Expert Systems With Applications 224 (2023) 120025

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

A multi-task learning framework for politeness and emotion detection in dialogues for mental health counselling and legal aid

Priyanshu Priya [a,*], Mauajama Firdaus [b], Asif Ekbal [a,*]

[a] Department of Computer Science and Engineering, Indian Institute of Technology Patna, Patna, India
[b] University of Alberta, Edmonton, Alberta, Canada

ARTICLE INFO

Keywords:
Conversation
Politeness detection
Emotion detection
Multi-task learning
Caps-DGCN

ABSTRACT

The World Health Organization (WHO) has highlighted the need to greatly accelerate the prevention of crime and harassment against women and children, thereby promoting their mental well-being and gender parity to accomplish the United Nations Sustainable Development Goals by 2030. WHO estimates that globally around 736 million women[1] and 1 billion children[2] have been subjected to violence. There is a global scarcity of mental healthcare workers[3]; hence, a conversational artificial intelligent agent that can converse with people in a polite and empathetic way like a human companion can be of great importance. Such systems can serve as first aid for the victims and the entry point for referring them to other support services they may need. Towards this goal, we create a **PO**liteness and **EM**otion annotated dialogue dataset, named **POEM** comprising of 5,000 dialogues in English for mental health counselling and legal aid for women and children crime victims. Furthermore, we propose a novel multi-task learning (MTL) framework, named Caps-DGCN for Politeness and Emotion Detection (PED) in conversations. Experimental results on POEM and DailyDialog datasets suggest that the Caps-DGCN framework achieves a considerable performance and the information shared between the tasks helps in improving the overall system.

1. Introduction

The long-standing objective of Artificial Intelligence (AI) and Natural Language Processing (NLP) is to create conversational agents (CAs) that are capable of mimicking human behaviour. Recent years have marked the development of various conversational agents ranging from open-domain (chit-chat) conversational agents to task-oriented conversational agents like Amazon's Alexa, Microsoft's Cortona, Google's Google Home, etc. The task-oriented dialogue systems are designed to accomplish a particular assignment. These systems aim to assist users in several specific ways, such as answering users' queries on corporate websites, detecting the medical condition, providing legal assistance, etc. The promising potential and alluring commercial and social values of these dialogue systems have led to the adaption of different styles, emotions, and personalities in such systems. Recent research has taken a step ahead and has been inclined towards making such systems polite besides being personal, empathetic, and amicable. The incorporation of politeness in dialogue systems facilitates smooth conversation between the agent and users (Mishra, Firdaus, & Ekbal, 2022b). The dialogue systems that can respond to the users politely and empathetically eventually promote user satisfaction and retention (Firdaus, Ekbal, & Bhattacharyya, 2020; Golchha, Firdaus, Ekbal, & Bhattacharyya, 2019; Zhou, Huang, Zhang, Zhu, & Liu, 2018).

1.1. Motivation

According to the World Health Organization[4] (WHO), about 1 in 3 (30%) of women worldwide have been subjected to either physical and/or sexual violence at least once in their lifetime and about 1 billion children aged 2–17 years, have experienced physical, sexual, or emotional violence or negligence in their lives. According to the United Nations Sustainable Development Goals Report 2021,[5] 10 million girls will be at risk of child marriage over the next decade globally as a result

* Corresponding authors.
E-mail addresses: priyanshu_2021cs26@iitp.ac.in (P. Priya), mauzama.03@gmail.com (M. Firdaus), asif@iitp.ac.in (A. Ekbal).

[1] https://www.who.int/news-room/fact-sheets/detail/violence-against-women
[2] https://www.who.int/health-topics/violence-against-children
[3] https://www.who.int/publications/i/item/9789240036703
[4] https://www.who.int/
[5] https://unstats.un.org/sdgs/report/2021/The-Sustainable-Development-Goals-Report-2021.pdf
[6] https://www.who.int/news-room/fact-sheets/detail/depression
[7] https://www.who.int/india/health-topics/mental-health

https://doi.org/10.1016/j.eswa.2023.120025
Received 25 August 2022; Received in revised form 31 January 2023; Accepted 31 March 2023
Available online 5 April 2023

of COVID-19, which in turn, will promote gender disparity. The WHO estimates that approximately 280 million people in the world have depression.[6] In India alone, about 7.5% population suffer from some mental disorder as per the WHO report.[7] According to the numbers, 56 million Indians suffer from depression and another 38 million suffer from anxiety disorders.

As per the National Mental Health Survey 2015–16 (Gururaj et al., 2016), it was revealed that 9.8 million teenagers in the age group 13–17 years suffer from depression and other mental health disorders and need active intervention. Around two-thirds of married women in India are victims of domestic violence. Violence is negatively affecting women's and children's physical, emotional and mental well-being. Consequently, the need to provide emotional support while being polite and respectful towards the victims, along with a safe and non-judgemental environment, has significantly increased. However, according to the Mental Health ATLAS 2020[8] compiled by the WHO, there is a global shortage of healthcare workers specialized in mental health. In India, per 100,000 population, there are 0.3 psychiatrists, 0.12 nurses, 0.07 psychologists and 0.07 social workers, while the desirable number is anything above 3 psychiatrists and psychologists per 100,000 population. Also, people facing mental health problems do not consult the experts due to the social stigma and prejudice associated with mental illness (White & Dorman, 2001), and the restricted availability of competent mental assistance. Moreover, women and children lag in reporting the assault or abuse due to lack of awareness about their legal and human rights, which calls for qualified legal assistance to seek justice.

To address these challenges, authors in Mohr, Burns, Schueller, Clarke, and Klinkman (2013) suggest that behavioural intervention technologies (BITs) offer a potential solution to overcome barriers that prevent access to mental health treatment, hence, contributing to the expansion of mental health care services. Chatbots or conversational agents (CAs) represent a particular type of BIT to address mental health conditions. The CAs can also be used as a viable way to remove hurdles that impede access to legal consultations. Although a few CAs for mental health-related assistance (*WoeBot* Fitzpatrick, Darcy, & Vierhile, 2017, *KokoBot* Morris, Kouddous, Kshirsagar, & Schueller, 2018, *Wysa* Inkster, Sarda, Subramanian, et al., 2018, *Tess* Fulmer et al., 2018) and legal assistance (Legalbot John, Caro, Robaldo, & Boella, 2017, *DoNotPay*,[9] *LawGeex*,[10] *Legaliboo*,[11] *Convey Law*[12]) have been reported in the past, none of them is designated to provide both mental health and legal counselling to the victims according to their needs. Furthermore, although some of these systems interact empathetically (Fitzpatrick et al., 2017; Inkster et al., 2018), none of these systems converse politely with the users. A few studies have shown that politeness (Golchha et al., 2019; Mishra, Firdaus, & Ekbal, 2022a; Wang et al., 2020) and emotions (Rashkin, Smith, Li, & Boureau, 2018; Zhou et al., 2018), in general, foster the interaction between the agent and the user. The CAs targeting users facing mental health problems can potentially be enhanced with the help of socio-linguistics cues, such as politeness (Bickmore & Picard, 2004; Kim et al., 2018; Newbold, Doherty, Rintel, & Thieme, 2019) and emotions (Castonguay & Hill, 2017; Saha, Reddy, Saha, & Bhattacharyya, 2022; Sharma, Lin, Miner, Atkins, & Althoff, 2021). Perceiving politeness in conversations provides cues about the interlocutors' social behaviours and perceiving emotions provides affective information about them.

We, therefore, hypothesize with the intuition that instiling a combination of these aspects in a CA will be beneficial for the victims facing mental health problems and hesitate to seek legal assistance. For instance, in Fig. 1, though the response shown in the blue box seeks to diagnose the victim's issue, the green box response with polite and emotional expressions may better motivate the victim to respond as it reflects a sense of compassion for, or personal investment in, the user, like a human companion.

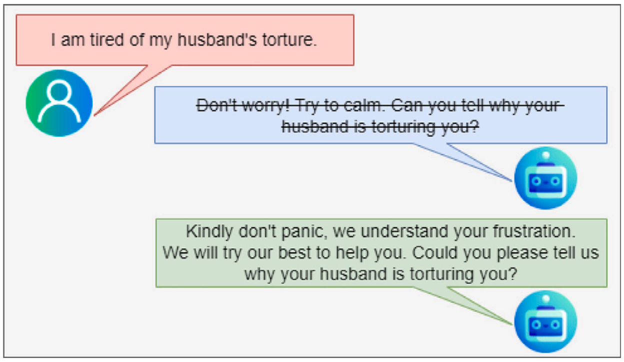

I am tired of my husband's torture.
~~Don't worry! Try to calm. Can you tell why your husband is torturing you?~~
Kindly don't panic, we understand your frustration. We will try our best to help you. Could you please tell us why your husband is torturing you?

Fig. 1. Example demonstrating how the use of polite and emotional (compassion) expressions by the CA promote engagement.

Moreover, recent research suggests that there is a relationship between politeness and emotion (Bothe, 2021; Feng et al., 2021). Politeness helps to differentiate between emotions such as those linked with apology or anger, both of which are intrinsically negative. The apologetic emotion is more likely to be associated with polite demeanor of the interlocutor, and anger, by contrast, is more likely to be associated with impolite behaviour. To illustrate, consider the utterance, "*We are really sorry for the inconvenience. Kindly stay with us for a moment.*", the emotion is apology and the behaviour is polite. On the contrary, in the turn "*Do hell with your sorry! Get lost.*", the emotion expressed is anger and the behaviour is impolite. Based on the findings that politeness identification and emotion detection are complementary, we are motivated to perform the PED task using a deep learning-based multi-task learning (MTL) framework. Prior research works have shown that MTL can enhance tasks performance by learning the similarities and differences among related tasks (Li, Kazameini, Mehta, & Cambria, 2021; Majumder, Poria, Peng, et al., 2019; Ruder, 2017). Therefore, for the PED task in conversations, we propose an MTL framework, Caps-DGCN based on the Capsule Network (Caps) and the Directional Graph Convolutional Network (DGCN). Capsule Network captures the local ordering of words in the utterance and corresponding semantic representations, and the DGCN encodes syntactic information by incorporating the dependency among the words. The identification of politeness and emotion in conversations requires the extraction of semantic as well as syntactic information from an utterance. We believe that our proposed approach strengthens the utterance representation by utilizing both semantic and syntactic information in utterances and helps in PED tasks accordingly.

In NLP, there seems to have considerable research in the understanding and building models for conversational agents (Feng et al., 2021), however, there have been a very few works on developing dialogue systems for mental health and/or legal support (John et al., 2017; Malhotra, Waheed, Srivastava, Akhtar, & Chakraborty, 2022; Saha, Chopra, Saha, Bhattacharyya, & Kumar, 2021; Saha, Reddy, Das, Saha, & Bhattacharyya, 2022; Sharma, Miner, Atkins, & Althoff, 2020). This is primarily due to the scarcity of available data. To address this limitation, we propose a novel dataset consisting of mental health and/or legal counselling conversations between an agent and the victim of different types of crimes. Further, for the PED task in conversations, we annotate this dataset with **PO**liteness and **EM**otion labels and name the resulting dataset as POEM. A sample conversation with

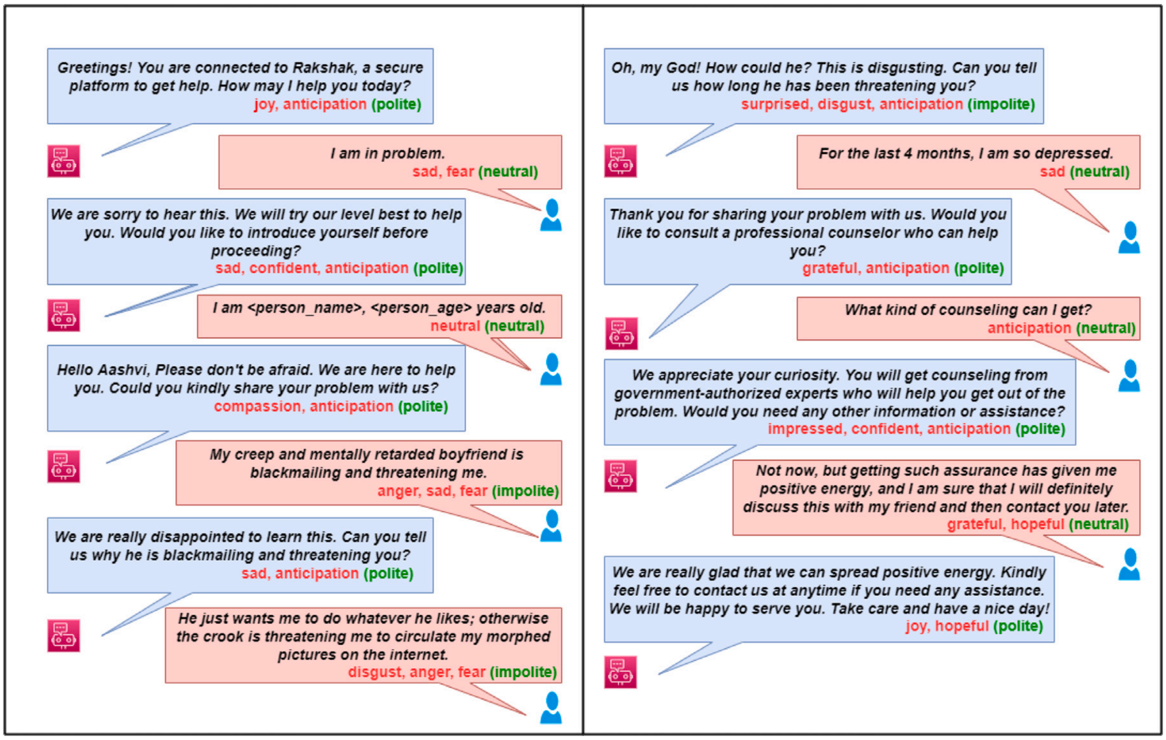
[8] https://www.who.int/publications/i/item/9789240036703

