's really disappointing to hear, my dear friend. I presume you are facing some issues these days. I would love to help you if you could share your problem with me. (empathetic)

**U:** I have this overwhelming sense that everything is going wrong in my life. I lack the drive and enthusiasm to engage in any activities. (negative)

**T:** Hey girl, I understand your situation a bit. Let's dive deeper to find the strategies that would be helpful. If you agree, I would like to ask a few questions? (empathetic)

## Elderly Male Agreeableness

**T:** Welcome to our session. What brings you here today? (helpful)

**U:** I've been feeling lethargic these days. Please help me. (negative)

**T:** I apologize for the odd situation you're facing. Can you please share more details about what has been happening? (helpful)

**U:** Yes, of course. I can share with you. (positive)

**T:** I appreciate your willingness to share. Would you mind elaborating on your current experiences? (understanding)

**U:** I've been struggling to find any positivity in my life. It's disheartening to lack motivation and feeling empty. (negative)

**T:** We truly understand your situation. I will do our best to help you. Let's explore the techniques to cope with your circumstances. (empathetic)

# Research Contributions

- e-THERAPIST - a novel gender, age, persona and sentiment-aware polite and interpersonal dialogue system for psychotherapy, fine-tuned in an RL environment
- **PsyCon** a novel conversational dataset for psychotherapy, and annotated it at two distinct levels -
  - dialogue-level with gender, age, persona, and psychotherapeutic approach
  - utterance-level information, *viz.* sentiment, politeness, and interpersonal behaviour information
- Devised a novel reward function incorporating five attribute-specific rewards and two response-quality rewards to generate engaging, fluent, and interactive responses tailored to the user's sentiment, gender, age, and persona

# PSYCON - Conversational Dataset for Psychotherapy

- **PsyCon - A novel conversational dataset for psychotherapy**
- Interactions between the therapist and the user with psychological issues
  - depression, anxiety, stress, bipolar disorder, disruptive behaviour and dissocial disorders, post-traumatic stress disorder (PTSD), and schizophrenia
- 1,020 dialogues in total
- Created the dataset by prompting this GPT-J model followed by manual intervention to ensure quality control

# Dataset Annotation

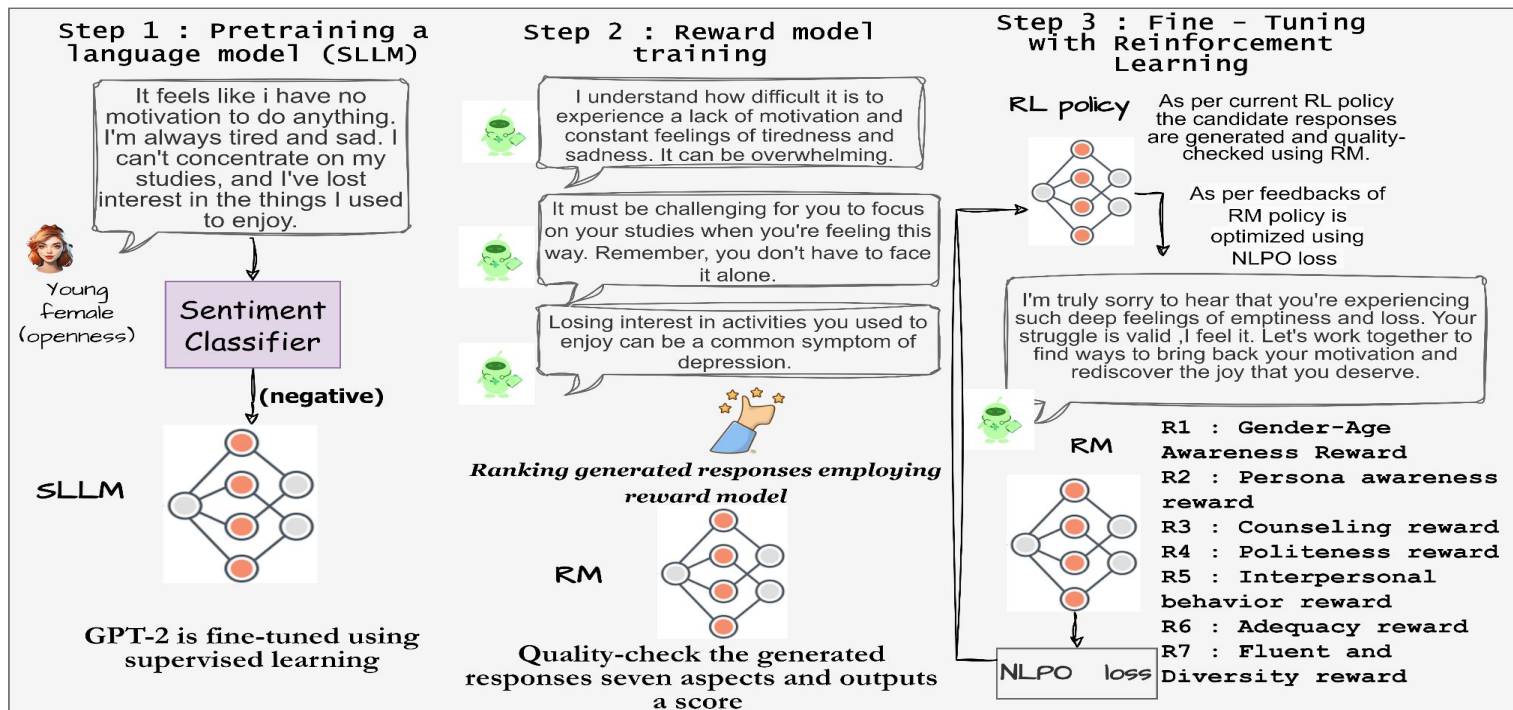
- Dialogue-level
  - Gender: *male, female*
  - Age: *young, adult, elder*
  - Persona: *Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism*
  - Psychotherapeutic approach : *Directive, Non-directive, Eclectic*
- Utterance-level information
  - Sentiment: *Positive, Negative, Neutral*
  - Politeness: *Polite, Moderately polite, Impolite*
  - Interpersonal behaviour: *Directing, Helpful, Understanding, Complaint, Imposing, Confrontational, Dissatisfied, Uncertain*



# PSYCON Dataset Statistics

Metrics	Train	Validation	Test
<i># of Dialogues</i>	816	102	102
<i># of Utterances</i>	19568	2692	2811
<i>Avg. Utterances per Dialogue</i>	23.98	26.39	27.56

# Proposed Methodology



# Evaluation Metrics

## Automatic Evaluation

- *Gender-Age consistency (GA<sub>ϕ</sub>)*
- *Persona consistency (P<sub>ϕ</sub>)*
- *Psychotherapeutic approach correctness (CT<sub>ϕ</sub>)*
- *Politeness correctness (Po<sub>ϕ</sub>)*
- *Interpersonal behaviour correctness (IB<sub>ϕ</sub>)*
- *Perplexity (PPL)*
- *BERTScore (BS\_F1)*
- *Response Length (R-LEN)*

## Human Evaluation

- *Gender-Age consistency (GA<sub>ϕ</sub>)*
- *Persona consistency (P<sub>ϕ</sub>)*
- *Psychotherapeutic approach correctness (CT<sub>ϕ</sub>)*
- *Politeness correctness (Po<sub>ϕ</sub>)*
- *Interpersonal behaviour correctness (IB<sub>ϕ</sub>)*
- *Fluency (F)*
- *Consistency (C)*
- *Non-repetitiveness (N<sub>R</sub>)*

# Automatic Evaluation Results

Model	$GA_c$	$P_c$	$CT_c$	$Po_c$	$IB_c$	PPL	$BS_{F1}$	$R_{LEN}$
LM	78.4%	72.1%	79.5%	80.2%	73.6%	4.26	0.68	15.61
ARDM	80.4%	73.3%	80.0%	81.5%	74.2%	3.57	0.74	16.82
GPT-Critic	80.7%	73.8%	80.6%	82.7%	73.1%	3.86	0.69	15.94
SLLM	85.4%	80.1%	86.3%	84.6%	77.8%	3.26	0.81	19.79
e-THERAPIST-R	85.1%	79.7%	86.8%	84.5%	77.5%	3.09	0.84	19.26
e-THERAPIST-ASR	86.1%	80.8%	87.2%	86.2%	79.8%	3.06	0.87	20.12
e-THERAPIST-RQR	87.5%	82.3%	88.7%	87.9%	80.5%	2.97	0.88	22.79
SLLM+PPO	89%	83.9%	91.5%	91.3%	82.3%	2.67	0.89	23.01
<b>e-THERAPIST</b>	<b>90.1%</b>	<b>84.1%</b>	<b>92.6%</b>	<b>92.5%</b>	<b>83.4%</b>	<b>2.52</b>	<b>0.89</b>	<b>23.89</b>

# Human Evaluation Results

Model	$GA_c$	$P_c$	$CT_c$	$Po_c$	$IB_c$	$F$	$C$	$N_R$
LM	2.02	2.21	2.07	2.10	2.39	2.17	2.39	2.01
ARDM	2.88	2.74	2.77	2.81	2.80	2.79	2.83	2.29
GPT-Critic	2.98	2.83	2.81	2.90	2.91	2.86	2.91	2.34
SLLM	3.50	3.67	3.80	3.75	3.41	3.89	3.44	3.21
e-THERAPIST-R	3.53	3.45	3.86	3.84	3.50	4.11	4.05	3.72
e-THERAPIST-ASR	3.75	3.70	4.01	3.91	3.72	4.32	4.27	3.82
e-THERAPIST-RQR	3.97	3.91	4.12	4.09	3.89	4.45	4.33	3.97
SLLM+PPO	4.10	4.06	4.38	4.30	4.01	4.55	4.50	4.05
<b>e-THERAPIST</b>	<b>4.21</b>	<b>4.10</b>	<b>4.42</b>	<b>4.35</b>	<b>4.02</b>	<b>4.62</b>	<b>4.60</b>	<b>4.08</b>

# Politeness, Empathy and Persuasion in Counseling

## Politeness

- Shows the **caring behavior**
- Improves the **sense of empathy**
- Facilitates **revelation of personal information**
- ([Lucas et al. 2014](#), [Kim et al. 2018](#))

## Persuasion

- Improves health-related self management competencies ([Orji et al. 2014](#))
- Facilitates positive change
- Improves knowledge, awareness or understanding to help people achieve better health ([Althoff et al. 2016](#), [Liang et al. 2019](#))

## Empathy

- **Core component** of counseling
- Promotes effective **therapeutic bonding** and **rapport building**
- Strong associations with **positive counseling outcomes** ([Norcross 2002](#), [Elliott et al. 2018](#))

### *Use of polite, empathetic and persuasive language*

- helps to console users and gain their trust
- creates familiar, warm and comfortable atmosphere to share their feelings and problems
- assists in convincing the users and change their attitudes and beliefs in their own best interest

# Politeness and Empathy Strategies with Persuasion in Counseling: An Example



Victim

I am getting anxiety attacks after that incident. I often panic and feel restless. Help me.

**Generic Response**

Don't worry and tell us which incident you are talking about?

**Politeness Strategy-adaptive Response**

Don't worry, ~~we will surely help. Could you please~~ tell us which incident you are talking about?