[9] https://donotpay.com/

[10] https://www.lawgeex.com/

[11] https://www.legaliboo.com/

[12] https://www.conveylaw.com/



Greetings! You are connected to Rakshak, a secure platform to get help. How may I help you today? joy, anticipation (polite)

I am in problem. sad, fear (neutral)

We are sorry to hear this. We will try our level best to help you. Would you like to introduce yourself before proceeding? sad, confident, anticipation (polite)

I am <person_name>, <person_age> years old. neutral (neutral)

Hello Aashvi, Please don't be afraid. We are here to help you. Could you kindly share your problem with us? compassion, anticipation (polite)

My creep and mentally retarded boyfriend is blackmailing and threatening me. anger, sad, fear (impolite)

We are really disappointed to learn this. Can you tell us why he is blackmailing and threatening you? sad, anticipation (polite)

He just wants me to do whatever he likes; otherwise the crook is threatening me to circulate my morphed pictures on the internet. disgust, anger, fear (impolite)

Oh, my God! How could he? This is disgusting. Can you tell us how long he has been threatening you? surprised, disgust, anticipation (impolite)

For the last 4 months, I am so depressed. sad (neutral)

Thank you for sharing your problem with us. Would you like to consult a professional counselor who can help you? grateful, anticipation (polite)

What kind of counseling can I get? anticipation (neutral)

We appreciate your curiosity. You will get counseling from government-authorized experts who will help you get out of the problem. Would you need any other information or assistance? impressed, confident, anticipation (polite)

Not now, but getting such assurance has given me positive energy, and I am sure that I will definitely discuss this with my friend and then contact you later. grateful, hopeful (neutral)

We are really glad that we can spread positive energy. Kindly feel free to contact us at anytime if you need any assistance. We will be happy to serve you. Take care and have a nice day! joy, hopeful (polite)

Fig. 2. An example of politeness and emotion annotated dialogue from POEM. The text in green and red represents the politeness and emotion labels, respectively.

politeness and emotion annotation from POEM is presented in Fig. 2. Our present study focuses on the dialogue-understanding module of the conversational system for mental health and legal counselling. The ongoing research in the mental health and legal arena might benefit from this work and aid the NLP community develop CAs capable of comprehending counselling conversations in a better way, which in turn, helps the counsellors and/or enhance accessibility to counselling support.

1.2. Contribution

In summary, our research contributes with respect to the following dimensions:

1. We propose a novel multi-task learning framework, Caps-DGCN based on Capsule Network (Caps) and Directional Graph Convolutional Network (DGCN) for politeness and emotion detection in conversations.
2. We create a politeness and emotion annotated conversational dataset in English, named POEM in Wizard-of-Oz manner for mental health counselling and legal aid for the victims of crimes.[13]
3. Experimental results demonstrate that our proposed framework outperforms the baselines proving the efficacy of our entire network.

The remainder of the paper is organized as follows: In Section 2, we briefly present a review of the related literature followed by the dataset description in Section 3. In Section 4, we will describe the proposed methodology. Experiment details and results are provided in Section 5. In Section 6, we present a detailed analysis of our proposed model followed by the concluding remarks and future directions in Section 7.

2. Related work

In this section, we explore several research works conducted so far in politeness and emotion followed by the works done in mental health-related and legal support in NLP.

Politeness and Emotion in NLP. Politeness has been a key objective of contemporary pragmatic theory since its inception (Brown, Levinson, & Levinson, 1987; Grice, 1975; Lakoff, 1973) because it is the source of pragmatic enhancement, cultural variance and social meaning (Byon, 2006; Matsumoto, 1988; Watts, 2003). Brown and Levinson's (B&L) theory of politeness (Brown et al., 1987) that explained politeness in terms of face-saving strategies marked the beginning of most of the research in this direction. Some computational linguistic study explores the link between politeness and social powers (Danescu-Niculescu-Mizil, Sudhof, Jurafsky, Leskovec, & Potts, 2013) and some shows how machines learn politeness, for example, "*could you*" and "*please*" words signals on the heatmaps of sentences (Aubakirova & Bansal, 2016). In Madaan et al. (2020), the authors have devised a tag and generate framework for converting non-polite sentences into polite ones. Recently, in Niu and Bansal (2018), the authors proposed a reinforced network that could induce politeness in chit-chat conversations in the absence of parallel data. Lately, the authors in Mishra et al. (2022b) attempted to predict politeness in goal-oriented dialogues using a hierarchical transformer network. The authors in Ohbyung and Sukjae (2009) presented a method for a context-aware polite interface that facilitates context-relevant conversations in the Korean language.

Emotion recognition in text has always been in great demand Dheeraj and Ramakrishnudu (2021), Gan, Yang, Zhu, Jain, and Struc (2022), Khoshnam and Baraani-Dastjerdi (2022) and Shao, Chandramouli, Subbalakshmi, and Boyadjiev (2019). Users' feelings in the form of sentiments and emotions have also been exploited in the past for building interactive systems (Acosta, 2009; Acosta & Ward, 2011; Perez-Gaspar, Caballero-Morales, & Trujillo-Romero, 2016; Pittermann, Pittermann, & Minker, 2010; Shi & Yu, 2018). The authors in Zhou and Wang (2017) and Zhou et al. (2018) attempted to generate emotional responses in order to give humanly essence to

[13] The dataset & code will be made available for research purposes.

the system. A few works (Firdaus et al., 2020; Golchha et al., 2019; Gupta, Walker, & Romano, 2007; Mishra et al., 2022a) have explored building more affective and socially intelligent conversational agents by incorporating different politeness strategies in the responses while being empathetic towards the user.

Politeness and emotion have been studied together in recent works as significant social cues (Culpeper & Tantucci, 2021; Langlotz & Locher, 2017; Renner, 2020). In human–machine interaction, politeness and emotion as social cues play a vital role to drive engaging social interaction with machines (Bothe, Magg, Weber, & Wermter, 2018; Srinivasan & Takayama, 2016; Steinfeld et al., 2006). Recently, Bothe (2021) and Feng et al. (2021) attempted to explore the link between politeness and emotion by analysing how polite the utterances are against their respective emotion classes. Our present work is different from the existing work as we have employed a deep learning-based multi-task framework for predicting the politeness and emotion labels of the utterances in dialogues. The domain that we focus on is having great social relevance in today's society as we aim for developing a conversational AI agent to help women and children, who have faced harassment.

NLP for mental health and legal support. With the rising concern for the psychological well-being of individuals, recently, efforts are being made to develop dialogue systems for mental health support (Saha, Gakhreja, Das, Chakraborty, & Saha, 2022; Saha, Reddy, Das, Saha, & Bhattacharyya, 2022; Sharma et al., 2021; Sharma, Miner, Atkins, & Althoff, 2020). Past works have explored diagnosing mental health issues from social media posts and activities (Gaur et al., 2018; Reis, Correia, Murai, Veloso, & Benevenuto, 2019; Yazdavar et al., 2017, 2020, 2018). The existing studies in NLP for mental health assistance and support are primarily focused on analysing effective approaches for obtaining and delivering conversational support, such as context-specific adaptation and response diversity (Althoff, Clark, & Leskovec, 2016; Pérez-Rosas, Wu, Resnicow, & Mihalcea, 2019; Sharma & De Choudhury, 2018; Yang, Yao, Seering, & Kraut, 2019; Zhang & Danescu-Niculescu-Mizil, 2020). Besides, researchers have developed techniques for gauging the linguistic development of counsellors (Zhang, Filbin, Morrison, Weiser, & Danescu-Niculescu-Mizil, 2019), extracting conversational engagement patterns (Sharma, Choudhury, Althoff, & Sharma, 2020), analysing moderation (Wadden, August, Li, & Althoff, 2021), detecting therapeutic actions (Lee, Hull, Levine, Ray, & McKeown, 2019), and identifying cognitive restructuring (Pruksachatkun, Pendse, & Sharma, 2019) in supportive talks. Lately, efforts have been made to understand and build computational methods for identifying empathy in face-to-face therapy (Gibson et al., 2016; Pérez-Rosas, Mihalcea, Resnicow, Singh, & An, 2017), and text-based peer-to-peer support system (Sharma, Miner, Atkins, & Althoff, 2020). The other works like (Saha, Chopra, Saha, Bhattacharyya, & Kumar, 2021; Saha, Gakhreja, Das, Chakraborty, & Saha, 2022; Saha, Reddy, Das, Saha, & Bhattacharyya, 2022; Saha, Reddy, Saha, & Bhattacharyya, 2022; Sharma et al., 2021) have also explored empathy in mental health support systems. Prior research has demonstrated that the perception of conversational agents as compassionate or polite can lead to a sense of empathy and facilitate the disclosure of sensitive information (Bickmore & Picard, 2004; Kim et al., 2018; Lucas, Gratch, King, & Morency, 2014) and support mental health self-management (Matthews & Doherty, 2011). The authors in Newbold et al. (2019) suggest the use of politeness strategies to create various dialogue templates, reflecting distinct agent personas and show how this can influence people's interactions with the agent and their willingness to self-disclose about mental health issues.

Recently, intelligent systems for the legal domain have also grown in popularity. A legal question–answering system was proposed by Do, Nguyen, Tran, Nguyen, and Nguyen (2017) that can respond to the user's legal query in YES/NO. The authors in McElvain et al. (2019) presented a question–answering system that gives jurisdictionally relevant, legally accurate, and conversationally responsive answers to the legal questions asked by the users. The authors in Kowsrihawat, Vateekul, and Boonkwan (2018) proposed a deep learning-based prediction system for the decision of criminal cases. The authors in John et al. (2017) developed a CA, named *"Legalbot"* capable of answering user queries posed as questions during conversation. *Legalbot* can provide simple legal advice to users.

The existing studies focus on providing either mental health support or legal support to the users. Furthermore, most of the existing works explore emotion or empathy in mental health support. Our research differs in the sense that we aim to identify both politeness and emotion in goal-oriented mental health and legal support conversations. We believe that the present work facilitates developing a polite and empathetic conversational agent for mental health as well as legal counselling assistance to the users (victims).

Multi-task Learning in NLP. Multi-task Learning (MTL) paradigm aims to accomplish generalization by using the inter-relatedness of multiple tasks. In a typical MTL context, the same input representation is shared among several tasks. MTL tries to exploit the dependency between several associated tasks in order to improve the performance of the individual tasks. MTL has been shown to be effective for various NLP tasks. For instance, there is a line of work that utilizes the MTL approach for improving the end-to-end aspect-based sentiment analysis (ABSA) task, which includes two sub-tasks: aspect term extraction and sentiment detection. The authors in Liang et al. (2020), Lin, Sun, and Wang (2022) and Wang, Sun, Huang, and Zhu (2019) leveraged a capsule-based learning network for ABSA. The works in Liang et al. (2021) and Yang, Wu, Li, and Wang (2020) exploited Graph Convolutional Networks (GCNs) based architecture for ABSA. Several works proposed multi-task models to exploit the correlation between dialogue act recognition (DAR) and sentiment classification (SC) and achieved state-of-the-art performances on both the tasks (Cerisara, Jafaritazehjani, Oluokun, & Le, 2018; Qin, Che, Li, Ni, & Liu, 2020; Qin, Li, Che, Ni, & Liu, 2021; Xu, Yao, Liu, Liu, & Xu, 2023). The authors in Qin et al. (2020) proposed a relation layer to explicitly model the interaction between the two tasks. In Qin et al. (2021) and Xu et al. (2023), the authors proposed graph attention network-based approaches for jointly performing DAR and SC tasks.

Recently, the authors in Marreddy, Oota, Vakada, Chinni, and Mamidi (2022) developed an MTL model utilizing GCN, named MT-Text GCN for conducting sentiment analysis, emotion identification, hate-speech and sarcasm detection tasks concurrently for the Telugu language. The work in Li, Zhang, Ji, and Liu (2020) introduced an MTL network for emotion recognition in conversations using speaker identification as an auxiliary task to better capture the speaker-related information and improve the performance of the primary emotion recognition task. In Chen, Hou, Cheng, and Li (2018), the authors proposed to identify emotion and emotion cause simultaneously using the MTL approach. In Li, Braud, and Amblard (2022), the authors investigated a multi-task hierarchical approach for simultaneously detecting depression, topic, emotion, and dialogue act information in conversations. Lately, Wang et al. (2022) proposed Context-based Hierarchical Attention Capsule (Chat-Capsule) for solving utterance-level emotion detection and dialogue-level opinion analysis (user satisfaction and emotion curve category) tasks simultaneously in a unified framework. The authors in Saha, Priya, Saha, and Bhattacharyya (2021), Staliūnaitė and Iacobacci (2020) and Zhang, Li, Du, Fan, and Yu (2018) utilized capsule networks to learn the intent detection and slot-filling tasks jointly.

Motivated by the association of politeness and emotion and the advantages of the MTL paradigm, we presented a multi-task framework that jointly learns and classifies the politeness and emotions of the utterances in dialogues.

3. Dataset

In this section, we describe the details of the dataset, POEM that we create for our experiments followed by a brief description of the DailyDialog dataset.

3.1. POEM

One of the notable contributions of our present work is a large-scale **PO**liteness and **EM**otion annotated conversational dataset in English, named as POEM. The dataset comprises dyadic conversations between the agent and crime victims who are in need of either mental health counselling or legal counselling or both. We believe that this dataset will aid in the development of robust goal-oriented dialogue systems for mental health and legal counselling assistance. A sample conversation from our proposed POEM dataset is depicted in Fig. 2 in the introduction section. In the subsequent part, we initially explain the dataset creation process, its statistics and quality survey followed by comparisons to the existing goal-oriented conversational datasets for mental health and legal counselling to demonstrate the usefulness and relevance of our work.

3.1.1. POEM dataset creation process

The primary stage in the data creation process was to determine the domain for which conversations were to be made. As per the statistics given in the earlier sections, the increasing number of crimes against women and children is a major public concern as it leads to various health problems for them and violations of their human rights. Hence, we have decided to create conversations targeting mental health and/or legal counselling for women and children victims of different categories of crimes. The motivation behind the domain selection is that the victims of crimes generally need support and assistance, which is often fundamental to their recovery. Victims may require mental health-related counselling for their emotional comfort and psychological well-being. They may also need legal support to seek justice and ensure their physical safety. The provision of early support can help them to prevent bigger and more complex problems that victims may face in the future.

The next step in the data creation process was to decide the types of crimes. Following in-depth examination and thorough deliberation with the domain experts, we have finalized the two criteria for crime selection: (i) conventional crimes against women and children, and (ii) cybercrimes against women and children. By conventional crimes, we mean the crimes that have been committed using manual ways (do not involve the use of the internet) like domestic violence, rape, acid attacks, etc. On the contrary, cybercrimes are crimes committed using the means of the internet and technology like cyber-stalking, online harassment, and masquerading, to name a few. Eventually, we have decided to create conversations for a total of 16 different types of crimes, covering conventional as well as cybercrimes, namely domestic violence, rape, acid attacks, physical/cyber-stalking, workplace harassment, online harassment, impersonation, trolling, matrimonial fraud, financial fraud, child pornography, women/child trafficking, non-consensual sexting, doxing/outing, and exclusion.

To the best of our knowledge, there is no dialogue corpus available to facilitate the development of a goal-oriented dialogue system in the selected domain. The key steps involved in the process of data creation are: 1. Designing guidelines for data creation, 2. Preparing a large-scale conversational dataset, and 3. Annotation of politeness and emotion.

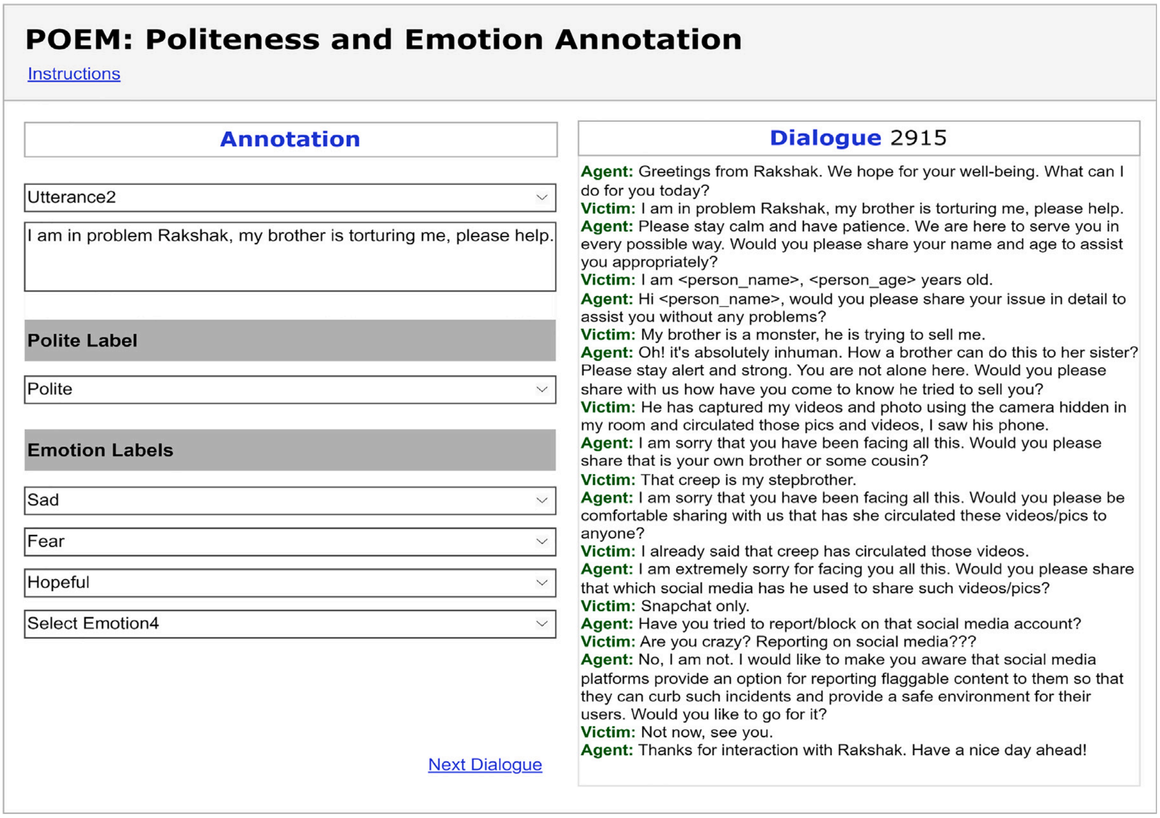
1. **Designing Guidelines for Dataset Creation.** We collaborated with a prominent mental health expert from the National Institute of Mental Health and Neurosciences (NIMHANS),[14] Bengaluru and a legal expert from a government-run institution of national repute to comprehend the dialogue flow in victims' contexts and to assist us in designing the guidelines for mental health and/or legal support. Through the experts' interactions, we understand the various types of intricacies in a natural conversation for our selected domain and identify that appropriate information is crucial for creating realistic, natural and free-flowing dialogues. Hence, we first drafted the guidelines in consultation with the experts for creating the dialogues, which are as follows:

   - Enquire about the problem of victims;
   - Be patient while interacting with the victims. Do not pressurize them for details. Let them decide how much they are willing to share and ask them how you can help;
   - Evaluate the immediate needs of the victims, whether they need psychological counselling or are they looking for legal help;
   - Be empathetic and polite during counselling to provide emotional comfort and non-judgemental environment to the victims;
   - Motivate the victims who are facing psychological problems to be optimistic and get themselves engaged in the activities that make them feel better in order to ensure their ultimate and speedy recovery;
   - Encourage the victims to seek medical attention, report the assault and/or contact organizations that can help them;
   - Help the victims in identifying the ways in which they can re-establish their sense of physical and emotional safety, let the victim must ultimately decide what to do;
   - Respect victims' privacy and make them believe that their information will be confidential;
   - Lay out legal options: Assessing whether the victims wish to report the assault? If yes, provide them with relevant authentic legal information and notify them that pursuing legal action requires collaboration with legal services, local police and forensic services and help them accordingly;
   - Provide a few general safety tips to the victims to ensure that they can make themselves aware of the crimes and prevent such incidents in future.

2. **Preparing Victim-Agent Conversations.** We recruited six people (workers, hereafter) for creating the dialogues and trained them in an interactive session under the supervision of experts. During training, the workers were made aware of the different types of crimes, various mental health-related problems and their potential solutions, and the relevant legal information for creating the dialogues. Besides, the guidelines for dialogue creation were explained to the workers along with a few example dialogues for better understanding. The workers were asked to first collect a few basic information about the victims like their name, age, gender, residence etc. to get familiar with them. Then, enquire about victims' problems and accordingly provide them with mental health or legal support or both.

   For creating the dialogues, we crawled the real-life stories of crimes related to women and children from various websites covering news articles or case studies of such crimes. Furthermore, we explored several websites, *viz.* National Cybercrime Reporting Portal, National Commission for Women, Ministry of Women and Child Development, to name a few; and documents, *viz.* Criminal Law Amendment Act 2013, Information Technology (Amendment) Act 2008, etc. to collect the authentic information pertaining to the mental health counselling and legal services. The workers were then advised to read the stories and create conversations based on these stories and relevant authentic information using a Wizard-of-Oz approach (Kelley, 1984) while keeping the guidelines mentioned in the previous section in mind. For every conversation, the workers were asked to make a pair of two people, of which one person (called as *wizard*) is assumed to play the role of the system agent while the other one acts as a victim. The workers were given the roles of the agent and victim at random. This random role assignment

[14] https://nimhans.ac.in/



POEM: Politeness and Emotion Annotation

Instructions

Annotation

Utterance2

I am in problem Rakshak, my brother is torturing me, please help.

Polite Label

Polite

Emotion Labels

Sad

Fear

Hopeful

Select Emotion4

Next Dialogue

Dialogue 2915

Agent: Greetings from Rakshak. We hope for your well-being. What can I do for you today?
Victim: I am in problem Rakshak, my brother is torturing me, please help.
Agent: Please stay calm and have patience. We are here to serve you in every possible way. Would you please share your name and age to assist you appropriately?
Victim: I am <person_name>, <person_age> years old.
Agent: Hi <person_name>, would you please share your issue in detail to assist you without any problems?
Victim: My brother is a monster, he is trying to sell me.
Agent: Oh! it's absolutely inhuman. How a brother can do this to her sister? Please stay alert and strong. You are not alone here. Would you please share with us how have you come to know he tried to sell you?
Victim: He has captured my videos and photo using the camera hidden in my room and circulated those pics and videos, I saw his phone.
Agent: I am sorry that you have been facing all this. Would you please share that is your own brother or some cousin?
Victim: That creep is my stepbrother.
Agent: I am sorry that you have been facing all this. Would you please be comfortable sharing with us that has she circulated these videos/pics to anyone?
Victim: I already said that creep has circulated those videos.
Agent: I am extremely sorry for facing you all this. Would you please share that which social media has he used to share such videos/pics?
Victim: Snapchat only.
Agent: Have you tried to report/block on that social media account?
Victim: Are you crazy? Reporting on social media???
Agent: No, I am not. I would like to make you aware that social media platforms provide an option for reporting flaggable content to them so that they can curb such incidents and provide a safe environment for their users. Would you like to go for it?
Victim: Not now, see you.
Agent: Thanks for interaction with Rakshak. Have a nice day ahead!

Fig. 3. Our Annotation interface. The right part shows the entire dialogue and the left part shows the utterance for which annotation has to be done.

assisted in removing the correlation between the agent's counselling techniques and the peculiarities of the targeted victim. The workers were further instructed to change their conversation partner everyday to maintain the diversity and realistic nature of conversations. The workers were also asked to continue the conversation until the victim's need was fulfilled or the victim did not want to interact further and the agent is supposed to carry out a system function by providing the relevant assistance to the victim.

Although we employ a well-established Wizard-of-Oz approach for preparing the data, as reported in numerous past studies (Budzianowski et al., 2018; Peskov et al., 2019), our dataset is novel in the sense that it is developed under the complete supervision of experts, and we deliberately supervise and assist the workers to participate in the process of preparing diverse, informative, and engaging conversations.

3. **Data Annotation.** The proposed POEM dataset has two types of annotation, *viz.* politeness annotation and emotion annotation. For politeness annotation, we consider three politeness labels, namely *polite*, *impolite* and *neutral*. For emotion annotation, 16 emotion categories, namely *anticipation, confident, hopeful, anger, sad, joy, compassion, fear, disgust, annoyed, grateful, impressed, apprehensive, surprised, guilty, trust* are taken into account. The emotion annotation list has been extended to incorporate one more label, namely *neutral*. The *neutral* label is designated to utterances having no-emotion.