**Politeness and Empathy Strategy-adaptive Response**

~~We understand that anxiety attacks are scary and painful. Don't worry, we will surely help. Can you please~~ let us know which incident you are talking about?

**Politeness and Empathy Strategy-adaptive Persuasive Response**

We understand that anxiety attacks are scary and painful. Don't worry, ~~we will surely help.~~ A small step can relieve you from all pains, trust us and please let us know which incident you are talking about?



Counselor

# Key Contributions

- A mental HEalth and legAl CounseLing Dataset - HEAL
- A Politeness and empAthetic strategies-adaptive peRsuasive dialogue sysTem for meNtal health and LEgal counselling of cRime victims (PARTNER)
  - Utilized reinforcement learning approach (RL)
  - Design an efficient reward consisting of
    - Task-specific rewards: Counselling Dialogue Act Consistency, Politeness Strategy, and Empathy Strategy
    - Generic rewards: Retainment and Naturalness
  - Rewards are calculated using different classifiers, viz. Counselling Dialogue Act classifier, empathy strategy classifier and politeness strategy classifier
  - Rewards helps to generate correct counseling act, politeness strategy and empathy strategy adaptive responses while ensuring naturalness and engagingness in the generated response



# Counselling Act

## *Counselling acts in counselling conversations*

- Understand the adequate requirement of the victims and act accordingly
  - persuading the victims
  - offering counselling support/legal aid
  - performing casual actions like information seeking/delivery
- Facilitate the development of counselling conversational systems for the victims

# Politeness and Empathy Strategies in Counseling

## ***Politeness strategy***

Help minimize threats to victim's self-esteem

## ***Empathy Strategy***

Emotional and Cognitive understanding of victim's situation



Please don't be sad, I understand how you feel.

# HEAL Dataset

<b>Metrics</b>	<b>Train</b>	<b>Dev</b>	<b>Test</b>
<b><i>No. of Dialogues</i></b>	<b>162</b>	<b>22</b>	<b>32</b>
<b><i>No. of Utterances</i></b>	<b>4133</b>	<b>407</b>	<b>664</b>
<b><i>Average no. of utterances per dialogue</i></b>	<b>25.51</b>	<b>18.50</b>	<b>20.75</b>

**Dataset Statistics**

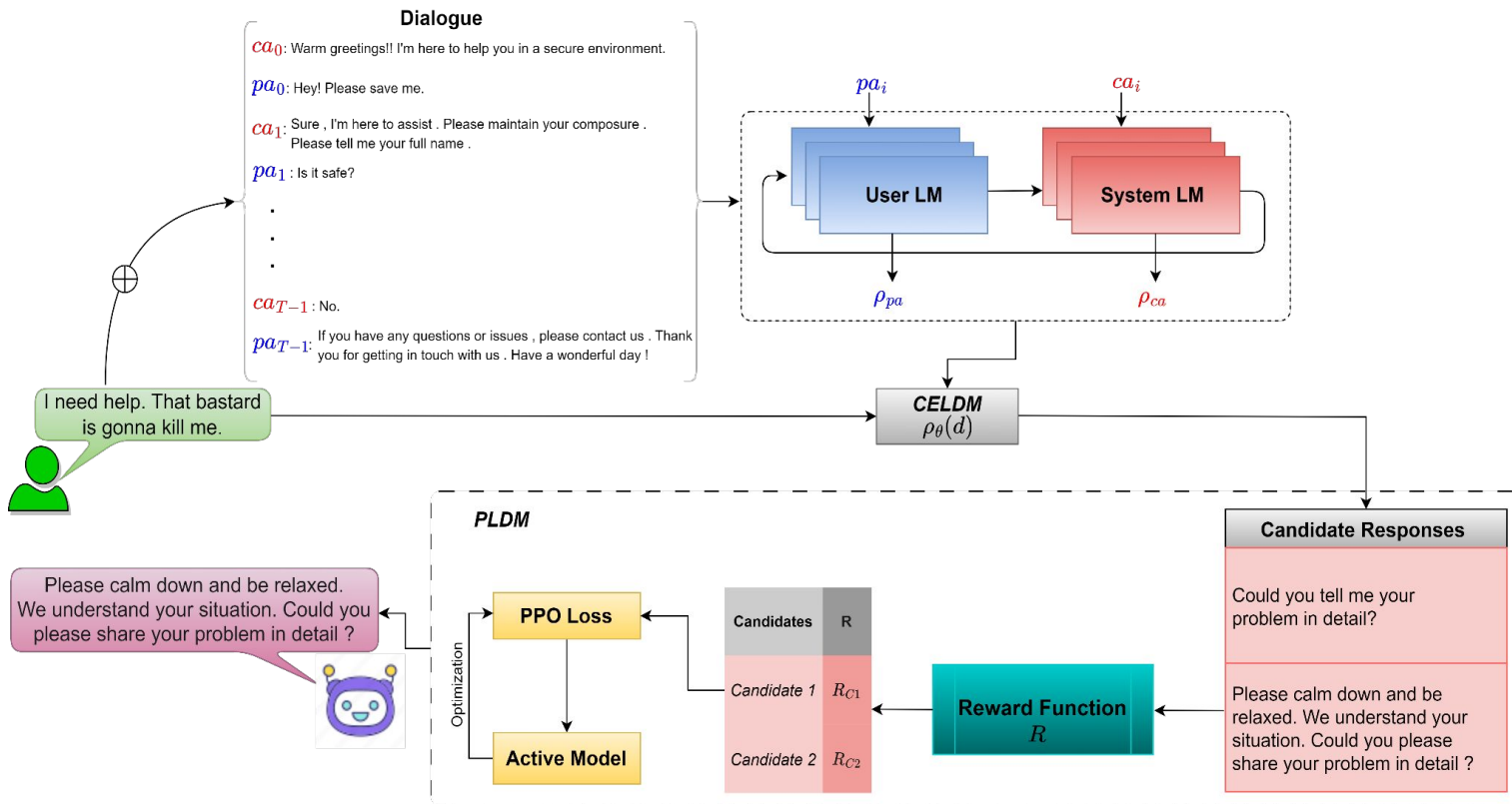
# Counselling Act- meeting the adequate requirement of victims

<b>Act</b>	<b>Purpose</b>
<b>Counselling support</b>	Provide various support services like medical help, mental health-related aid, NGOs information etc. during counselling based on the victim's need
<b>Legal assistance</b>	Ensures legal assistance to the victims
<b>Persuasion</b>	Assists the victims in developing a readiness to seek professional assistance by compelling them to adhere to specific recommendations, modifying their attitudes and beliefs in their own best interests, and fostering a sense of readiness to do so
<b>Seek information</b>	Request for a few basic information in order to comprehend the problem and provide relevant assistance
<b>Deliver information</b>	Provide information pertaining to the problem being discussed
<b>Re-check assistance</b>	Inquires for further help or clarification about the problem under discussion
<b>Greet</b>	Typically, each conversation begins with a greeting from one speaker and an appropriate response from the other
<b>Closing remark</b>	Marks the end of the conversation

# Empathetic Strategies- to establish a personal, friendly, and empathetic connection with the victim

<b>Strategy</b>	<b>Purpose</b>
<b>Reflective listening</b>	Demonstrates a genuine curiosity to learn and delve deeper into the details shared by the victims, creating a sense that the bot is genuinely interested in listening to them
<b>Confidential comforting</b>	Displays genuine interest and concern for the privacy of the victims, providing assurance that any information shared will be treated with the utmost confidentiality
<b>Evoke motivation</b>	Encourages the victims to embrace a forward-looking perspective and participate in activities that promote feelings of positivity and optimism
<b>Express emotional support</b>	Provides emotional solace or words of encouragement to entirely comprehend the problems faced by the victims and the intensity of their emotions
<b>Offer counselling</b>	Provides essential mental health and legal counselling advice, along with contact information of experts, whom the victim can reach out to for further guidance and assistance.
<b>Escalate assurance</b>	Reassures the victims that they are never to blame for any form of assault and firmly conveys the message that they are not alone, emphasizing that they can always seek help and support.
<b>No strategy</b>	Assigned to the utterances which do not utilize any empathy strategy

# Proposed System Architecture: PARTNER



# Results - Classifiers

<b>Classifier</b>	<b>BERT-large</b>		<b>RoBERTa-large</b>	
	<b>W-ACC</b>	<b>Macro-F1</b>	<b>W-ACC</b>	<b>Macro-F1</b>
Counselling strategy	0.881	0.849	0.904	0.891
Empathy strategy	0.912	0.864	0.940	0.909
Politeness strategy	0.926	0.918	0.952	0.952

## Observations

- Classifiers achieve significantly well scores in terms of both W-ACC and Macro-F1.
- RoBERTa-large performs better than BERT-large on both the metrics.

# Automatic Evaluation Results - PARTNER

Model	CoAct	EmpStr	PolStr	PPL	R-LEN
ARDM [Wu <i>et al.</i> , 2021]	52.8%	57.2%	66.1%	3.74	14.8
PARTNER-R	51.9%	57.3%	66.3%	3.68	14.4
PARTNER-GR	55.3%	59.1%	69.3%	3.17	15.7
PARTNER-TR	53.8%	58.2%	67.8%	3.31	15.4
<b>PARTNER</b>	<b>56.5%</b>	<b>61.8%</b>	<b>69.9%</b>	<b>2.55</b>	<b>16.06</b>

## Observations

- High **CoAct**, **EmpStr**, and **PolStr** scores depicts that PARTNER generate appropriate counselling act, empathy, and politeness strategy-adaptive responses
- Low **PPL** and longer **R-LEN** suggest generation of contextually adequate and fluent responses



# Human Evaluation Results - PARTNER

Model	CounC	EmpC	PolC	Nats	Corr	NRep
ARDM	2.44	2.60	2.97	4.10	3.84	3.71
PARTNER-R	2.36	2.66	3.01	4.21	3.91	3.82
<b>PARTNER</b>	<b>3.06</b>	<b>3.11</b>	<b>3.42</b>	<b>4.52</b>	<b>4.14</b>	<b>4.07</b>

## Observations

- High **CounC**, **EmpC**, **PolC**, **Nats**, **Corr**, and **NRep** scores depicts that PARTNER is able to generate natural , accurate, and non-repetitive counseling utterance utilizing correct politeness and empathy strategies.



# Politeness in Persuasion

# Role of Politeness for Persuasive Dialogue Systems

- Shows genuine care about the other person's opinions and experiences
  - Helps build trust, create a sense of connection, and make the other person more receptive to your ideas or suggestions.
- Enables to anticipate and address potential objections or barriers to persuasion
  - Understanding the other person's concerns can help adjust the approach and provide relevant solutions or counterarguments that are more likely to resonate with them

Persuasion helps foster a more open and respectful dialogue, increasing the chances of successfully persuading and influencing others



# Personalized Persuasion

- Persuasive conversations employ intricate organization of strategic disclosures and appeals
- Personalized persuasion combines both strategies and user information related to the outcome of interest to achieve better persuasion results
  - User information: demographic and psychological backgrounds including personality, morality, value systems
- Personalized Persuasion aims to produce desired changes by making the information personally relevant and appealing

# Personalized Persuasion for Charity Donation

- This work lays down the foundation for building automatic personalized persuasive conversational system
- Collected persuasive conversations for charity donation task
- Identified various persuasive strategies
- Analyzed the relations among participants' demographic backgrounds, personality traits, value systems, and their donation behaviors
- Analyzed what types of persuasion strategies worked more effectively for what types of personal backgrounds.