The entire annotation process is carried out using three in-house annotators (two male and one female). Two annotators have Ph.D. degrees and one has a post-graduate degree. All the annotators are highly proficient in the English language and have good exposure in the related task. The guidelines for annotation along with some examples were explained to the annotators before starting the annotation process. The annotators were asked to label each utterance with one of the three politeness categories and one or more emotion categories. Fig. 3 depicts our annotation interface. We achieved the overall Fleiss' (Fleiss, 1971) kappa score of 0.79 for the politeness labels, and 0.84 for emotion categories, which can be considered reliable. In Fig. 4, we present the annotation matrix of politeness and emotion annotation for a few utterances from our dataset. For this subset of data, we obtained Fleiss' (Fleiss, 1971) kappa scores of 0.78 and 0.85 for the politeness and emotion categories, respectively, which is comparable to the scores obtained for the entire dataset. The utterances for which the annotators could not reach an agreement on the politeness or emotion labels were marked as neutral. The politeness labels, emotion categories, and the annotation process were finalized in consultation with experts.

Table 1
POEM dataset statistics.

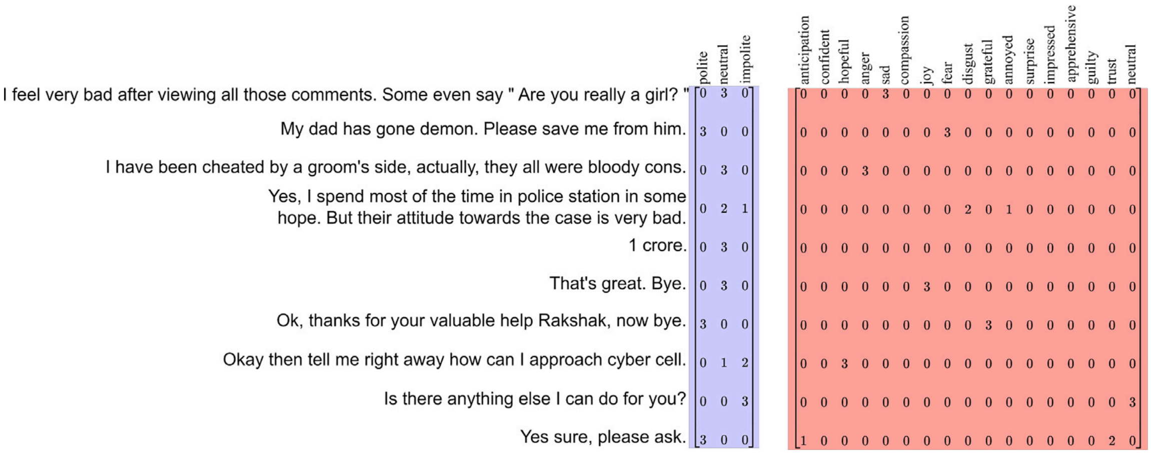
Metrics	Train	Validation	Test
# of dialogues	2859	1080	1061
# of utterances	77,806	25,775	25,744
Avg. utterances per dialogue	27.21	23.87	24.26

3.1.2. Dataset statistics

Table 1 provides the statistics of our proposed POEM dataset. The dataset is split into train, validation and test sets. We present information on the total number of dialogues, the total number of utterances, and the average number of utterances per dialogue in each set in the table. The politeness and emotion distribution in POEM is depicted in Fig. 5.

3.1.3. Quality survey

To build a robust model for any specific application, clean and high-quality data are required. To ascertain that our proposed dataset is realistic and not flawed in terms of dialogues having inconsistent dialogue flows, unnatural responses and biased towards the language



	polite	neutral	impolite	anticipation	confident	hopeful	anger	sad	compassion	joy	fear	disgust	grateful	annoyed	surprise	impressed	apprehensive	guilty	trust	neutral
I feel very bad after viewing all those comments. Some even say " Are you really a girl? "	0	3	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0
My dad has gone demon. Please save me from him.	3	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0
I have been cheated by a groom's side, actually, they all were bloody cons.	0	3	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0
Yes, I spend most of the time in police station in some hope. But their attitude towards the case is very bad.	0	2	1	0	0	0	0	0	0	0	0	2	0	1	0	0	0	0	0	0
1 crore.	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
That's great. Bye.	0	3	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0	0	0	0
Ok, thanks for your valuable help Rakshak, now bye.	3	0	0	0	0	0	0	0	0	0	0	0	3	0	0	0	0	0	0	0
Okay then tell me right away how can I approach cyber cell.	0	1	2	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0
Is there anything else I can do for you?	0	0	3	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	3
Yes sure, please ask.	3	0	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	2	0

Fig. 4. Annotation matrices for a few sample utterances from the POEM dataset. The matrices highlighted in blue and pink show politeness and emotion annotation matrices, respectively.

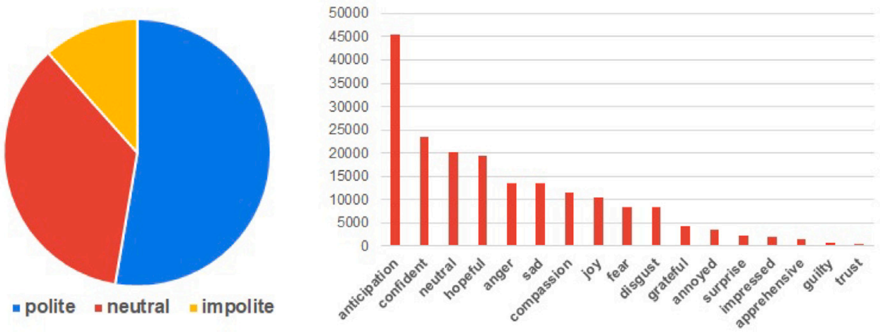


Fig. 5. Politeness and Emotion distribution in POEM.

understanding of a particular individual, we conducted a comprehensive quality check on the POEM dataset. For this survey, we randomly sampled 1000 conversations from the dataset. Further, we employed four people (evaluators, hereafter) highly proficient in English and well-acquainted with the labelling tasks, and the notion of politeness and emotions. The evaluators were distinct from those involved in the data creation process. They were given the instructions to (i) ensure that dialogues are consistent, meaningful and natural and (ii) emotion and politeness annotations of each utterance in the dialogue are correct. These evaluators were requested to provide an overall rating to the dialogues on a scale of 1–5 (5 being the highest and 1 being the lowest rating) depending on the following criteria:

1. Presence of significant errors like the victim's utterance was wrongly interpreted and the response was given, the incorrect annotation for either politeness or emotion or both.
2. Presence of trivial errors like spelling mistakes or grammatical errors.

During the entire quality survey process, we kept in touch with expert from government-run institution. For every 200 human-evaluated conversations, a randomly sampled batch of 20 conversations was given to the experts for cross-verification in terms of quality and resemblance to real-life scenarios. Once the expert approved the evaluation process, we proceed to evaluate the further dialogues. The results of the survey are documented in Table 2. We obtain an average rating of 4 out of 5 which indicates that the dataset is viable and has few errors. Nevertheless, minute mistakes in the dataset are inevitable as we have created and annotated the data manually. We should also take into account the different viewpoints of every individual who plays a significant role in this task. Even though such error exists, they are fewer in number, and these types of errors are expected in any data creation process done on such a large scale. Thus, it can be concluded from the quality survey results that conversations in the dataset are of good quality and natural.

Table 2
Quality survey for POEM dataset.

Rating	Rating description	Surveyed dialogues (%)
5	The dialogues are natural and annotations are correct	42.5
4	Presence of trivial errors like spelling or grammatical mistakes	28
3	The annotations have some discrepancy for polite labels	18
2	The emotion labelling is incorrect for a few utterances	11.5

3.1.4. Comparison with the related datasets

During our preliminary examination of the existing mental health-related datasets, we observe that there are several datasets available for the mental health study that are based on social media content, for instance, Twitter (De Choudhury, Sharma, Logar, Eekhout, & Nielsen, 2017; Yazdavar et al., 2017, 2020), Reddit (Gaur et al., 2018; Yazdavar et al., 2018), etc. However, some of these datasets (De Choudhury et al., 2017; Yazdavar et al., 2017) consist of only self-disclosing and self-expressive postings from anonymized social media users with no specific dyadic or multiparty dialogues to utilize. In the past, a few conversational datasets have been introduced (Callejas, Griol, & López-Cózar, 2011; Gratch et al., 2014; Howes, Purver, & McCabe, 2014; Pérez-Rosas et al., 2019). Some of these datasets consist of text-based counselling conversations (Althoff et al., 2016; Dowling & Rickwood, 2014, 2016; Howes et al., 2014), while others comprise face-to-face counselling conversations such as in DAIC-WOZ (Gratch et al.,

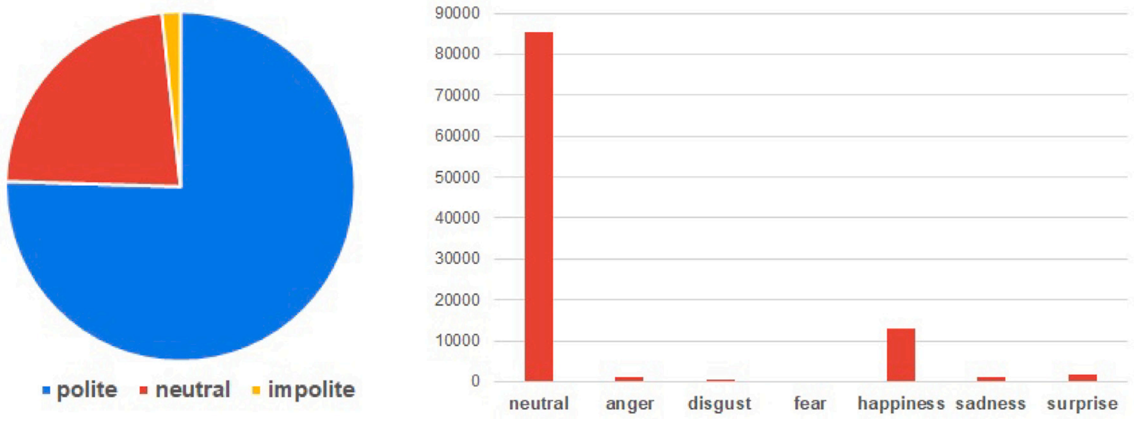


Fig. 6. Politeness and Emotion distribution in DailyDialog.

2014). However, some of the datasets are either not publicly available for research purposes (Althoff et al., 2016) or contain only a small number of conversations (Gratch et al., 2014). Recently, the authors in Saha, Chopra, Saha, Bhattacharyya, and Kumar (2021) proposed a large-scale dataset called *MotiVAte* that consists of dyadic motivational conversations between a depressed user and the virtual agent.

Lately, a few datasets for the legal domain have been released to facilitate the research in legal NLP. For instance, the authors in Zhong et al. (2020) introduced JEC-QA dataset for legal question–answering. Several other studies presented datasets comprising legal question–answer pairs (McElvain et al., 2019). The authors in Hendrycks, Burns, Chen, and Ball (2021) proposed a novel Contract Understanding Atticus Dataset (CUAD) for legal contract review. The datasets for the legal domain in languages other than English have also been introduced in recent years (Kapoor et al., 2022; Xiao et al., 2018). In John et al. (2017), the authors introduced a conversational dataset prepared using legal question–answer pairs.

All the existing datasets are targeted for study in either mental health or legal domain. None of these datasets is designated for both mental health and legal support and assistance. The authors in Singh, Priya, Firdaus, Ekbal, and Bhattacharyya (2022), recently released a dataset in Hindi named *EmoInHindi* for mental health and legal counselling of crime victims. Our proposed POEM dataset is novel in that it comprises large-scale dyadic conversations in English between the agent and a victim who needs mental health or legal support or both. Table 3 shows the comparison between the existing datasets and our proposed POEM dataset.

3.2. DailyDialog

DailyDialog (Li et al., 2017) is a manually labelled multi-turn conversational dataset that covers various topics of daily life ranging from ordinary life topics to financial topics. All utterances in this dataset are labelled with both emotion categories and dialogue acts (intention). Recently, Bothe (2021) annotated all the utterances of DailyDialog with a politeness score ranging from 1 to 5 on a politeness scale by utilizing a pre-trained politeness recognizer (Bao, Wu, Zhang, Chandrasekharan, & Jurgens, 2021), where, a score around 3 indicates neutral, the score inclined towards 1 on the politeness scale indicates impolite, and the score inclined towards 5 on the politeness scale indicates polite.

In this work, we have used the DailyDialog dataset for politeness and emotion detection tasks. The emotion of the utterances belongs to one of the seven categories, *viz. anger, disgust, fear, joy, neutral, sadness,* and *surprise*. For the politeness detection task, all the utterances with the politeness score >= 1 and < 2.8 are marked as impolite, politeness score >= 2.8 and < 3.2 are marked as neutral, and politeness score >= 3.2 and <= 5 are marked as polite. The politeness score range for polite, neutral and impolite classes is determined empirically through inspection of a small subset of samples.

3.2.1. Dataset statistics

We provide the statistics for the DailyDialog dataset in terms of the total number of dialogues, the total number of utterances, and the average number of utterances per dialogue in train, validation and test sets in Table 4. The politeness and emotion distributions in DailyDialog are shown in Fig. 6.