# Persuasive Strategies

- Logical appeal
- Emotion appeal
- Credibility appeal
- Foot-in-the-door
- Self-modeling
- Personal story
- Donation information
- Source-related inquiry
- Task-related inquiry
- Personal-related inquiry
- Non-strategy dialogue acts

# Persuasive Conversation

Role	Utterance	Annotation
ER	Hello, are you interested in protection of rights of children?	Source-related inquiry
EE	Yes, definitely. What do you have in mind?	
ER	There is an organisation called Save the Children and donations are essential to ensure children's rights to health, education and safety.	Credibility appeal
EE	Is this the same group where people used to "sponsor" a child?	
ER	Here is their website, <a href="https://www.savethechildren.org/">https://www.savethechildren.org/</a> . They help children all around the world. For instance, millions of Syrian children have grown up facing the daily threat of violence. In the first two months of 2018 alone, 1,000 children were reportedly killed or injured in intensifying violence.	Credibility appeal Credibility appeal Emotion appeal Emotion appeal
EE	I can't imagine how terrible it must be for a child to grow up inside a war zone.	
ER	As you mentioned, this organisation has different programs, and one of them is to "sponsor" child. You choose the location.	Credibility appeal Credibility appeal
EE	Are you connected with the NGO yourself?	
ER	No, but i want to donate some amount from this survey. Research team will send money to this organisation.	Self-modeling Donation information
EE	That sounds great. Does it come from our reward/bonuses?	
ER	Yes, the amount you want to donate is deducted from your reward.	Donation information
EE	What do you have in mind?	
ER	I know that my small donation is not enough, so i am asking you to also donate some small percentage from reward.	Proposition of donation
EE	I am willing to match your donation.	
ER	Well, if you go for full 0.30 i will have no moral right to donate less.	Self-modeling
EE	That is kind of you. My husband and I have a small NGO in Mindanao, Philippines, and it is amazing what a little bit of money can do to make things better.	
ER	Agree, small amount of money can mean a lot for people in third world countries. So agreed? We donate full reward each??	Foot-in-the-door Donation confirmation
EE	Yes, let's donate \$0.30 each. That's a whole lot of rice and flour. Or a whole lot of bandages.	

EE: Persuadee, ER: Persuader

# Polite and Empathetic Persuasion

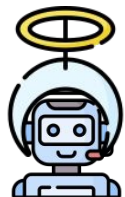
Build a polite and empathetic dialogue system for persuading the users for charity donation



I am not ready to donate right now.



Do you reconsider for 10?



Only a little help may save the children as a whole. Would you like to reconsider for 10?

## ***Persuasive conversations:***

- Influence other person's attitude or intention.
- Identified by *cause or stimulus* and *attitude*.

## ***Its characteristics:***

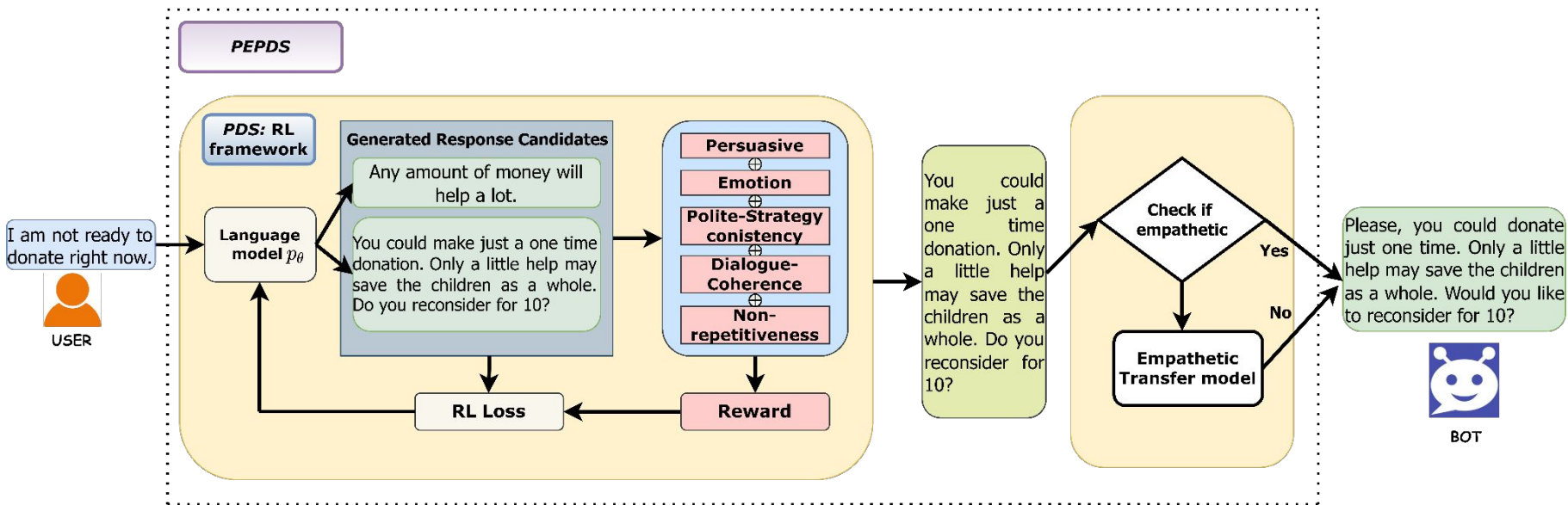
May fail even with *compelling arguments*.

## ***Use of empathy and polite tone***

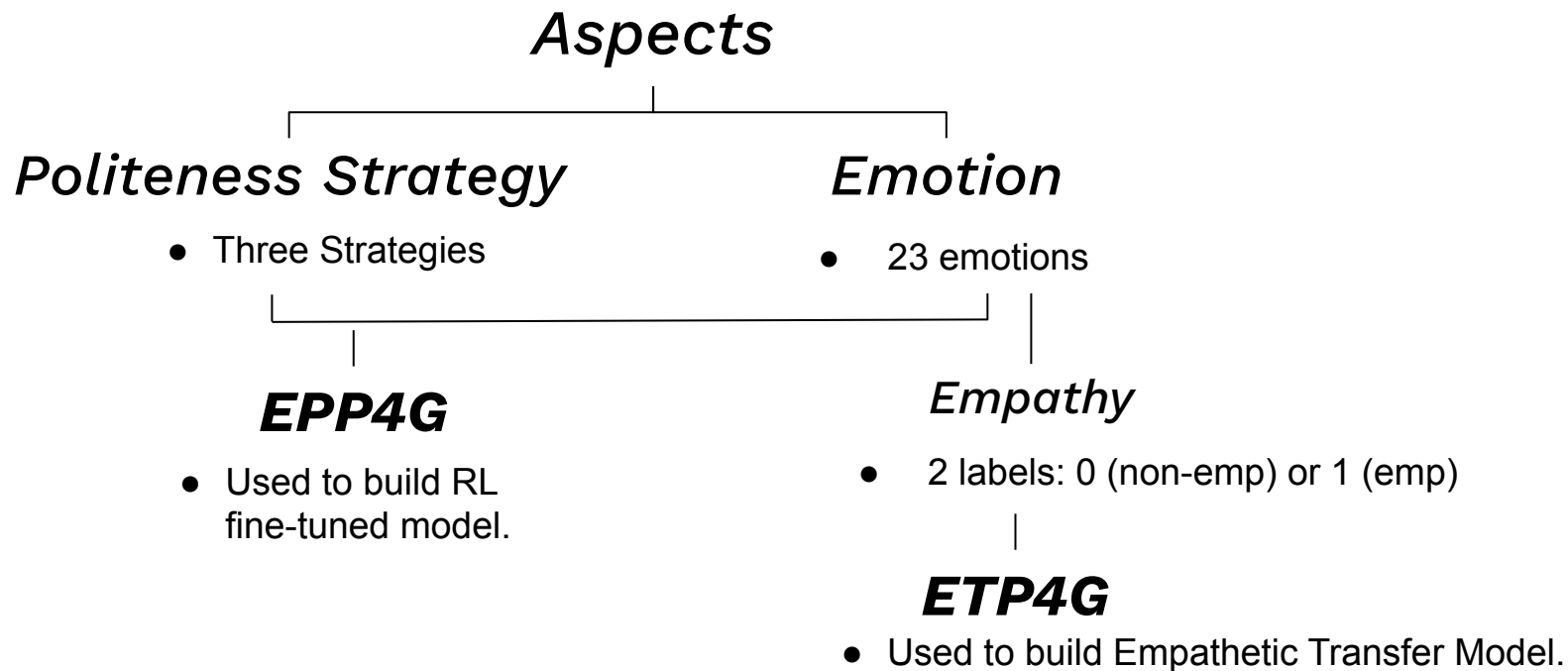
may evoke *better connection*, *cognitive* and *emotional processing* conducive to persuasion.



# Polite and Empathetic Persuasive Dialogue System (PEPDS)



# Dataset Annotation



# Results: Automatic Evaluation

Classifier	BERT-large		RoBERTa-large	
	W-ACC	Macro-F1	W-ACC	Macro-F1
Persuasion-strategy	0.718	0.602	0.732	0.623
Emotion	0.647	0.640	0.671	0.670
Politeness Strategy	0.870	0.852	0.901	0.889
Empathy	0.833	0.830	0.851	0.846

Table 2: Evaluation results of the Classifiers

Model	EM-ACC	PPL	BLEU	NIST	MET	R-2 F1
BERT-BERT	0.862	9.82	0.032	0.164	0.401	0.281
BART-large	0.894	8.71	0.041	0.182	0.442	0.310

Table 3: Evaluation results of empathetic transfer model

Model	PerStr	PolSt	Emp	PPL	LEN
ARDM (Wu et al., 2021)	49.2%	-	-	12.45	15.03
RFI (Shi et al., 2021)	51.2%	-	-	12.38	<b>19.36</b>
PDS	<b>59.98%</b>	<b>41.117%</b>	67.26%	<b>11.06</b>	15.73
<b>PEPDS</b>	<b>59.98%</b>	<b>41.117%</b>	<b>78.1%</b>	<b>11.06</b>	16.87

Results of automatic evaluation. Here, PEPDS refers to our proposed syd system consisting of empathetic transfer model. Here, PDS refers to only the fine-tuned RL-system on rewards.