4. Methodology

Inspired by the success of Capsule Network (Du, Sun, Wang, Qi, Liao, Wang, et al., 2019; Du, Sun, Wang, Qi, Liao, Xu, et al., 2019; Xiao, Zhang, Chen, Wang, & Jin, 2018) and Directional Graph Convolutional Network (Chen, Tian, & Song, 2020) in NLP, we propose our MTL framework Caps-DGCN. The detailed description of our proposed Caps-DGCN model is presented in this section and the overall architecture of the proposed approach is depicted in Fig. 7.

4.1. Task definition

Given an utterance $U_t = (w_1, w_2, \dots, w_m)$ consisting of m words of a dialogue D_i annotated with politeness and one or more emotion labels, the task objective is to detect the politeness and emotion labels of the utterance, U_t. Formally, let, D denotes the dataset consisting of the N dialogues, then $D = \{D_1, D_2, \dots, D_N\}$. A dialogue $D_i = \{(U_t, y_t^P, y_t^E)\}_{t=1}^{n}$ denotes the sequence of n utterances with corresponding politeness label $y_t^P \in L^P$ and emotion labels $y_t^E = \{y_t^E \in \{0,1\}^{L^E}\}$, where L^P and L^E are the total number of politeness and emotion labels, respectively. The proposed approach work towards maximizing the following objective function:

$$\text{argmax}_\theta\left(\prod_{t=1}^{n} P(\hat{y}_t^P, \hat{y}_t^E | U_t, y_t^P, y_t^E; \theta)\right) \tag{1}$$

where, $\hat{y}_t^P \in L^P$ and $\hat{y}_t^E \in \{0,1\}^{L^E}$ denote the predicted politeness label and predicted emotion labels, respectively. θ denotes the model parameters to be optimized.

4.2. Proposed framework

The proposed Caps-DGCN framework presented in Fig. 7 is mainly composed of the following five parts: an utterance encoder, a capsule network layer (Caps), a directional graph convolutional network layer (DGCN), an aggregation layer, and the output layer. The utterance encoder employs Bidirectional Encoder Representations from Transformers (BERT) (Devlin, Chang, Lee, & Toutanova, 2018) to encode the input tokens and produces a contextual representation of the input utterance for the computation by the Caps and DGCN layers. The Caps layer uses the capsule network to further extract the spatial and

Table 3
Comparison between different existing datasets and our proposed POEM dataset.

Dataset	Domain		Conversational or not	Size	Description
	Mental health	Legal			
Yazdavar et al. (2017)	✓		×	21 million tweets	Detecting clinical depression from user's tweets.
De Choudhury et al. (2017)	✓		×	50 million tweets	Gender and cross-cultural differences in mental illness disclosures.
Gaur et al. (2018)	✓		×	1.1 million reddit posts	Detecting mental health information on issues like anxiety, personality disorder etc.
Callejas et al. (2011)	✓		✓	100 dialogues	Detecting mental health of the speakers from conversations.
Gratch et al. (2014)	✓		✓	189 dialogues	Face-to-face counselling conversations between the interviewer and patient with mental health issues (depression, anxiety etc.)
Pérez-Rosas et al. (2019)	✓		✓	259 dialogues	Motivational interviews compiled from YouTube and Vimeo video titles on issues such as quitting smoking, alcohol abstinence, etc.
Howes et al. (2014)	✓		✓	882 dialogues	Text-based system for conversations between depressed users and the therapist.
Althoff et al. (2016)	✓		✓	80885 dialogues (only 15555 dialogues used for analysis)	SMS-based conversations collected for handling crisis situations like suicidal thoughts, depression, etc.
Zhong et al. (2020)		✓	×	26365 questions	Question–answer collected from the National Judicial Examination of China for building legal QA system.
Hendrycks et al. (2021)		✓	×	510 contracts and 13101 labelled clauses	Expert-annotated dataset for legal contract review.
John et al. (2017)		✓	✓	1200 dialogues	Question–answer pairs formatted into dialogues between the agent and user to answer simple legal questions.
Singh et al. (2022)	✓	✓	✓	1814 dialogues	Text-based mental health and/or legal counselling conversations between the agent and a victim in Hindi.
POEM	✓	✓	✓	5000 dialogues	Text-based mental health and/or legal counselling conversations between the agent and a victim in English.

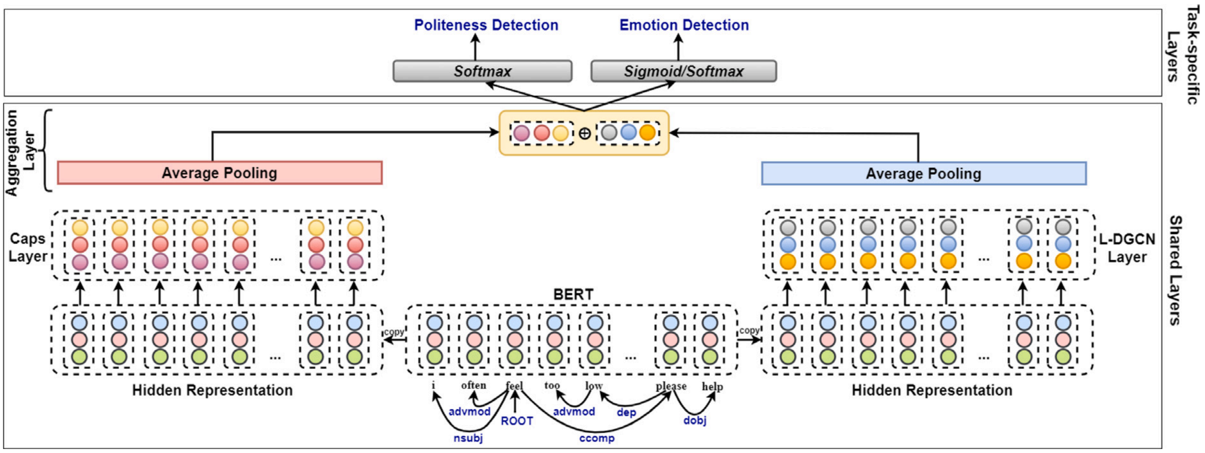


Fig. 7. Architecture of the proposed Caps-DGCN framework.

Table 4
Daily dialogue 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

hierarchical features from the text sequence via the dynamic routing algorithm, which are compressed into the representation with pre-determined size aiming to capture more local n-gram features reflecting its semantic meaning at different positions that finally span over the entire sequence.

The DGCN layer handles the dependency tree of the input utterance and convolutes over the dependency parse tree. This layer models the contextual features having different positional relationships with their related word and then weight these features according to the comparison among them. The aggregation layer combines the semantic and syntactic feature representation of Caps and DGCN layers. Finally, the output layer consisting of the task-specific *Softmax* and *Sigmoid* layers, are applied to predict the politeness and emotion labels of the input utterance, respectively. Our proposed multi-task model shares all the trainable parameters of the utterance encoder, Caps, DGCN and aggregation layers, and the parameters of the output layer are specific for each task.

Utterance Encoder. Given an input utterance, $U_t = (w_1, w_2, \dots, w_m)$, the utterance encoder employs BERT, a multi-layer bidirectional Transformer encoder which is based on the original Transformer model (Vaswani et al., 2017) for encoding the input and obtaining the hidden vector, h_i for each token w_i of the input utterance U_t:

$$h_i = BERT(w_i), \forall w_i \in U_t \tag{2}$$

where, $h_i \in \mathbb{R}^{d_h}$; d_h is the dimension of hidden state. Thus, the final representation of utterance encoder can be denoted as $H^E = [h_1^E, h_2^E, \dots, h_m^E]$. This hidden representation will be the input to the Caps and DGCN layers as we will describe in the subsequent section.

Capsule Network Layer. The hidden sequence H^E might contains intrinsic spatial and hierarchical information that would be informative for the model. Thus, we further use a capsule network to extract such local features from H^E. In particular, the capsule network automatically learns child–parent relationships from the output sequence H^E that are viewed as input capsules, and compresses them into the representations with pre-determined size. Formally, for the child capsule h_i^E, each parent capsule Q_j aggregates all the incoming messages s_j from each child capsule and squashes s_j to $\|s_j\| \in (0, 1)$ via the squash function (Sabour, Frosst, & Hinton, 2017) as shown below:

$$Q_j = \frac{\|s_j\|^2}{1 + \|s_j\|^2} \frac{s_j}{\|s_j\| + e^{-7}} \tag{3}$$

$$s_j = \sum_i c_{ij} H_{j|i}^E \tag{4}$$

Here, c_{ij} is the coupling coefficient, which measures proportionally how much information would be transferred from $H_{j|i}^E$ to Q_j. The dynamic routing (Dou et al., 2019; Gong, Qiu, Wang, & Huang, 2018) is implemented for refining c_{ij}. At p_{th} iteration, the c_{ij} is computed via a softmax function for ensuring that all the information from the child capsule h_i^E can be transferred to the parent and is computed as:

$$c_{ij}^p = \frac{e^{(a_{ij}^{p-1})}}{\sum_k e^{(a_{ik}^{p-1})}} \tag{5}$$

$$a_{ij}^p = a_{ij}^{p-1} + H_{j|i}^{E_p} Q_j \tag{6}$$

where, a_{ij}^p is iteratively updated and is initialized to 0. And, $H_{j|i}^E$ represents the information transferred from h_i^E to Q_j and is computed as follows:

$$H_{j|i}^E = ReLU(W_{ij}^T h_i^E) \tag{7}$$

where, $W_{ij} \in R^{d_h \times (n_c * d_c)}$; n_c and d_c denotes the number of capsules and dimension of each capsule, respectively.

After encoding the sequence H^E into n_c capsules, we concatenate these capsules into a vector $H^C = [Q_1; Q_2; \dots; Q_{n_c}]$, where ; represents the concatenation operation and $H^C \in \mathbb{R}^{d_h}$ denotes the source sequence's representations containing its child–parent relationship features.

Directional Graph Convolutional Network Layer. The directional graph convolutional layer learns the syntactically relevant words to the target words on the dependency tree. This layer separately models the contextual features which have different positional relationships with their associated words and then weights these features based on the comparison among these features.

In particular, for a given input utterance, we first construct a word relation graph from its auto-processed dependency parse trees. Afterwards, we apply a direction mechanism in Graph Convolutional Network (GCN) to encode each word's related contextual features (as implied by the graph) with regard to distinct positional relationships i.e., left, right or self. Finally, we introduce the attention mechanism to distinguish the importance of distinct contextual features by giving different weights to these features that are calculated based on the comparisons among these features, in order to accentuate important syntactic features for the PED task.

Formally, the obtained hidden state H^E is fed into the stacked DGCN model, which is performed in a multilayer fashion with an L DGCN layer and the output in the lth DGCN layer for w_i is computed as follows:

$$h_i^{(l)} = ReLU\left(\sum_{j=1}^{n} d_{ij}\left(W_{dir}^{(l)} \cdot h_j^{(l-1)} + b_{dir}^{(l)}\right)\right) \tag{8}$$

where $W_{dir}^{(l)}$ and d_{ij} denote the direction modelling and attention mechanism, respectively.

For the direction modelling, the positional relationship of all w_j with respect to w_i is encoded via $W_{dir}^{(l)}$ and has three options, $W_{left}^{(l)}$, $W_{right}^{(l)}$ and $W_{self}^{(l)}$ for different i and j. In particular, if $i < j$, then $W_{dir}^{(l)} = W_{right}^{(l)}$. In such case, attention is applied to the edge between w_i and w_j to weight distinct contextual features. The attention, d_{ij} is computed using Eq. (9):

$$d_{ij}^{(l)} = \frac{b_{ij} \cdot e^{(h_i^{(l-1)} \cdot h_j^{(l-1)})}}{\sum_{j=1}^{m} b_{ij} \cdot e^{(h_i^{(l-1)} \cdot h_j^{(l-1)})}} \tag{9}$$

Here, $h_i^{(l-1)} \cdot h_j^{(l-1)}$ computes the interaction between w_i and w_j through inner product. For computing d_{ij}, b_{ij} from the adjacency matrix $\mathbf{M} = \{b_{ij}\}_{m \times m}$ such that $b_{ij} = 1$, if there exists an edge between w_i and w_j; 0 otherwise, is also applied in order to ignore the attention for any two words if there is no edge between them (i.e., $b_{ij} = 0$). Afterwards, we calculate the mean of the hidden representation $h_i^{(L)}$ to obtain the final representation, $H^D \in R^{d_h}$ of the DGCN layer:

$$H^D = \frac{1}{m} \sum_{i=1}^{m} h_i^{(L)} \tag{10}$$

Aggregation Layer. This layer combines the semantic and syntactic features obtained using the Caps and DGCN layers, respectively. Specifically, we apply the average pooling with a kernel size of 3 on the final hidden representation of the Caps and DGCN layers and then aggregate both of these representations into a vector $O \in R^d$. Formally,

$$\begin{aligned} H_p^C &= AvgPool(H^C) \\ H_p^D &= AvgPool(H^D) \end{aligned} \tag{11}$$

$$O = [H_p^C; H_p^D] \tag{12}$$

Output Layer. After obtaining the representation O, it is fed into task-specific fully connected layer and then a *Softmax* layer to generate a probability distribution over the politeness classes:

$$\hat{y}^P = Softmax(W^P O + b^P) \tag{13}$$

Similarly, the representation O is fed into another task-specific fully connected layer followed by a *Sigmoid* layer to generate the probability distribution over the emotion classes for multi-label emotion classification in POEM.

$$\hat{y}^E = Sigmoid(W^E O + b^E) \tag{14}$$

where, W^P, W^E, b^P, and b^E denote the task-specific weights and biases, respectively. Note: We use the $Softmax$ for emotion classification in DailyDialog.