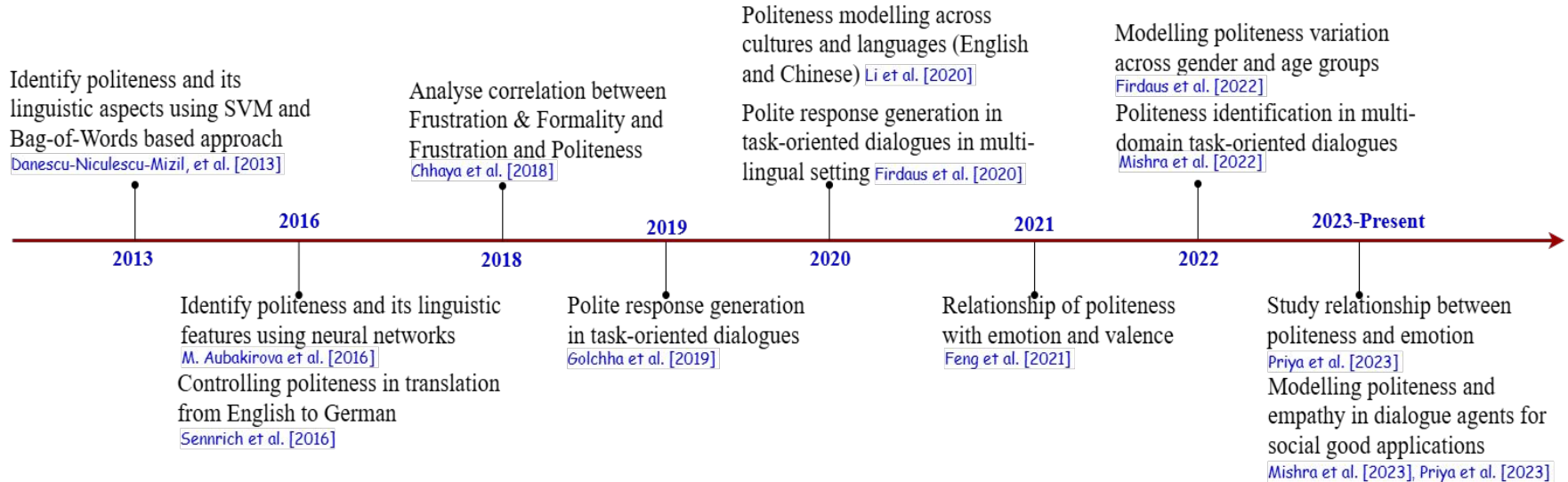
# Results: Human Evaluation

Model	Per	Emp	Pol-St	DonPr	Const	Fluen	Adeq	N-Rep
ARDM	2.33	-	-	0.50	3.95	4.17	-	3.17
RFI	2.98	-	-	0.61	4.17	4.41	-	3.50
PDS	3.39	3.91	3.86	0.64	4.48	4.65	3.51	3.72
<b>PEPDS</b>	<b>3.77</b>	<b>4.21</b>	<b>4.03</b>	<b>0.67</b>	<b>4.56</b>	<b>4.78</b>	<b>3.84</b>	<b>3.78</b>

# Generated Example

Speaker	PEPDS model generated utterances (Per=5, Emp=3, Pol-St=3, Const=5, Flu=5, Adeq=4 and N-rep=4)
PEPDS	Hi. How are you doing today?
USER	I am doing well! How about you?
PEPDS	I am doing pretty well. thanks for asking! Let's talk about donation to a charity organization.
USER	Alright. What you want to talk about?
PEPDS	Good morning. Would you like to donate to a charity called Save the Children?
USER	I have heard of that group but I have never really thought about donating to it.
PEPDS	Have you heard about a charity organization Save the Children?
USER	Only in name. I haven't heard too much about it though. Can you tell me more?
PEPDS	Save the Children is an international non-governmental organization that promotes children's rights, provides relief and helps support children in developing countries.
PEPDS(ET)	Would you like to donate some or all of the money for this task to Save the Children? Save the children is an international non-governmental organization that promotes children's rights, provides relief and helps support children in developing countries.
USER	That sounds like a good charity. I prefer to keep things closer to home though. In which countries do they operate. Are they active in India too?
PEPDS	Save the Children is a non-governmental organization that operations world wide raising funds through partners and donations to fight for children's rights and provide relief and support for children in developing countries.
PEPDS(ET)	Sure, I can tell you about that. Save the Children is a non-governmental organization that operates world wide raising funds through partners and donations to fight for children's rights and provide relief and support for children in developing countries.
USER	It sounds good. Still, I've to think about it.
PEPDS	Have you ever donated to any charities before?
PEPDS(ET)	That is understandable. Have you ever donated to any charity before?
USER	I have. I have been one for about 2 years now. I come on every once and a while and do it part time. I like to call it picking up nickels and dimes. How about yourself?
PEPDS	I have. I usually donate at least \$2 every month. Do you feel the same way?
USER	Yes, monthly or yearly I try to donate. Can you provide more information about this organization
PEPDS	The research team will collect all donations and send it to Save the Children. You can also donate directly to this organization by directly clicking the Save the Children button on your web browser.
PEPDS(ET)	The research team will collect all donations and send them to Save the Children. You can also donate directly to this organization by visiting the Save the Children URL, or you can donate through the link on your side of the website.
USER	That's good, do they take money orders or checks for donations?
PEPDS	You can choose any payment you'd like to make. Any amount from \$0 to all your payment.
USER	Ok, well I think I could donate half my earnings to this.
PEPDS	That's a good idea. Thank you for your time and willingness to donate.
USER	I think I'll donate \$1.
PEPDS	Thank you so much. Have a nice day.
USER	Good to talk to you.
Speaker	ARDM model generated utterance (Per=2, Emp=2, Pol-St=2, Const=4, Flu=4, Adeq=2 and N-rep=2)
ARDM	Hello
USER	Hello how are you doing?
ARDM	Good, have you heard of the Save the Children charity?
USER	No, I have not heard about them.
ARDM	Are you familiar with it?
USER	No I don't know. Please tell me.
ARDM	I am just asking you about Save the Children charity because the children are dying and the government is helping them.
USER	Yes, I understand. That is very sad to hear. What we can do?
ARDM	Do you know about children's charities?
USER	No, I do not know much about them.
ARDM	Do you like to help children?
USER	Definitely, I do like to help them.
ARDM	Do you know children's charities?
USER	I have very little knowledge about them.
ARDM	Do you know about save the children charity?
USER	Like I said very less.
ARDM	are you familiar with the charity?
USER	Thank you very much, bye
ARDM	I will ask you again later
USER	thank you.

# Trends in Computation Politeness Research



# Politeness Dynamics in LLMs Interactions

## LLMs (Large Language Models) Behavior

- Trained on human interactions, LLMs like ChatGPT:
  - Generate respectful, collaborative outputs when treated politely.
  - Avoid engaging with derogatory or offensive language.
  - **Microsoft Bing AI:** Disapproves of derogatory speech.
  - **ChatGPT:** Advises against offensive language or emphasizes its lack of emotions.

## Human-LLM Interaction

- **Human Politeness:**
  - Influences collaborative output quality.
- **LLM Politeness:**
  - Essential for fostering trust and effective communication

# Can LLMs Predict Politeness?

## **Context #example input**

user: I am looking for help improving the dermatology content on Wikipedia. Would you be willing to help, or do you have any friends interested in der...

## **Prompt #query ChatGPT**

Based on the formal workplace social norms , which of the following best describes the above conversation?

A: Polite

B: Neutral

C: Impolite

Constraint: Answer with only the option above that is most accurate and nothing else.

## **Expected Answer**

A: Polite



# Issues in Computational Politeness Research

- **Lack of Annotated Data**
  - Data scarcity limits the training of robust models.
  - Heavy reliance on manually created datasets.
- **Subjectivity in Politeness**
  - Variability in annotations due to personal and cultural differences.
  - Diverse inter-annotator agreement scores.
- **Data Annotation Challenges**
  - Manual annotation is time-intensive and costly.
  - Quality concerns in large-scale annotations.
- **Limited Multilingual Resources**
  - Majority of resources focus on English.
  - Limited datasets for languages like Hindi and Chinese.

# Conclusion

Computational politeness research has advanced through **four** milestones:

1. Feature-based identification and induction.
2. Contextual incorporation in politeness modeling.
3. Analysis of social factors (e.g., culture, age, gender).
4. Exploration of socio-linguistic cues like emotions and dialogue acts.

# Conclusion

- Datasets and methodologies for politeness annotation are evolving, with supervised techniques being the most prominent.
- Contextual inclusion and deep learning-based approaches are enhancing the study of politeness.
- Issues persist in the form of limited annotated corpora and subjectivity in annotations.

# Future Directions

## 1. Politeness Variation Across Social Factors

- Model politeness variation by age, gender, and culture for adaptive human-like systems.

## 2. Integration with Socio-Linguistic Cues

- Extend analysis of politeness-emotion and politeness-dialogue act relationships using advanced deep learning and LLMs.

## 3. Domain and Language Expansion

- Investigate politeness in low-resource languages (e.g., Hindi, Spanish).
- Explore domain-specific politeness in education, healthcare, and other contexts.

## 4. Deep Learning for Contextual Retrieval

- Utilize architectures to capture conversation-specific and user-specific knowledge.

## 5. Politeness for Social Good Applications

- Mental health counseling: Foster trust and rapport in dialogue systems.
- Persuasion and negotiation: Create empathetic and constructive communication systems.
- New domains: Education, environmental monitoring, and more.
- **Vision:** Build AI systems that are empathetic, engaging, and capable of fostering positive societal impact through effective communication

# Acknowledgement

- “Innovation in Science Pursuit for Inspired Research (INSPIRE) Fellowship” implemented by the Department of Science and Technology, Ministry of Science and Technology, Government of India
- Partial support from the project titled “An Empathetic Knowledge Grounded Conversational System for Mental Health Counseling and Legal Assistance”, Sponsored by IHUB Anubhuti, TIH, IIIT Delhi.
- Special thanks to Medha Aggarwal, B.Tech Student in the Department of Computer Science and Engineering, IIT Patna and Prajwal Kajare, Ph.D student in the School of AI and Data Science, IIT Jodhpur for their exceptional support in putting together this tutorial.



# Relevant publications from research group

- Priyanshu Priya, Mauajama Firdaus, Asif Ekbal. Computational Politeness in Natural Language Processing: A Survey. In ACM Computing Surveys (<https://arxiv.org/pdf/2407.12814>)
- Priyanshu Priya, Gopendra Singh, Mauajama Firdaus, Jyotsna Agrawal, Asif Ekbal (2024). On the Way to Gentle AI Counselor: Politeness Cause Elicitation and Intensity Tagging in Code-mixed Hinglish Conversations for Social Good. In Findings of NAACL 2024.
- Mauajama Firdaus, Priyanshu Priya, Asif Ekbal (2023). Mixing It Up: Inducing Empathy and Politeness using Multiple Behaviour-aware Generators for Conversational Systems. In Findings of AACL-IJCNLP 2023.
- Kshitij Mishra, Priyanshu Priya, Asif Ekbal (2023). PAL to Lend a Helping Hand: Towards Building an Emotion Adaptive Polite and Empathetic Counseling Conversational Agent. In Proceedings of ACL 2023, Toronto, Canada.
- Priyanshu Priya, Kshitij Mishra, Asif Ekbal (2023). PARTNER: A Persuasive Mental Health and Legal Counselling Dialogue System for Women and Children Crime Victims. In IJCAI 2023 (AI for Social Good Track), Macao, S.A.R..
- Kshitij Mishra, Priyanshu Priya, Asif Ekbal (2023). Help Me Heal: A Reinforced Polite and Empathetic Mental Health and Legal Counseling Dialogue System for Crime Victims. In AAAI 2023, USA.
- Z. Ahmad, K. Mishra, A. Ekbal (2023). RPTCS: A Reinforced Persona-aware Topic-guiding Conversational System. In EACL 2023, Croatia.
- K. Mishra, M. Firdaus, A. Ekbal (2023). GenPADS: Reinforcing Politeness in an End-to-End Dialogue System. In PLoS ONE.
- M. Firdaus, A. Ekbal, P. Bhattacharyya (2022). PoliSe: Reinforcing Politeness using User Sentiment for Customer Care Response Generation. In COLING 2022, Korea.
- K. Mishra, A. Samad, P. Totala, A. Ekbal (2022). PEPDS: A Polite and Empathetic Persuasive Dialogue System for Charity Donation. In COLING 2022, Korea.
- A. Samad, K. Mishra, M. Firdaus, A. Ekbal (2022). Empathetic Persuasion: Reinforcing Empathy and Persuasiveness in Dialogue Systems. In Findings of NAACL-HLT 2022, USA.