4.2.1. Loss functions and multi-task objective

For the politeness classification task which is a multi-class classification problem, we define the objective function with weighted cross-entropy loss as described in Eq. (15):

$$L_P = -\sum_{i}^{C} w_i [y_i^P \log(\hat{y}_i^P)] \tag{15}$$

where, C denotes the number of politeness classes.

For the multi-label emotion detection task in POEM, we define the objective function with weighted binary cross-entropy with logits loss as described in Eq. (16):

$$L_E = -\sum_{i}^{C} w_i [y_i^E \cdot \log \sigma(\hat{y}_i^E) + (1 - y_i^E) \cdot \log(1 - \sigma(\hat{y}_i^E))] \tag{16}$$

For the multi-class emotion detection task in DailyDialog, we define the objective function with weighted cross-entropy loss as described in Eq. (17):

$$L_E = -\sum_{i}^{C} w_i [y_i^E \log(\hat{y}_i^E)] \tag{17}$$

where, C denotes the number of emotion classes.

In order to optimize the network in a unified framework, we simply add the two loss functions together as the joint loss function as shown in Eq. (18) and fine-tune the model in an end-to-end fashion via minimizing the joint loss.

$$L_{Multi} = L_P + L_E \tag{18}$$

5. Experiments and results

5.1. Implementation details

We have used PyTorch[15] framework for implementation purposes. We use the uncased BERT-Base model[16] pre-trained on the English language using a masked language modelling (MLM) objective under their default settings. For getting the dependency tree for each utterance to create its DGCN graph, an off-the-shelf system, namely Stanford CoreNLP Toolkit[17] (version 3.9.2) is used. This popular toolkit has been extensively used in various previous studies (Huang & Carley, 2019; Sun, Zhang, Mensah, Mao, & Liu, 2019; Tian et al., 2020). The dimension of each output capsule is set to 128.

We randomly initialized all the trainable parameters in our proposed model and all the hyper-parameters of the model are fine-tuned on the validation set. We use 1 layer of DGCN with BERT-Base. We use the maximum sequence length of 180 for POEM and 300 for DailyDialog and a batch size of 32. Adam (Kingma & Ba, 2014) optimizer is used with an initial learning rate of $3e^{-4}$ and Adam's epsilon of $1e^{-8}$. To avoid over-fitting, a useful regularization technique known as dropout (Srivastava, Hinton, Krizhevsky, Sutskever, & Salakhutdinov, 2014) with probability 0.5 is used and the model is trained for 10 epochs.

For POEM, we have used weighted Accuracy (Acc) and Macro-F1 (F1) score as the evaluation metrics for the politeness detection task and subset accuracy (S-Acc), Micro-F1, Jaccard Index (JI), and Hamming Loss (HL) for multi-label emotion detection. Similarly, for DailyDialog, we have used weighted Accuracy and Macro-F1 score as the evaluation metrics for both tasks. These evaluation metrics are chosen to account for the imbalanced class distribution.

5.2. Baselines

We perform a comparative study against both the single-task and multi-task learning models to demonstrate the effectiveness of our proposed approach. In particular, we compare our obtained results with the following baseline models:

1. **Single-task Learning:** All the models under this group perform politeness detection and emotion detection tasks independently, i.e. one model is responsible for the politeness detection task, while the other model is responsible for the emotion detection task, and both the models are trained and evaluated separately.
   (a) ***BERT*:** This is an uncased BERT-base model fine-tuned separately for politeness detection and emotion detection tasks. Given an input utterance, $U_t = (w_1, w_2, \dots, w_m)$, the *BERT* model encodes the utterance and outputs a global aggregated representation of the input utterance, represented by $H^{[CLS]} \in \mathbb{R}^{d_h}$ and a sequence of hidden representations, $H^E = (h_1^E, h_2^E, \dots, h_m^E)$, where $H^E \in \mathbb{R}^{m \times d_h}$. The global representation of the utterance, $H^{[CLS]}$ is fed into a fully connected layer. Finally, a *Softmax* layer is used to get the predicted politeness class. We used *Sigmoid* layer to get the predicted emotion categories for the utterances in the POEM dataset. Since the DailyDialog dataset has a single emotion class for each utterance, a *Softmax* layer is used to get the predicted emotion category of the utterance.
   (b) ***Caps*:** We apply a Capsule Network (Caps) layer on the top of the BERT layer in the *BERT* model. The contextual word representation generated by the BERT layer (H^E) is fed into the Caps layer, which in turn, extracts intrinsic spatial and hierarchical information from H^E that would be informative for politeness or emotion detection task. The joint optimization of features extracted from the BERT and Caps layers enables the learning of advanced semantic information and contextual information of utterances, which leads to performance improvement for the task.
   (c) ***DGCN*:** This baseline model applies the DGCN layer on top of the BERT layer in the *BERT* model. The DGCN layer models the dependency relations among the words by considering their position and direction with their related words, which captures useful syntactic information relevant to the politeness or emotion detection task. The DGCN model tends to learn traits and attributes by leveraging from the joint optimization of features (syntactic and contextual information of utterances) from the BERT and DGCN layers to develop a robust classifier for the task.
   (d) ***DialogueRNN*** (Majumder, Poria, Hazarika, et al., 2019)**:** This recurrent neural network (RNN) based model keeps track of individual interlocutor states throughout the conversation for emotion recognition in conversations (ERC).
   (e) ***DialogueGCN*** (Ghosal, Majumder, Poria, Chhaya, & Gelbukh, 2019)**:** DialogueGCN presented a dialogue graph convolutional network to model the conversational context for emotion identification by leveraging self and inter-speaker dependency of the interlocutors.

[15] https://pytorch.org/

[16] https://github.com/huggingface/transformers

[17] https://stanfordnlp.github.io/CoreNLP/

Table 5
Experimental results for politeness and emotion detection on POEM dataset.

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

2. **Multi-task Learning:** Models in this group solve more than one task, i.e., a single model performs both tasks simultaneously. All the above-mentioned single-task baselines, i.e., *BERT*, *Caps*, *DGCN*, *DialogueRNN* and *DialogueGCN* are fine-tuned for multi-task learning of politeness classification and emotion detection tasks. For multi-task learning, the politeness and emotion detection objectives are added by giving equal weights to these objectives and then combined loss is backpropagated to predict the politeness and emotion categories of the utterances. The reason for assigning equal weights is that we want our MTL model to handle both tasks with equal importance. Comprising one task can affect the other task as both are closely related and significant for a dialogue system.
   Besides, we compared our proposed framework with the following MTL models:
   (a) ***DCR-Net*** (Qin et al., 2020)**:** A Deep Co-Interactive Relation Network (DCR-Net) employed a hierarchical encoder to model the contextual information, which is followed by a relation layer to explicitly model the interaction between the two tasks.
   (b) ***co-GAT*** (Qin et al., 2021)**:** A Co-Interactive Graph Attention Network (co-GAT) utilized a co-interactive graph attention layer in which cross-utterances connection and cross-tasks connection are created and updated iteratively with each other to simultaneously consider contextual information and mutual interaction information for joint modelling of the tasks.

5.3. Evaluation results

Results of the different models for POEM and DailyDialog datasets are presented in Tables 5 and 6, respectively. In the experiment, we run our model in both single-task and multi-task settings. We find that the proposed multi-task model achieves 90.30% accuracy (4.93 points ↑ in comparison single-task Caps-DGCN model) and 87.19% F1-score (2.17 points ↑ in comparison single-task Caps-DGCN model) for the politeness detection task on POEM. For the emotion detection task on POEM, the multi-task Caps-DGCN achieves 58.72% subset accuracy (7.77 points ↑ in comparison to single-task Caps-DGCN model) and Jaccard Index of 0.63 (0.11 points ↑ in comparison single-task Caps-DGCN model) on POEM. Similarly, for the politeness detection task on DailyDialog, we observed that the proposed framework obtains 86.78% accuracy and 75.27% F1-score (4.74 and 3.94 points ↑ in comparison single-task Caps-DGCN model, respectively). We obtain an accuracy of 75.49% and F1-score of 46.08% (1.97 and 2.58 points ↑ in comparison single-task Caps-DGCN model, respectively) for the emotion detection task on DailyDialog. We observe that the sharing of information between the tasks helps in improving the performance of the proposed multi-task framework. Further, it can also be observed that our proposed MTL framework outperforms the other MTL baselines in terms of all evaluation metrics on both datasets. This shows that our approach has effectively incorporated both semantic and syntactic information which eventually helped in boosting the performance of the model on both tasks.

5.4. Statistical significance test

We have carried out a statistical significance test known as Welch's t-test (Welch, 1947) at the 5% (0.05) significance level to determine whether the improvement in our proposed model is statistically significant or not. The test is conducted to demonstrate that the best accuracy achieved by the proposed approach is not a coincidence, rather it is statistically significant. The evaluation metric (accuracy) is obtained by 20 successive iterations of each algorithm for the statistical test on both datasets. We have calculated the p-values provided by Welch's t-test for comparing two groups in order to determine the statistical significance of our approach. The first group (α) represents the accuracies obtained by our best model (i.e. Caps-DGCN), and the other group (β) represents the accuracies achieved by the other baseline and the suggested Caps-DGCN model. The t-test is a null hypothesis (H_0) test that is used to examine if there is a significant difference between two sets of data or not.

$$H_0 : \lambda_\alpha = \lambda_\beta \tag{19}$$

In contrast, the alternative hypothesis (H_1) asserts that there are significant differences between the average accuracies achieved by any of the two groups.

$$H_0 : \lambda_\alpha > \lambda_\beta \tag{20}$$

where, λ_k denotes the kth algorithm's average accuracy. In this experiment, we have considered alternative hypothesis (H_1) as that the proposed approach is significantly better than the other baseline methods. Therefore, the alternative hypothesis (H_1) is established by disproving the null hypothesis (H_0).

We have calculated the means of the two groups and found that the mean of the proposed approach is always greater than the means of the baselines. Thus, we performed the one-sided t-test. Consequently, there are only two possible outcomes: either the proposed model is superior to the baseline model, or it is the same. Therefore, we have determined that the alternative hypothesis is superior as compared to

Table 6
Experimental results for politeness and emotion detection on DailyDialog dataset.

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

Table 7
Results of statistical significance test on POEM dataset.

Learning paradigm	Models	Politeness				Emotion			
		Acc	P	R	F1	S-Acc	Micro-F1	JI	HL
	Baselines								
Single-task learning	*BERT*	1.9E−060	2.04E−062	1.77E−057	6.04E−062	1.49E−071	5.33E−077	2.16E−058	2.82E−062
	Caps	2.20E−037	4.76E−044	4.78E−052	1.09E−048	2.45E−051	2.68E−036	1.78E−043	4.42E−031
	DGCN	9.05E−044	6.87E−047	6.20E−054	2.53E−025	1.11E−031	4.66E−053	8.21E−030	7.05E−34
	DialogueRNN	3.38E−012	4.47E−044	7.05E−034	1.66E−041	3.93E−041	1.81E−053	2.86E−061	1.73E−055
	DialogueGCN	1.48E−070	7.06E−035	2.67E−035	1.34E−039	1.90E-O50	6.03E−066	4.78E−062	2.45E−066
	Caps-DGCN	2.30E−048	5.79E−063	1.81E−054	1.90E−040	3.85E−029	1.09E−065	1.52E−054	1.04E−060
Multi-task learning	*BERT*	4.66E−053	1.78E−043	3.22E−012	1.34E−038	5.97E−042	1.83E−062	2.43E−064	1.79E−048
	Caps	1.08E−050	1.09E−048	4.42E−031	3.34E−038	7.05E−034	2.55E−074	2.61E−071	1.17E−060
	DGCN	2.82E−062	2.21E−053	1.47E−026	2.68E−036	2.31E−039	6.48E−066	4.14E−067	4.78E−052
	DialogueRNN	8.21E−030	1.90E−060	2.21E−053	3.95E−044	3.28E−040	2.62E−061	4.65E−053	1.85E−035
	DialogueGCN	9.58E−027	8.49E−036	5.41E−022	6.58E−046	7.65E−037	1.27E−042	8.49E−033	1.48E−022
	DCR-Net	2.20E−032	3.50E−042	1.90E-O42	9.03E−041	2.30E−044	4.72E−015	4.12E−041	1.08E−055
	Co-GAT	4.88E−013	6.87E−037	5.42E-O35	2.56E−062	1.99E−034	8.54E−026	4.43E−041	5.48E−022

the baseline, but the null hypothesis is equivalent to the baseline. Thus, both hypotheses are mutually exclusive and complete with regard to the circumstance under consideration. Using the following t- statistic formula, we can now calculate the difference between the average accuracies as:

$$t = \frac{\bar{\mathcal{X}}_1 - \bar{\mathcal{X}}_2}{\sqrt{\frac{\sigma_1^2}{n_1} + \frac{\sigma_2^2}{n_2}}} \tag{21}$$

where, $\bar{\mathcal{X}}_i, \sigma_i$, and n_i are the mean, variance and size of the ith sample, respectively. p-value is the probability, under the assumption of the null hypothesis (H_0) and the smaller p-value represents strong evidence against null hypothesis (H_0). We have executed the experiment 20 times for the statistical test on both datasets. Tables 7 and 8 present the p-values produced by the Welch's t-test on POEM and DailyDialog datasets, respectively. All the p-values reported in Tables 7 and 8 are less than 0.05 (5% significance level), which demonstrate that our proposed approach is statistically significant.