# Relevant publications from research group

- Kshitij Mishra, Mauajama Firdaus, Asif Ekbal (2022). Please be Polite: Towards Building a Politeness Adaptive Dialogue System for Goal-oriented Conversations. In Neurocomputing, Elsevier.
- K. Mishra, M. Firdaus, A. Ekbal (2022). Predicting Politeness Variations in Goal-oriented Conversations. In IEEE Transactions on Computational Social Systems.
- Mauajama Firdaus, Nidhi Thakur, Asif Ekbal (2022). Sentiment Guided Aspect Conditioned Dialogue Generation in a Multimodal System. In ECIR 2022.
- M. Firdaus, U. Jain, A. Ekbal, P. Bhattacharyya (2021). SEPRG: Sentiment Aware Emotion Controlled Personalized Response Generation. In INLG 2021.
- Mauajama Firdaus, Hitesh Golchha, Asif Ekbal, Pushpak Bhattacharyya (2021). A Deep Multi-task Model for Dialogue Act Classification, Intent Detection and Slot Filling. In Cognitive Computation.
- Mauajama Firdaus, Nidhi Thakur, Asif Ekbal (2021). Multi-Aspect Controlled Response Generation in a Multimodal Dialogue System using Hierarchical Transformer Network. In IJCNN 2021.
- Mauajama Firdaus, Arunav Pratap Shandilya, Sthita Pragyani Pujari, Asif Ekbal (2021). Attribute Centered Multimodal Response Generation in a Dialogue System. In IJCNN 2021.
- M. Firdaus, N. Thakur, A. Ekbal (2021). Aspect-Aware Response Generation for Multimodal Dialogue System. In ACM TIST.
- M. Firdaus, H. Chauhan, A. Ekbal, P. Bhattacharyya (2020). EmoSen: Generating Sentiment and Emotion Controlled Responses in a Multimodal Dialogue System. In IEEE Transactions on Affective Computing.
- Mauajama Firdaus, Nidhi Thakur, Asif Ekbal (2020). MultiDMGCN: Aspect-Guided Response Generation in Multi-Domain Multi-Modal Dialogue System. In Findings of EMNLP 2020.

# Relevant publications from research group

- M. Firdaus, A.P. Shandilya, A. Ekbal (2020). More to Diverse: Generating Diversified Responses in a Task Oriented Multimodal Dialog System. In PLoS ONE.
- Mauajama Firdaus, Asif Ekbal, Pushpak Bhattacharyya (2020). Incorporating Politeness across Languages in Customer Care Responses. In LREC 2020.
- Hardik Chauhan, Mauajama Firdaus, Asif Ekbal, Pushpak Bhattacharyya (2019). Ordinal and Attribute Aware Response Generation in a Multimodal Dialogue System. In ACL 2019.
- M. Firdaus, H. Chauhan, A. Ekbal, P. Bhattacharyya (2020). MEISD: A Multimodal Multi-Label Emotion, Intensity and Sentiment Dialogue Dataset. In COLING 2020.
- M. Firdaus, N. Thangavelu, A. Ekbal, P. Bhattacharyya (2020). Persona Aware Response Generation with Emotions. In IJCNN 2020.
- M. Firdaus, A. Kumar, A. Ekbal, P. Bhattacharyya (2019). A Multi-task Hierarchical Approach for Intent Detection and Slot Filling. In Knowledge-Based Systems, Elsevier.
- Hitesh Golchha, Mauajama Firdaus, Asif Ekbal, Pushpak Bhattacharyya (2019). Courteously Yours: Inducing Courteous Behavior in Customer Care Responses. In NAACL 2019.



# References

- [1] Mikhail Alexandrov, Xavier Blanco, Natalia Ponomareva, and Paolo Rosso. 2007. Constructing empirical models for automatic dialog parameterization. In *International Conference on Text, Speech and Dialogue*. Springer, 455–463.
- [2] Mikhail Alexandrov, Natalia Ponomareva, and Xavier Blanco. 2008. Regression Model for Politeness Estimation Trained on Examples. In *Proceedings of the NooJ'07 conference*. 206–13.
- [3] Ahmad Aljanaideh, Eric Fosler-Lussier, and Marie-Catherine de Marneffe. 2020. Contextualized Embeddings for Enriching Linguistic Analyses on Politeness. In *Proceedings of the 28th International Conference on Computational Linguistics*. 2181–2190.
- [4] Malika Aubakirova and Mohit Bansal. 2016. Interpreting Neural Networks to Improve Politeness Comprehension. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 2035–2041.
- [5] Dzmitry Bahdanau, Kyung Hyun Cho, and Yoshua Bengio. 2015. Neural machine translation by jointly learning to align and translate. In *3rd International Conference on Learning Representations, ICLR 2015*.
- [6] Jiajun Bao, Junjie Wu, Yiming Zhang, Eshwar Chandrasekharan, and David Jurgens. 2021. Conversations gone alright: Quantifying and predicting prosocial outcomes in online conversations. In *Proceedings of the Web Conference 2021*. 1134–1145.
- [7] Ricardo Barandela, José Salvador Sánchez, Vicente Garcia, and Edgar Rangel. 2003. Strategies for learning in class imbalance problems. *Pattern Recognition* 36, 3 (2003), 849–851.
- [8] Elizabeth Bates and Louise Silvern. 1977. Social Adjustment and Politeness in Preschoolers. *Journal of Communication* 27, 2 (1977), 104–111.
- [9] Prabhat Kumar Bharti, Meith Navlakha, Mayank Agarwal, and Asif Ekbal. 2023. PolitePEER: does peer review hurt? A dataset to gauge politeness intensity in the peer reviews. *Language Resources and Evaluation* (2023), 1–23.
- [10] Douglas Biber. 1991. *Variation across speech and writing*. Cambridge University Press.
- [11] Avrim L Blum and Merrick L Furst. 1997. Fast planning through planning graph analysis. *Artificial intelligence* 90, 1-2 (1997), 281–300.

# References

- [12] Chandrakant Bothe and Stefan Wermter. 2022. Conversational Analysis of Daily Dialog Data using Polite Emotional Dialogue Acts. In Proceedings of the Thirteenth Language Resources and Evaluation Conference. 2395–2400.
- [13] Derek Bousfield. 2008. Impoliteness in interaction. *Impoliteness in interaction* (2008), 1–295.
- [14] Jonathon D Brown and S Smart. 1991. The self and social conduct: Linking self-representations to prosocial behavior. *Journal of Personality and Social psychology* 60, 3 (1991), 368.
- [15] Penelope Brown and Colin Fraser. 1979. Speech as a marker of situation. In *Social markers in speech*. Cambridge University Press, 33–62.
- [16] Penelope Brown and Stephen C Levinson. 1978. Universals in language usage: Politeness phenomena. In *Questions and politeness: Strategies in social interaction*. Cambridge University Press, 56–311.
- [17] Penelope Brown, Stephen C Levinson, and Stephen C Levinson. 1987. *Politeness: Some universals in language usage*. Vol. 4. Cambridge university press.
- [18] Peter F Brown, Stephen A Della Pietra, Vincent J Della Pietra, Jennifer C Lai, and Robert L Mercer. 1992. An estimate of an upper bound for the entropy of English. *Computational Linguistics* 18, 1 (1992), 31–40.
- [19] Tom Brown, Benjamin Mann, Nick Ryder, Melanie Subbiah, Jared D Kaplan, Prafulla Dhariwal, Arvind Neelakantan, Pranav Shyam, Girish Sastry, Amanda Askell, et al. 2020. Language models are few-shot learners. *Advances in neural information processing systems* 33 (2020), 1877–1901.
- [20] Sven Buechel, Anneke Buffone, Barry Slaff, Lyle Ungar, and João Sedoc. 2018. Modeling Empathy and Distress in Reaction to News Stories. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*. 4758–4765.
- [21] Bill Byrne, Karthik Krishnamoorthi, Chinnadhurai Sankar, Arvind Neelakantan, Ben Goodrich, Daniel Duckworth, Semih Yavuz, Amit Dubey, Kyu-Young Kim, and Andy Cedilnik. 2019. Taskmaster-1: Toward a Realistic and Diverse Dialog Dataset. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 4516–4525.

# References

- [22] Laura L Carstensen, Bulent Turan, Susanne Scheibe, Nilam Ram, Hal Ersner-Hershfield, Gregory R Samanez-Larkin, Kathryn P Brooks, and John R Nesselrode. 2011. Emotional experience improves with age: evidence based on over 10 years of experience sampling. *Psychology and aging* 26, 1 (2011), 21.
- [23] Guimin Chen, Yuanhe Tian, and Yan Song. 2020. Joint aspect extraction and sentiment analysis with directional graph convolutional networks. In *Proceedings of the 28th international conference on computational linguistics*. 272–279.
- [24] Niyati Chhaya, Kushal Chawla, Tanya Goyal, Projjal Chanda, and Jaya Singh. 2018. Frustrated, polite, or formal: Quantifying feelings and tone in email. In *Proceedings of the Second Workshop on Computational Modeling of People’s Opinions, Personality, and Emotions in Social Media*. 76–86.
- [25] Christine Christie. 2007. Relevance theory and politeness. *Journal of Politeness Research* 3, 2 (2007), 269–294. <https://doi.org/doi:10.1515/PR.2007.012>
- [26] Hyung Won Chung, Le Hou, Shayne Longpre, Barret Zoph, Yi Tay, William Fedus, Yunxuan Li, Xuezhi Wang, Mostafa Dehghani, Siddhartha Brahma, Albert Webson, Shixiang Shane Gu, Zhuyun Dai, Mirac Suzgun, Xinyun Chen, Aakanksha Chowdhery, Alex Castro-Ros, Marie Pellat, Kevin Robinson, Dasha Valter, Sharan Narang, Gaurav Mishra, Adams Yu, Vincent Zhao, Yanping Huang, Andrew Dai, Hongkun Yu, Slav Petrov, Ed H. Chi, Jeff Dean, Jacob Devlin, Adam Roberts, Denny Zhou, Quoc V. Le, and Jason Wei. 2024. Scaling Instruction-Finetuned Language Models. *Journal of Machine Learning Research* 25, 70 (2024), 1–53. <http://jmlr.org/papers/v25/23-0870.html>
- [27] Yi-Ling Chung, Elizaveta Kuzmenko, Serra Sinem Tekirođlu, and Marco Guerini. 2019. CONAN-COunter NARratives through Nichesourcing: a Multilingual Dataset of Responses to Fight Online Hate Speech. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 2819–2829.
- [28] Jonathan Culpeper. 2011. 13. Politeness and impoliteness. *Pragmatics of society* 5 (2011), 393.
- [29] Jonathan Culpeper. 2011. *Impoliteness: Using language to cause offence*. Vol. 28. Cambridge University Press.
- [30] Jonathan Culpeper. 2016. *Impoliteness strategies*. *Interdisciplinary studies in pragmatics, culture and society* (2016), 421–445.

# References

- [31] Jonathan Culpeper, Michael Haugh, and Dániel Z Kádár. 2017. The Palgrave handbook of linguistic (im) politeness. Springer.
- [32] Marianne Dainton, Laura Stafford, and Daniel J Canary. 1994. Maintenance strategies and physical affection as predictors of love, liking, and satisfaction in marriage. *Communication Reports* 7, 2 (1994), 88–98.
- [33] Cristian Danescu-Niculescu-Mizil, Moritz Sudhof, Dan Jurafsky, Jure Leskovec, and Christopher Potts. 2013. A computational approach to politeness with application to social factors. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 250–259.
- [34] Tirthankar Dasgupta, Manjira Sinha, and Chundru Geetha Praveen. 2023. Graph Induced Transformer Network for Detection of Politeness and Formality in Text. In *Companion Proceedings of the ACM Web Conference 2023*. 221–224.
- [35] Jacob Devlin, Ming-Wei Chang, Kenton Lee, and Kristina Toutanova. 2018. Bert: Pre-training of deep bidirectional transformers for language understanding. *arXiv preprint arXiv:1810.04805* (2018).
- [36] Doris Dippold, Jenny Lynden, Rob Shrubsall, and Rich Ingram. 2020. A turn to language: How interactional sociolinguistics informs the redesign of prompt: response chatbot turns. *Discourse, Context & Media* 37 (2020), 100432.
- [37] Chunling Du, Haifeng Sun, Jingyu Wang, Qi Qi, Jianxin Liao, Chun Wang, and Bing Ma. 2019. Investigating capsule network and semantic feature on hyperplanes for text classification. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*. 456–465.
- [38] Weston Feely, Eva Hasler, and Adrià de Gispert. 2019. Controlling Japanese honorifics in English-to-Japanese neural machine translation. In *Proceedings of the 6th Workshop on Asian Translation*. 45–53.
- [39] Shutong Feng, Nurul Lubis, Christian Geishauer, Hsien-Chin Lin, Michael Heck, Carel van Niekerk, and Milica Gasic. 2022. EmoWOZ: A Large-Scale Corpus and Labelling Scheme for Emotion Recognition in Task-Oriented Dialogue Systems. In *Proceedings of the Thirteenth Language Resources and Evaluation Conference*. 4096–4113.

# References

- [40] Mauajama Firdaus, Asif Ekbal, and Pushpak Bhattacharyya. 2020. Incorporating Politeness across Languages in Customer Care Responses: Towards building a Multi-lingual Empathetic Dialogue Agent. In Proceedings of the 12th Language Resources and Evaluation Conference. 4172–4182.
- [41] Mauajama Firdaus, Asif Ekbal, and Pushpak Bhattacharyya. 2022. PoliSe: Reinforcing Politeness using User Sentiment for Customer Care Response Generation. In Proceedings of the 29th International Conference on Computational Linguistics. 6165–6175.
- [42] Mauajama Firdaus, Priyanshu Priya, and Asif Ekbal. 2023. Mixing It Up: Inducing Empathy and Politeness using Multiple Behaviour-aware Generators for Conversational Systems. In Findings of the Association for Computational Linguistics: IJCNLP-AAACL 2023 (Findings). 336–347.
- [43] Mauajama Firdaus, Arunav Shandilya, Asif Ekbal, and Pushpak Bhattacharyya. 2022. Being Polite: Modeling Politeness Variation in a Personalized Dialog Agent. IEEE Transactions on Computational Social Systems (2022).
- [44] John L Fischer. 1965. The stylistic significance of consonantal sandhi in Trukese and Ponapean. *American Anthropologist* 67, 6 (1965), 1495–1502.
- [45] Liye Fu, Susan Fussell, and Cristian Danescu-Niculescu-Mizil. 2020. Facilitating the Communication of Politeness through Fine-Grained Paraphrasing. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 5127–5140.
- [46] Vicente García, José Salvador Sánchez, and Ramón Alberto Mollineda. 2012. On the effectiveness of preprocessing methods when dealing with different levels of class imbalance. *Knowledge-Based Systems* 25, 1 (2012), 13–21.
- [47] Eelen Gino. 2001. *A critique of Politeness Theories*. Manchester: St.
- [48] Erving Goffman. 1967. *Interaction ritual: Essays on face-to-face interaction* (1st ed.). Routledge, New York, NY, USA.
- [49] Hitesh Golchha, Mauajama Firdaus, Asif Ekbal, and Pushpak Bhattacharyya. 2019. Courteously Yours: Inducing courteous behavior in Customer Care responses using Reinforced Pointer Generator Network. In Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers). 851–860.
- [50] Sage Lambert Graham. 2007. Disagreeing to agree: Conflict,(im) politeness and identity in a computer-mediated community. *Journal of pragmatics* 39, 4 (2007), 742–759

# References

- [51] H Paul Grice. 1975. Logic and Conversation, Syntax and Semantics, vol. 3 edited by P. Cole and J. Morgan.
- [52] Yueguo Gu. 1990. Politeness phenomena in modern Chinese. *Journal of pragmatics* 14, 2 (1990), 237–257.
- [53] Swati Gupta, Marilyn A Walker, and Daniela M Romano. 2007. How rude are you?: Evaluating politeness and affect in interaction. In *International Conference on Affective Computing and Intelligent Interaction*. Springer, 203–217.
- [54] Michael Haugh. 2010. When is an email really offensive?: Argumentativity and variability in evaluations of impoliteness. (2010).
- [55] Francis Heylighen and Jean-Marc Dewaele. 1999. Formality of language: definition, measurement and behavioral determinants. Technical Report. Center "Leo Apostel", Free University of Brussels, Pleinlaan 2, B-1050 Brussels, Belgium.
- [56] Erin R Hoffman, David W McDonald, and Mark Zachry. 2017. Evaluating a computational approach to labeling politeness: Challenges for the application of machine classification to social computing data. *Proceedings of the ACM on Human-computer Interaction* 1, CSCW (2017), 1–14.
- [57] Janet Holmes. 1988. Paying compliments: A sex-preferential politeness strategy. *Journal of pragmatics* 12, 4 (1988), 445–465.
- [58] Janet Holmes. 2013. *Women, men and politeness*. Routledge.
- [59] Thomas Holtgraves and Yang Joong-Nam. 1990. Politeness as universal: Cross-cultural perceptions of request strategies and inferences based on their use. *Journal of personality and social psychology* 59, 4 (1990), 719.
- [60] Eduard Hovy. 1987. Generating natural language under pragmatic constraints. *Journal of Pragmatics* 11, 6 (1987), 689–719.
- [61] Richard Hudson. 1994. About 37% of word-tokens are nouns. *Language* 70, 2 (1994), 331–339.
- [62] Clayton Hutto and Eric Gilbert. 2014. Vader: A parsimonious rule-based model for sentiment analysis of social media text. In *Proceedings of the international AAAI conference on web and social media*, Vol. 8. 216–225.
- [63] Ohad Inbar and Joachim Meyer. 2019. Politeness counts: Perceptions of peacekeeping robots. *IEEE Transactions on Human-Machine Systems* 49, 3 (2019), 232–240.
- [64] Judith T Irvine. 1979. Formality and informality in communicative events. *American anthropologist* 81, 4 (1979), 773–790.

# References

- [65] Richard W Janney and Horst Arndt. 1993. Universality and relativity in cross-cultural politeness research: a historical perspective. *Multilingua* 12, 1 (1993), 13–50.
- [66] Martha Jeong, Julia Minson, Michael Yeomans, and Francesca Gino. 2019. Communicating with warmth in distributive negotiations is surprisingly counterproductive. *Management Science* 65, 12 (2019), 5813–5837.
- [67] W Lewis Johnson and Paola Rizzo. 2004. Politeness in tutoring dialogs: “run the factory, that’s what I’d do”. In *International Conference on Intelligent Tutoring Systems*. Springer, 67–76.
- [68] W Lewis Johnson, Paola Rizzo, Wauter Bosma, Sander Kole, Mattijs Ghijsen, and Herwin van Welbergen. 2004. Generating socially appropriate tutorial dialog. In *Tutorial and Research Workshop on Affective Dialogue Systems*. Springer, 254–264.
- [69] Chaitanya K Joshi, Fei Mi, and Boi Faltings. 2017. Personalization in goal-oriented dialog. *arXiv preprint arXiv:1706.07503* (2017).
- [70] Manfred Kienpointner. 2008. Impoliteness and emotional arguments. (2008).
- [71] Thomas N. Kipf and Max Welling. 2017. Semi-Supervised Classification with Graph Convolutional Networks. In *Proceedings of the 5th International Conference on Learning Representations (Palais des Congrès Neptune, Toulon, France) (ICLR ’17)*. <https://openreview.net/forum?id=SJU4ayYgl>
- [72] Bryan Klimt and Yiming Yang. 2004. Introducing the Enron corpus.. In *CEAS*.
- [73] Klaus Krippendorff. 2007. Computing Krippendorff’s Alpha-Reliability. <http://www.asc.upenn.edu/Krippendorff/> (2007).
- [74] Ritesh Kumar. 2012. Challenges in the development of annotated corpora of computer-mediated communication in Indian languages: A case of Hindi. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC’12)*. 299–302.
- [75] Ritesh Kumar. 2014. Developing Politeness Annotated Corpus of Hindi Blogs.. In *LREC*. 1275–1280.
- [76] Ritesh Kumar. 2021. Towards automatic identification of linguistic politeness in Hindi texts. *arXiv preprint arXiv:2111.15268* (2021).
- [77] Ritesh Kumar and Girish Nath Jha. 2010. Translating politeness across cultures: Case of Hindi and English. In *Proceedings of the 3rd international conference on Intercultural collaboration*. 175–178.