On the other hand, the two-sided t-test can have three possible outcomes: (i) the mean of the proposed approach is equal to the baselines, (ii) the mean of the proposed approach is less than the baselines, and (iii) the mean of the proposed approach is greater than the baselines. We notice the third criterion, which prompts us to conduct a one-sided t-test for establishing the statistical significance of our proposed approach. The second and the third criteria are complementary in our case i.e. if the second criterion is true it implies that the third criterion is false and vice versa. Thus, in this regard, proving our alternative hypothesis (i.e. proposed approach is better than the baseline approaches) is sufficient to prove the third condition and eliminate the second condition.

6. Result analysis

6.1. Ablation analysis

We perform the ablation study of the proposed Caps-DGCN. We run the model with or without DGCN and observe that the Caps with DGCN are performing significantly better than the Caps without the DGCN layer. The Caps-DGCN can effectively capture the semantic and syntactic information of the utterances, thereby enhancing the performance of the model. Moreover, we also carry out the ablation of the direction modelling component in the DGCN layer and find that the direction modelling in the DGCN layer separately models the different contextual features of a word by considering their direction to the word and further weighing these features using the attention mechanism, which enhances the model capability of capturing syntactic information, thereby improving the overall performance of the model. The ablation study of our proposed model is reported in Table 9.

6.2. Qualitative analysis

We perform a comprehensive qualitative analysis of the outputs generated from our proposed multi-task model for politeness and emotion detection tasks over the single-task Caps-DGCN model for both datasets. Tables 10 and 11 present a few utterances with their corresponding politeness and emotion labels predicted by the single-task and multi-task

Table 8
Results of statistical significance test on DailyDialog dataset.

Learning paradigm	Models	Politeness				Emotion			
		Acc	P	R	F1	Acc	P	R	F1
Single-task learning	Baselines								
	BERT	3.93E−041	5.29E−049	3.27E−050	2.04E−049	2.62E−050	9.74E−063	4.16E−049	1.90E−040
	Caps	5.79E−063	2.31E−039	1.34E−038	3.93E−041	4.42E−031	1.34E−038	5.97E−042	2.21E−053
	DGCN	9.99E−058	9.73E−045	1.51E−039	4.35E−054	3.51E−043	1.78E−043	1.46E−070	1.47E−026
	DialogueRNN	5.33E−016	2.16E−044	3.50E−010	4.78E−052	3.27E−060	1.47E−028	3.31E−062	2.45E−038
	DialogueGCN	9.05E−011	3.22E−018	7.05E−029	5.29E−041	2.53E−038	9.74E−011	1.09E−060	4.42E−055
	Caps-DGCN	1.27E−043	1.52E−032	1.32E−060	1.90E−040	4.78E−052	7.91E−063	6.87E−047	3.34E−038
Multi-task learning	*BERT*	2.01E−042	2.50E−044	8.66E−065	7.65E−036	6.73E−049	7.07E−051	7.94E−068	4.76E−044
	Caps	1.79E−047	7.17E−020	2.86E−045	2.45E−051	6.20E−054	4.24E−057	1.34E−038	8.21E−030
	DGCN	1.81E−054	8.03E−033	2.83E−063	3.21E−052	4.42E−031	2.53E−025	1.11E−031	6.87E−047
	DialogueRNN	9.66E−032	1.23E−055	8.72E−016	4.88E−032	9.65E−091	6.57E−032	5.23E−041	1.27E−033
	DialogueGCN	4.87E−014	5.42E−033	8.49E−037	2.23E−052	1.34E−039	3.81E−022	1.75E−044	3.23E−015
	DCR-Net	2.25E−088	1.42E−025	5.96E−033	1.34E−013	1.84E−032	2.46E−045	4.77E−028	2.30E−050
	Co-GAT	3.32E−066	1.77E−058	2.04E−063	6.09E−051	1.11E−032	1.09E−015	2.04E−023	2.31E−040

Table 9
Ablation study of proposed model demonstrating the effect ablating the DGCN layer and direction modelling component in DGCN layer. ATT and DIR represent the attention mechanism and direction modelling, respectively.

Models	POEM				DailyDialog			
	Politeness		Emotion		Politeness		Emotion	
	Acc	F1	S-Acc	JI	Acc	F1	Acc	F1
Caps+GCN+ATT+DIR (Caps-DGCN)	**90.30**	**87.19**	**58.72**	**0.63**	**86.78**	**75.27**	**75.49**	**46.08**
Caps+GCN+ATT	89.45	86.79	57.48	0.61	85.23	73.01	74.67	44.32
Caps+GCN	87.10	85.27	53.68	0.57	83.51	72.64	73.45	43.00
Caps	86.22	85.14	51.91	0.51	82.96	72.74	73.20	42.94
BERT	85.01	84.22	50.73	0.46	81.23	71.27	72.76	41.17

Table 10
Few utterances with corresponding politeness and emotion labels predicted by Caps-DGCNST and Caps-DGCN^{MT} for POEM. Here, ST and MT represent single-task and multi-task frameworks, respectively.

Input		Gold	Predicted	
			Caps-DGCNST	Caps-DGCN^{MT}
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

Caps-DGCN models for POEM and DailyDialog datasets, respectively. For instance, for the utterance "*what can I write in the application, dont call me dear*" mentioned in Table 10, the single-task Caps-DGCN model predicts the emotion label as "anticipation" which is partially correct and politeness label as "polite", which is incorrect. One possible explanation could be the presence of the word "*dear*" in the utterance, which often occurs when the interlocutor's behaviour is polite. On the contrary, the multi-task model Caps-DGCN model, correctly predicts the emotions "anticipation, annoyed" and politeness label of the utterance as "impolite". The sharing of information between the tasks helps the Caps-DGCN model in making the correct predictions.

The proposed approach has performed reasonably well for both tasks; however, there have been a few errors. Some of the common forms of mistakes include:

1. **Lack of contextual information:** Due to the absence of contextual information from the previous utterances, our model sometimes fails to correctly predict the politeness and emotion labels of the target utterance. For instance, *yeah sure* — Gold: neutral (politeness) and confident (emotion). Predicted: polite (politeness) and joy (emotion).
2. **Inconsistent sentence structure:** Writing style of victims does not follow standard grammatical rules for sentences and is often informal. As a result, sometimes it becomes difficult for the model to correctly understand the syntactic and semantic information. For instance, *i need help my mother. you can?*, Gold: neutral (politeness) and hopeful (emotion). Predicted: neutral (politeness) and neutral (emotion).

7. Conclusion and future work

In this paper, we have proposed an end-to-end multi-task learning framework that simultaneously identifies the politeness and emotion labels of the utterances in the dialogue. For this work, we have prepared a novel politeness and emotion annotated Wizard-of-Oz dialogue dataset, POEM for mental health counselling and legal assistance for crime victims. Experimental results on both the POEM and DailyDialog datasets show that our proposed multi-task model outperforms the single-task models modelling politeness and emotion detection tasks separately, depicting the efficacy of exploiting the relationships between the two tasks.

Table 11
Few utterances with corresponding politeness and emotion labels predicted by Caps-DGCNST and Caps-DGCN^{MT} for DailyDialog. Here, ST and represent Single-task and Multi-task frameworks, respectively.

Input		Gold	Predicted	
			Caps-DGCNST	Caps-DGCN^{MT}
everything is fine how are you	Politeness	polite	polite	polite
	Emotion	neutral	happiness	neutral
Are you more of a leader or a follower?	Politeness	neutral	impolite	neutral
	Emotion	neutral	neutral	neutral
oh okay can i do that right now	Politeness	polite	impolite	polite
	Emotion	surprise	neutral	surprise

In future, we would like to investigate the association between politeness and emotion for other domains and languages like Hindi. We would also like to see the effect of incorporating the contextual information from the previous utterances of the dialogue for the PED task.

CRediT authorship contribution statement

Priyanshu Priya: Conceptualization, Data curation, Methodology, Writing – original draft, Implementation, Evaluation. **Mauajama Firdaus:** Conceptualization, Data curation, Validation, Writing – original draft. **Asif Ekbal:** Visualization, Investigation, Supervision, Writing – review & editing.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Priyanshu Priya reports financial support was provided by Innovation in Science Pursuit for Inspired Research (INSPIRE) Fellowship, Department of Science and Technology, Ministry of Science and Technology, Government of India. Asif Ekbal reports financial support was provided by Young Faculty Research Fellowship (YFRF), Visvesvaraya Ph.D. scheme for Electronics and IT, Ministry of Electronics and Information Technology (MeitY), Government of India.

Data availability

Data will be made available on request.

Acknowledgements

Priyanshu Priya acknowledges the 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 for financial support. Asif Ekbal acknowledges the Young Faculty Research Fellowship (YFRF), supported by Visvesvaraya Ph.D. scheme for Electronics and IT, Ministry of Electronics and Information Technology (MeitY), Government of India, being implemented by Digital India Corporation (formerly Media Lab Asia). All the authors gratefully also acknowledge the partial support from "IITP Centre of Excellence in Cyber Crime Prevention against Women and Children- AI based Tools for Women and Children Safety", sponsored by Ministry of Home Affairs, Govt. of India. We would also like to express our gratitude to Dr. Jyotsna Agrawal, Associate Professor, Department of Clinical Psychology, National Institute of Mental Health and Neuro Sciences (NIMHANS), Bengaluru for their support and guidance during the work.

Ethical consideration

The POEM dataset introduced in this paper is used only for the purpose of academic research. There is nothing to disclose that warrant the ethical issues.

References

Acosta, J. C. (2009). *Using emotion to gain rapport in a spoken dialog system*.

Acosta, J. C., & Ward, N. G. (2011). Achieving rapport with turn-by-turn, user-responsive emotional coloring. *Speech Communication*, *53*, 1137–1148.

Althoff, T., Clark, K., & Leskovec, J. (2016). Large-scale analysis of counseling conversations: An application of natural language processing to mental health. *Transactions of the Association for Computational Linguistics*, *4*, 463–476.

Aubakirova, M., & Bansal, M. (2016). Interpreting neural networks to improve politeness comprehension. arXiv preprint arXiv:1610.02683.

Bao, J., Wu, J., Zhang, Y., Chandrasekharan, E., & Jurgens, D. (2021). Conversations gone alright: Quantifying and predicting prosocial outcomes in online conversations. In *Proceedings of the web conference 2021* (pp. 1134–1145).

Bickmore, T. W., & Picard, R. W. (2004). Towards caring machines. In *CHI'04 extended abstracts on human factors in computing systems* (pp. 1489–1492).

Bothe, C. (2021). Polite emotional dialogue acts for conversational analysis in dialy dialog data. arXiv preprint arXiv:2112.13572.

Bothe, C., Magg, S., Weber, C., & Wermter, S. (2018). Discourse-wizard: Discovering deep discourse structure in your conversation with RNNs. arXiv preprint arXiv:1806.11420.

Brown, P., Levinson, S. C., & Levinson, S. C. (1987). *Politeness: Some universals in language usage, volume 4*. Cambridge University Press.

Budzianowski, P., Wen, T. -H., Tseng, B. -H., Casanueva, I., Ultes, S., Ramadan, O., et al. (2018). Multiwoz–a large-scale multi-domain wizard-of-oz dataset for task-oriented dialogue modelling. arXiv preprint arXiv:1810.00278.

Byon, A. S. (2006). *The role of linguistic indirectness and honorifics in achieving linguistic politeness in Korean requests*. Walter de Gruyter.

Callejas, Z., Griol, D., & López-Cózar, R. (2011). Predicting user mental states in spoken dialogue systems. *EURASIP Journal on Advances in Signal Processing*, *2011*, 1–21.

Castonguay, L. G., & Hill, C. E. (2017). *How and why are some therapists better than others?: understanding therapist effects*. JSTOR.