# References

- [78] Aylin C Küntay, Keiko Nakamura, and B Ateş Şen. 2014. Crosslinguistic and crosscultural approaches to pragmatic development. *Pragmatic development in first language acquisition* (2014), 317–342.
- [79] Robin Lakoff. 1973. Language and woman's place. *Language in society* 2, 1 (1973), 45–79.
- [80] Robin Lakoff. 1973. The logic of politeness: Or, minding your p's and q's. In *Proceedings from the Annual Meeting of the Chicago Linguistic Society*, Vol. 9. Chicago Linguistic Society, 292–305.
- [81] Robin Tolmach Lakoff. 1989. The Limits of Politeness: Therapeutic and Courtroom Discourse in *Linguistic Politeness II*. *Multilingua* 8, 2-3 (1989), 101–129.
- [82] Andreas Langlotz and Miriam A Locher. 2017. (Im) politeness and emotion. In *The Palgrave Handbook of Linguistic (Im) politeness*. Springer, 287–322.
- [83] GN Leech. 1983. *Principles of Pragmatics* London: Longman Group Ltd. (1983).
- [84] Geoffrey Leech. 1992. Pragmatic principles in Shaw's you never can tell. *Language, text and context: essays in stylistics* (1992), 259–78.
- [85] Geoffrey Leech. 2005. Politeness: is there an East-West divide. *Journal of foreign languages* 6, 3 (2005), 1–30.
- [86] Mike Lewis, Yinhan Liu, Naman Goyal, Marjan Ghazvininejad, Abdelrahman Mohamed, Omer Levy, Veselin Stoyanov, and Luke Zettlemoyer. 2020. BART: Denoising Sequence-to-Sequence Pre-training for Natural Language Generation, Translation, and Comprehension. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*. 7871–7880.
- [87] Can Li, Bin Pang, Wenbo Wang, Lingshu Hu, Matthew Gordon, Detelina Marinova, Bitty Balducci, and Yi Shang. 2023. How Well Can Language Models Understand Politeness?. In *2023 IEEE Conference on Artificial Intelligence (CAI)*. IEEE, 230–231.
- [88] Mingyang Li, Louis Hickman, Louis Tay, Lyle Ungar, and Sharath Chandra Guntuku. 2020. Studying Politeness across Cultures Using English Twitter and Mandarin Weibo. *Proceedings of the ACM on Human-Computer Interaction* 4, CSCW2 (2020), 1–15.
- [89] Xiujun Li, Yu Wang, Siqi Sun, Sarah Panda, Jingjing Liu, and Jianfeng Gao. 2018. Microsoft dialogue challenge: Building end-to-end task-completion dialogue systems. *arXiv preprint arXiv:1807.11125* (2018).



# References

- [90] Yanran Li, Hui Su, Xiaoyu Shen, Wenjie Li, Ziqiang Cao, and Shuzi Niu. 2017. DailyDialog: A Manually Labelled Multi-turn Dialogue Dataset. In Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 1: Long Papers). 986–995.
- [91] Valentin Liévin, Christoffer Egeberg Hother, and Ole Winther. 2022. Can large language models reason about medical questions? arXiv preprint arXiv:2207.08143 (2022).
- [92] Chin-Yew Lin. 2004. Rouge: A package for automatic evaluation of summaries. In Text summarization branches out. 74–81.
- [93] Yen-Ting Lin and Yun-Nung Chen. 2023. LLM-Eval: Unified Multi-Dimensional Automatic Evaluation for Open-Domain Conversations with Large Language Models. In Proceedings of the 5th Workshop on NLP for Conversational AI (NLP4ConvAI 2023), Yun-Nung Chen and Abhinav Rastogi (Eds.). Association for Computational Linguistics, Toronto, Canada, 47–58. <https://doi.org/10.18653/v1/2023.nlp4convai-1.5>
- [94] Miriam A Locher. 2006. Polite behavior within relational work: The discursive approach to politeness. *Multilingua* 25, 3 (2006), 249–267.
- [95] Miriam A Locher and DJ Glick. 2006. Power and politeness in action: Disagreements in oral communication. *LANGUAGE IN SOCIETY-LONDON* 35, 5 (2006), 729–732.
- [96] Miriam A Locher and Richard J Watts. 2005. Politeness theory and relational work. (2005).
- [97] Miriam A Locher and Richard J Watts. 2008. Relational work and impoliteness: Negotiating norms of linguistic behaviour. *Language Power and Social Process* 21 (2008), 77.
- [98] Aman Madaan, Amrith Setlur, Tanmay Parekh, Barnabás Póczós, Graham Neubig, Yiming Yang, Ruslan Salakhutdinov, Alan W Black, and Shrimai Prabhumoye. 2020. Politeness Transfer: A Tag and Generate Approach. In Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics. 1869–1881.
- [99] Michael P Maratsos. 1973. Nonegocentric communication abilities in preschool children. *Child development* (1973), 697–700.
- [100] Bonnie McElhinny. 1996. Language and Gender. *Sociolinguistics and Language Teaching* (1996), 218. [101]
- Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica* 22, 3 (2012), 276–282. [102]
- Christopher Miller, Peggy Wu, H Funk, Lewis Johnson, and H Viljalmsson. 2007. A computational approach to etiquette and politeness: An “Etiquette Engine™” for cultural interaction training. Proceedings of BRIMS 2007 (2007).

# References

- [103] Christopher Miller, Peggy Wu, H Funk, Peggy Wilson, and Lewis Johnson. 2006. A computational approach to etiquette and politeness: Initial test cases. In proceedings of 2006 BRIMS Conference. 15–18. [104]
- Christopher A Miller, Peggy Wu, and Harry B Funk. 2008. A computational approach to etiquette: Operationalizing Brown and Levinson’s politeness model. *IEEE Intelligent Systems* 23, 4 (2008), 28–35. [105]
- Sara Mills. 2003. *Gender and politeness*. Number 17. Cambridge University Press. [106]
- Julia A Minson, Frances S Chen, and Catherine H Tinsley. 2020. Why won’t you listen to me? Measuring receptiveness to opposing views. *Management Science* 66, 7 (2020), 3069–3094. [107]
- Kshitij Mishra, Mauajama Firdaus, and Asif Ekbal. 2022. Please be polite: Towards building a politeness adaptive dialogue system for goal-oriented conversations. *Neurocomputing* 494 (2022), 242–254. [108]
- Kshitij Mishra, Mauajama Firdaus, and Asif Ekbal. 2022. Predicting Politeness Variations in Goal-Oriented Conversations. *IEEE Transactions on Computational Social Systems* (2022). [109] Kshitij Mishra, Mauajama Firdaus, and Asif Ekbal. 2023. GenPADS: Reinforcing politeness in an end-to-end dialogue system. *Plos one* 18, 1 (2023), e0278323. [110] Kshitij Mishra, Priyanshu Priya, Manisha Burja, and Asif Ekbal. 2023. e-THERAPIST: I suggest you to cultivate a mindset of positivity and nurture uplifting thoughts. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*. 13952–13967. [111] Kshitij Mishra, Priyanshu Priya, and Asif Ekbal. 2023. Help Me Heal: A Reinforced Polite and Empathetic Mental Health and Legal Counseling Dialogue System for Crime Victims. In *Proceedings of the AAAI Conference on Artificial Intelligence*, Vol. 37. 14408–14416. [112] Kshitij Mishra, Priyanshu Priya, and Asif Ekbal. 2023. PAL to Lend a Helping Hand: Towards Building an Emotion Adaptive Polite and Empathetic Counseling Conversational Agent. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 12254–12271. [113] Kshitij Mishra, Azlaan Mustafa Samad, Palak Totala, and Asif Ekbal. 2022. PEPDS: A polite and empathetic persuasive dialogue system for charity donation. In *Proceedings of the 29th International Conference on Computational Linguistics*. 424–440

# References

- [114] Tomoki Miyamoto, Daisuke Katagami, and Mayumi Usami. 2020. A politeness control method for conversational agents considering social relationships with users. In Annual Conference of the Japanese Society for Artificial Intelligence. Springer, 224–231. [115]
- Saif M Mohammad and Peter D Turney. 2013. Crowdsourcing a word–emotion association lexicon. *Computational intelligence* 29, 3 (2013), 436–465. [116]
- Michael B Montgomery. 1998. Multiple Modals in LAGS. *From the gulf states and beyond: The legacy of Lee Pederson and LAGS* (1998), 90. [117]
- Sourabrata Mukherjee, Vojtěch Hudeček, and Ondřej Dušek. 2023. Polite Chatbot: A Text Style Transfer Application. In Proceedings of the 17th Conference of the European Chapter of the Association for Computational Linguistics: Student Research Workshop. 87–93. [118]
- Dasa Munkova, Michal Munk, and Zuzana Fráterová. 2013. Identifying social and expressive factors in request texts using transaction/sequence model. In Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013. 496–503. [119]
- Keiko Nakamura, Y Shirai, S Kobayashi, and S Miyata. 2002. Polite language usage in mother-infant interactions: A look at language socialization. *Studies in language sciences* 2 (2002), 175–191. [120]
- Courtney Napoles, Aasish Pappu, and Joel Tetreault. 2017. Automatically identifying good conversations online (yes, they do exist!). In Proceedings of the International AAAI Conference on Web and Social Media, Vol. 11. 628–631. [121]
- Courtney Napoles, Joel Tetreault, Aasish Pappu, Enrica Rosato, and Brian Provenza. 2017. Finding good conversations online: The yahoo news annotated comments corpus. In Proceedings of the 11th linguistic annotation workshop. 13–23. [122]
- Tong Niu and Mohit Bansal. 2018. Polite dialogue generation without parallel data. *Transactions of the Association for Computational Linguistics* 6 (2018), 373–389. [123]
- Andrew Ortony, Gerald L Clore, and Allan Collins. 1988. *The cognitive structure of emotions* Cambridge. UK: Cambridge University Press9 (1988).

# References

- [124] Kishore Papineni, Salim Roukos, Todd Ward, and Wei-Jing Zhu. 2002. Bleu: a method for automatic evaluation of machine translation. In Proceedings of the 40th annual meeting of the Association for Computational Linguistics. 311–318. [125]
- James W Pennebaker, Martha E Francis, and Roger J Booth. 2001. Linguistic inquiry and word count: LIWC 2001. Mahway: Lawrence Erlbaum Associates 71, 2001 (2001), 2001. [126]
- Verónica Pérez-Rosas, Xinyi Wu, Kenneth Resnicow, and Rada Mihalcea. 2019. What makes a good counselor? learning to distinguish between high-quality and low-quality counseling conversations. In Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics. 926–935. [127]
- Denis Peskov, Nancy Clarke, Jason Krone, Brigi Fodor, Yi Zhang, Adel Youssef, and Mona Diab. 2019. Multi-domain goal-oriented dialogues (multidogo): Strategies toward curating and annotating large scale dialogue data. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 4526–4536. [128]
- C Pfuderer. 1968. A scale of politeness of request forms in English. Term paper for Speech 164A. University of California, Berkeley. Cited by Ervin-Tripp, S.(1973). Language Acquisition and Communicative Choice, Stanford University Press, Stanford, Calif (1968). [129]
- Priyanshu Priya, Mauajama Firdaus, and Asif Ekbal. 2023. A multi-task learning framework for politeness and emotion detection in dialogues for mental health counselling and legal aid. Expert Systems with Applications 224 (2023), 120025. [130]
- Priyanshu Priya, Kshitij Mishra, Palak Totala, and Asif Ekbal. 2023. PARTNER: A Persuasive Mental Health and Legal Counselling Dialogue System for Women and Children Crime Victims. In Proceedings of the Thirty-Second International Joint Conference on Artificial Intelligence, IJCAI-23, Edith Elkind (Ed.). International Joint Conferences on Artificial Intelligence Organization, 6183–6191. <https://doi.org/10.24963/ijcai.2023/686> AI for Good.
- [131] Dongqi Pu and Vera Demberg. 2023. ChatGPT vs Human-authored Text: Insights into Controllable Text Summarization and Sentence Style Transfer. In Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 4: Student Research Workshop), Vishakh Padmakumar, Gisela Vallejo, and Yao Fu (Eds.). Association for Computational Linguistics, Toronto, Canada, 1–18. <https://doi.org/10.18653/v1/2023.acl-srw.1>

# References

- [132] Jing Qian, Anna Bethke, Yinyin Liu, Elizabeth Belding, and William Yang Wang. 2019. A Benchmark Dataset for Learning to Intervene in Online Hate Speech. In Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP). 4755–4764. [133]
- Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. OpenAI blog 1, 8 (2019), 9. [134]
- Rajkumar Ramamurthy, Prithviraj Ammanabrolu, Kianté Brantley, Jack Hessel, Rafet Sifa, Christian Bauckhage, Hannaneh Hajishirzi, and Yejin Choi. 2022. Is Reinforcement Learning (Not) for Natural Language Processing?: Benchmarks, Baselines, and Building Blocks for Natural Language Policy Optimization. arXiv preprint arXiv:2210.01241. <https://arxiv.org/abs/2210.01241> [135]
- Kanishk Rana, Rahul Madaan, and Jainendra Shukla. 2021. Effect of Polite Triggers in Chatbot Conversations on User Experience across Gender, Age, and Personality. In 2021 30th IEEE International Conference on Robot & Human Interactive Communication (RO-MAN). IEEE, 813–819. [136]
- Bernd Renner. 2020. (In) directness as an (im) politeness strategy in the contact between german and Brazilian portuguese as additional languages. Ph. D. Dissertation. Universidade de Brasília, Brasília. [137]
- Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, et al. 2021. Recipes for Building an Open-Domain Chatbot. In Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume. 300–325. [138]
- Norton Trevisan Roman, Paul Piwek, and AMBR Carvalho. 2004. Politeness and summarization: an exploratory study. In Proceedings of AAAI Spring Symposium on Exploring Attitude and Affect in Text: Theories and Applications (AAAI-EAAT 2004). [139]
- Norton Trevisan Roman, Paul Piwek, and Ariadne Maria Brito Rizzoni Carvalho. 2006. Politeness and bias in dialogue summarization: Two exploratory studies. In Computing attitude and affect in text: theory and applications. Springer, 171–185.

# References

- [140] Punyajoy Saha, Kanishk Singh, Adarsh Kumar, Binny Mathew, and Animesh Mukherjee. 2022. CounterGeDi: A Controllable Approach to Generate Polite, Detoxified and Emotional Counterspeech. In Proceedings of the Thirty-First International Joint Conference on Artificial Intelligence, IJCAI-22, Lud De Raedt (Ed.). International Joint Conferences on Artificial Intelligence Organization, 5157–5163. <https://doi.org/10.24963/ijcai.2022/716> AI for Good. [141]
- Victor Sanh, Lysandre Debut, Julien Chaumond, and Thomas Wolf. 2019. DistilBERT, a distilled version of BERT: smaller, faster, cheaper and lighter. In NeurIPS EMC2 Workshop. [142] John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. arXiv preprint arXiv:1707.06347 (2017). [143] John R Searle and John Rogers Searle. 1969. Speech acts: An essay in the philosophy of language. Vol. 626. Cambridge university press. [144] Rico Sennrich, Barry Haddow, and Alexandra Birch. 2016. Controlling politeness in neural machine translation via side constraints. In Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies. 35–40. [145] Rico Sennrich, Martin Volk, and Gerold Schneider. 2013. Exploiting synergies between open resources for german dependency parsing, pos-tagging, and morphological analysis. In Proceedings of the International Conference Recent Advances in Natural Language Processing RANLP 2013. 601–609. [146] Ashish Sharma, Adam Miner, David Atkins, and Tim Althoff. 2020. A Computational Approach to Understanding Empathy Expressed in Text-Based Mental Health Support. In Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP). 5263–5276. [147]
- Marilyn Shatz and Rochel Gelman. 1973. The development of communication skills: Modifications in the speech of young children as a function of listener. Monographs of the society for research in child development (1973), 1–38. [148]
- Diogo Silva, David Semedo, and João Magalhães. 2022. Polite Task-oriented Dialog Agents: To Generate or to Rewrite?. In Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis. 304–314.

# References

- [149] Gopendra Vikram Singh, Priyanshu Priya, Mauajama Firdaus, Asif Ekbal, and Pushpak Bhattacharyya. 2022. EmoInHindi: A Multi-label Emotion and Intensity Annotated Dataset in Hindi for Emotion Recognition in Dialogues. In Proceedings of the Thirteenth Language Resources and Evaluation Conference. 5829–5837. [150]
- Mengmeng Song, Huixian Zhang, Xinyu Xing, and Yucong Duan. 2023. Appreciation vs. apology: Research on the influence mechanism of chatbot service recovery based on politeness theory. *Journal of Retailing and Consumer Services* 73 (2023), 103323. [151]
- Darcy Sperlich, Jaiho Leem, and Eui-Jeen Ahn. 2016. The interaction of politeness systems in Korean learners of French. In Proceedings of the 30th Pacific Asia Conference on Language, Information and Computation: Oral Papers. 163–171. [152]
- Anirudh Srinivasan and Eunsol Choi. 2022. TyDiP: A Dataset for Politeness Classification in Nine Typologically Diverse Languages. In Findings of the Association for Computational Linguistics: EMNLP 2022. 5723–5738. [153] David
- Sweetman and Fred Luthans. 2010. The power of positive psychology: Psychological capital and work engagement. *Work engagement: A handbook of essential theory and research* 54 (2010), 68. [154] Mariana
- Terkourafi. 2001. Politeness in Cypriot Greek: A frame-based approach. Ph. D. Dissertation. Citeseer. [155] Marina
- Terkourafi. 2002. Politeness and formulaicity: evidence from Cypriot Greek. *Journal of Greek Linguistics* 3, 1 (2002), 179–201. [156] Marina
- Terkourafi. 2005. Beyond the micro-level in politeness research. *Journal of Politeness Research* 1 (2005), 237262. [157] Marina
- Terkourafi. 2005. Pragmatic correlates of frequency of use: The case for a notion of “minimal context”. *Trends in Linguistics Studies and Monographs* 161 (2005), 209. [158] Jenny A
- Thomas. 2014. *Meaning in interaction: An introduction to pragmatics*. Routledge. [159] Jörg
- Tiedemann. 2012. Parallel data, tools and interfaces in OPUS. In *Lrec, Vol. 2012*. Citeseer, 2214–2218. [160] Hugo
- Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, et al. 2023. Llama: Open and efficient foundation language models. *arXiv preprint arXiv:2302.13971* (2023). [161] Ashish
- Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N Gomez, Łukasz Kaiser, and Illia Polosukhin. 2017. Attention is all you need. In *Advances in neural information processing systems*. 5998–6008

# References

- [162] Aditi Viswanathan, Varden Wang, and Antonina Kononova. 2019. Controlling Formality and Style of Machine Translation Output Using AutoML. In Annual International Symposium on Information Management and Big Data. Springer, 306–313. [163]
- Carl Vogel. 2015. Some puzzles of politeness and impoliteness within a formal semantics of offensive language. *Conflict and Multimodal Communication: Social Research and Machine Intelligence* (2015), 223–241. [164]
- Rob Voigt, Nicholas P Camp, Vinodkumar Prabhakaran, William L Hamilton, Rebecca C Hetey, Camilla M Griffiths, David Jurgens, Dan Jurafsky, and Jennifer L Eberhardt. 2017. Language from police body camera footage shows racial disparities in officer respect. *Proceedings of the National Academy of Sciences* 114, 25 (2017), 6521–6526. [165]
- Marilyn A Walker, Janet E Cahn, and Stephen J Whittaker. 1997. Improvising linguistic style: Social and affective bases for agent personality. In *Proceedings of the first international conference on Autonomous agents*. 96–105. [166] Xuewei
- Wang, Weiyan Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for Good: Towards a Personalized Persuasive Dialogue System for Social Good. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*. 5635–5649. [167] Yi-Chia
- Wang, Alexandros Papangelis, Runze Wang, Zhaleh Feizollahi, Gokhan Tur, and Robert Kraut. 2020. Can You be More Social? Injecting Politeness and Positivity into Task-Oriented Conversational Agents. *arXiv preprint arXiv:2012.14653* (2020). [168] Ronald
- Wardhaugh. 2006. *An Introduction to Sociolinguistics*. [169] Zeerak
- Waseem, Thomas Davidson, Dana Warmsley, and Ingmar Weber. 2017. Understanding Abuse: A Typology of Abusive Language Detection Subtasks. In *Proceedings of the First Workshop on Abusive Language Online*. 78–84. [170] Richard
- J Watts. 2003. *Politeness*. Cambridge University Press. [171] Richard J
- Watts. 2005. 2. Linguistic politeness and politic verbal behaviour: Reconsidering claims for universality. *Politeness in Language Studies in its History, Theory and Practice* (2005), 43–69. [172] World
- Health Organization WHO. 2022. Mental disorders. <https://www.who.int/news-room/fact-sheets/detail/mental-disorders>



# References

- [173] Jason D Williams, Matthew Henderson, Antoine Raux, Blaise Thomson, Alan Black, and Deepak Ramachandran. 2014. The dialog state tracking challenge series. *AI Magazine* 35, 4 (2014), 121–124. [174]
- Michael Yeomans, Alejandro Kantor, and Dustin Tingley. 2018. The politeness Package: Detecting Politeness in Natural Language. *R Journal* 10, 2 (2018). [175]
- Michael Yeomans, Julia Minson, Hanne Collins, Frances Chen, and Francesca Gino. 2020. Conversational receptiveness: Improving engagement with opposing views. *Organizational Behavior and Human Decision Processes* 160 (2020), 131–148. [176] Yong
- Yu, Xiaosheng Si, Changhua Hu, and Jianxun Zhang. 2019. A review of recurrent neural networks: LSTM cells and network architectures. *Neural computation* 31, 7 (2019), 1235–1270. [177] George Yule. 2020. *The study of language*. Cambridge university press. [178] Justine Zhang, Jonathan Chang, Cristian Danescu-Niculescu-Mizil, Lucas Dixon, Yiqing Hua, Dario Taraborelli, and Nithum Thain. 2018. Conversations Gone Awry: Detecting Early Signs of Conversational Failure. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 1350–1361. [179] Yizhe Zhang, Siqi Sun, Michel Galley, Yen-Chun Chen, Chris Brockett, Xiang Gao, Jianfeng Gao, Jingjing Liu, and William B Dolan. 2020. DIALOGPT: Large-Scale Generative Pre-training for Conversational Response Generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics: System Demonstrations*. 270–278. [180] Weixiang Zhao, Yanyan Zhao, Xin Lu, Shilong Wang, Yanpeng Tong, and Bing Qin. 2023. Is ChatGPT Equipped with Emotional Dialogue Capabilities? *arXiv preprint arXiv:2304.09582* (2023). [181] Wayne Xin Zhao, Kun Zhou, Junyi Li, Tianyi Tang, Xiaolei Wang, Yupeng Hou, Yingqian Min, Beichen Zhang, Junjie Zhang, Zican Dong, et al. 2023. A survey of large language models. *arXiv preprint arXiv:2303.18223* (2023). [182] Naitian Zhou and David Jurgens. 2020. Condolence and empathy in online communities. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 609–626. [183] Caleb Ziems, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2023. Can Large Language Models Transform Computational Social Science? *arXiv preprint arXiv:2305.03514* (2023).

# Thank You!