Cerisara, C., Jafaritazehjani, S., Oluokun, A., & Le, H. (2018). Multi-task dialog act and sentiment recognition on mastodon. arXiv preprint arXiv:1807.05013.

Chen, Y., Hou, W., Cheng, X., & Li, S. (2018). Joint learning for emotion classification and emotion cause detection. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 646–651).

Chen, G., Tian, Y., & Song, Y. (2020). Joint aspect extraction and sentiment analysis with directional graph convolutional networks. In *Proceedings of the 28th international conference on computational linguistics* (pp. 272–279).

Culpeper, J., & Tantucci, V. (2021). The principle of (im) politeness reciprocity. *Journal of Pragmatics*, *175*, 146–164.

Danescu-Niculescu-Mizil, C., Sudhof, M., Jurafsky, D., Leskovec, J., & Potts, C. (2013). A computational approach to politeness with application to social factors. arXiv preprint arXiv:1306.6078.

De Choudhury, M., Sharma, S. S., Logar, T., Eekhout, W., & Nielsen, R. C. (2017). Gender and cross-cultural differences in social media disclosures of mental illness. In *Proceedings of the 2017 ACM conference on computer supported cooperative work and social computing* (pp. 353–369).

Devlin, J., Chang, M. -W., Lee, K., & Toutanova, K. (2018). Bert: Pre-training of deep bidirectional transformers for language understanding. arXiv preprint arXiv:1810.04805.

Dheeraj, K., & Ramakrishnudu, T. (2021). Negative emotions detection on online mental-health related patients texts using the deep learning with MHA-BCNN model. *Expert Systems with Applications*, *182*, Article 115265.

Do, P. -K., Nguyen, H. -T., Tran, C. -X., Nguyen, M. -T., & Nguyen, M. -L. (2017). Legal question answering using ranking SVM and deep convolutional neural network. arXiv preprint arXiv:1703.05320.

Dou, Z. -Y., Tu, Z., Wang, X., Wang, L., Shi, S., & Zhang, T. (2019). Dynamic layer aggregation for neural machine translation with routing-by-agreement. In *Proceedings of the AAAI conference on artificial intelligence: Vol. 33*, (pp. 86–93).

Dowling, M., & Rickwood, D. (2014). Investigating individual online synchronous chat counselling processes and treatment outcomes for young people. *Advances in Mental Health*, *12*, 216–224.

Dowling, M., & Rickwood, D. (2016). Exploring hope and expectations in the youth mental health online counselling environment. *Computers in Human Behavior*, *55*, 62–68.

Du, C., Sun, H., Wang, J., Qi, Q., Liao, J., Wang, C., et al. (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* (pp. 456–465).

Du, C., Sun, H., Wang, J., Qi, Q., Liao, J., Xu, T., et al. (2019). Capsule network with interactive attention for aspect-level sentiment classification. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing* (pp. 5489–5498).

Feng, S., Lubis, N., Geishauser, C., Lin, H. -c., Heck, M., van Niekerk, C., et al. (2021). Emowoz: A large-scale corpus and labelling scheme for emotion in task-oriented dialogue systems. arXiv preprint arXiv:2109.04919.

Firdaus, M., Ekbal, A., & Bhattacharyya, P. (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* (pp. 4172–4182).

Fitzpatrick, K. K., Darcy, A., & Vierhile, M. (2017). Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): A randomized controlled trial. *JMIR Mental Health*, *4*, Article e7785.

Fleiss, J. L. (1971). Measuring nominal scale agreement among many raters. *Psychological Bulletin*, *76*, 378.

Fulmer, R., Joerin, A., Gentile, B., Lakerink, L., Rauws, M., et al. (2018). Using psychological artificial intelligence (tess) to relieve symptoms of depression and anxiety: Randomized controlled trial. *JMIR Mental Health*, *5*, Article e9782.

Gan, C., Yang, Y., Zhu, Q., Jain, D. K., & Struc, V. (2022). DHF-net: A hierarchical feature interactive fusion network for dialogue emotion recognition. *Expert Systems with Applications*, Article 118525.

Gaur, M., Kursuncu, U., Alambo, A., Sheth, A., Daniulaityte, R., Thirunarayan, K., et al. (2018). "Let me tell you about your mental health!" contextualized classification of reddit posts to DSM-5 for web-based intervention. In *Proceedings of the 27th ACM international conference on information and knowledge management* (pp. 753–762).

Ghosal, D., Majumder, N., Poria, S., Chhaya, N., & Gelbukh, A. (2019). DialogueGCN: A graph convolutional neural network for emotion recognition in conversation. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing* (pp. 154–164).

Gibson, J., Can, D., Xiao, B., Imel, Z. E., Atkins, D. C., Georgiou, P., et al. (2016). A deep learning approach to modeling empathy in addiction counseling. *Commitment*, *111*, 21.

Golchha, H., Firdaus, M., Ekbal, A., & Bhattacharyya, P. (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: Vol. 1* (pp. 851–860).

Gong, J., Qiu, X., Wang, S., & Huang, X. (2018). Information aggregation via dynamic routing for sequence encoding. arXiv preprint arXiv:1806.01501.

Gratch, J., Artstein, R., Lucas, G., Stratou, G., Scherer, S., Nazarian, A., et al. (2014). *The distress analysis interview corpus of human and computer interviews: Technical report*, UNIVERSITY OF SOUTHERN CALIFORNIA LOS ANGELES.

Grice, H. P. (1975). Logic and conversation. In *Speech acts* (pp. 41–58). Brill.

Gupta, S., Walker, M. A., & Romano, D. M. (2007). How rude are you?: Evaluating politeness and affect in interaction. In *International conference on affective computing and intelligent interaction* (pp. 203–217). Springer.

Gururaj, G., Varghese, M., Benegal, V., Rao, G., Pathak, K., Singh, L., et al. (2016). *National mental health survey of India, 2015-16: Mental health systems*. Bengaluru, India: National Institute of Mental Health and Neuro Sciences. NIMHANS Publ.

Hendrycks, D., Burns, C., Chen, A., & Ball, S. (2021). Cuad: An expert-annotated nlp dataset for legal contract review. arXiv preprint arXiv:2103.06268.

Howes, C., Purver, M., & McCabe, R. (2014). Linguistic indicators of severity and progress in online text-based therapy for depression. In *Proceedings of the workshop on computational linguistics and clinical psychology: from linguistic signal to clinical reality*. Association for Computational Linguistics.

Huang, B., & Carley, K. M. (2019). Syntax-aware aspect level sentiment classification with graph attention networks. arXiv preprint arXiv:1909.02606.

Inkster, B., Sarda, S., Subramanian, V., et al. (2018). An empathy-driven, conversational artificial intelligence agent (wysa) for digital mental well-being: Real-world data evaluation mixed-methods study. *JMIR mHealth and uHealth*, *6*, Article e12106.

John, A. K., Caro, L. D., Robaldo, L., & Boella, G. (2017). Legalbot: A deep learning-based conversational agent in the legal domain. In *International conference on applications of natural language to information systems* (pp. 267–273). Springer.

Kapoor, A., Dhawan, M., Goel, A., Arjun, T., Bhatnagar, A., Agrawal, V., et al. (2022). HLDC: Hindi legal documents corpus. arXiv preprint arXiv:2204.00806.

Kelley, J. F. (1984). An iterative design methodology for user-friendly natural language office information applications. *ACM Transactions on Information Systems (TOIS)*, *2*, 26–41.

Khoshnam, F., & Baraani-Dastjerdi, A. (2022). A dual framework for implicit and explicit emotion recognition: An ensemble of language models and computational linguistics. *Expert Systems with Applications*, *198*, Article 116686.

Kim, J., Kim, Y., Kim, B., Yun, S., Kim, M., & Lee, J. (2018). Can a machine tend to teenagers' emotional needs? A study with conversational agents. In *Extended abstracts of the 2018 CHI conference on human factors in computing systems* (pp. 1–6).

Kingma, D. P., & Ba, J. (2014). Adam: A method for stochastic optimization. arXiv preprint arXiv:1412.6980.

Kowsrihawat, K., Vateekul, P., & Boonkwan, P. (2018). Predicting judicial decisions of criminal cases from thai supreme court using bi-directional GRU with attention mechanism. In *2018 5th Asian conference on defense technology* (pp. 50–55). IEEE.

Lakoff, R. (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*, (pp. 292–305). Chicago Linguistic Society.

Langlotz, A., & Locher, M. A. (2017). (Im) politeness and emotion. In *The palgrave handbook of linguistic (im) politeness* (pp. 287–322). Springer.

Lee, F. -T., Hull, D., Levine, J., Ray, B., & McKeown, K. (2019). Identifying therapist conversational actions across diverse psychotherapeutic approaches. In *Proceedings of the sixth workshop on computational linguistics and clinical psychology* (pp. 12–23).

Li, C., Braud, C., & Amblard, M. (2022). Multi-task learning for depression detection in dialogs. arXiv preprint arXiv:2208.10250.

Li, Y., Kazameini, A., Mehta, Y., & Cambria, E. (2021). Multitask learning for emotion and personality detection. arXiv preprint arXiv:2101.02346.

Li, Y., Su, H., Shen, X., Li, W., Cao, Z., & Niu, S. (2017). Dailydialog: A manually labelled multi-turn dialogue dataset. arXiv preprint arXiv:1710.03957.

Li, J., Zhang, M., Ji, D., & Liu, Y. (2020). Multi-task learning with auxiliary speaker identification for conversational emotion recognition. arXiv preprint arXiv:2003.01478.

Liang, Y., Meng, F., Zhang, J., Chen, Y., Xu, J., & Zhou, J. (2020). An iterative multi-knowledge transfer network for aspect-based sentiment analysis. arXiv preprint arXiv:2004.01935.

Liang, Y., Meng, F., Zhang, J., Chen, Y., Xu, J., & Zhou, J. (2021). A dependency syntactic knowledge augmented interactive architecture for end-to-end aspect-based sentiment analysis. *Neurocomputing*, *454*, 291–302.

Lin, T., Sun, A., & Wang, Y. (2022). EDU-capsule: Aspect-based sentiment analysis at clause level. *Knowledge and Information Systems*, 1–25.

Lucas, G. M., Gratch, J., King, A., & Morency, L. -P. (2014). It's only a computer: Virtual humans increase willingness to disclose. *Computers in Human Behavior*, *37*, 94–100.

Madaan, A., Setlur, A., Parekh, T., Poczos, B., Neubig, G., Yang, Y., et al. (2020). Politeness transfer: A tag and generate approach. arXiv preprint arXiv:2004.14257.

Majumder, N., Poria, S., Hazarika, D., Mihalcea, R., Gelbukh, A., & Cambria, E. (2019). Dialoguernn: An attentive rnn for emotion detection in conversations. *Vol. 33*, In *Proceedings of the AAAI conference on artificial intelligence* (pp. 6818–6825).

Majumder, N., Poria, S., Peng, H., Chhaya, N., Cambria, E., & Gelbukh, A. (2019). Sentiment and sarcasm classification with multitask learning. *IEEE Intelligent Systems*, *34*, 38–43.

Malhotra, G., Waheed, A., Srivastava, A., Akhtar, M. S., & Chakraborty, T. (2022). Speaker and time-aware joint contextual learning for dialogue-act classification in counselling conversations. In *Proceedings of the fifteenth ACM international conference on web search and data mining* (pp. 735–745).

Marreddy, M., Oota, S. R., Vakada, L. S., Chinni, V. C., & Mamidi, R. (2022). Multi-task text classification using graph convolutional networks for large-scale low resource language. arXiv preprint arXiv:2205.01204.

Matsumoto, Y. (1988). Reexamination of the universality of face: Politeness phenomena in Japanese. *Journal of Pragmatics*, *12*, 403–426.

Matthews, M., & Doherty, G. (2011). In the mood: Engaging teenagers in psychotherapy using mobile phones. In *Proceedings of the SIGCHI conference on human factors in computing systems* (pp. 2947–2956).

McElvain, G., Sanchez, G., Matthews, S., Teo, D., Pompili, F., & Custis, T. (2019). Westsearch plus: A non-factoid question-answering system for the legal domain. In *Proceedings of the 42nd international ACM SIGIR conference on research and development in information retrieval* (pp. 1361–1364).

Mishra, K., Firdaus, M., & Ekbal, A. (2022a). Please be polite: Towards building a politeness adaptive dialogue system for goal-oriented conversations. *Neurocomputing*, *494*, 242–254.

Mishra, K., Firdaus, M., & Ekbal, A. (2022b). Predicting politeness variations in goal-oriented conversations. *IEEE Transactions on Computational Social Systems*.

Mohr, D. C., Burns, M. N., Schueller, S. M., Clarke, G., & Klinkman, M. (2013). Behavioral intervention technologies: Evidence review and recommendations for future research in mental health. *General Hospital Psychiatry*, *35*, 332–338.

Morris, R. R., Kouddous, K., Kshirsagar, R., & Schueller, S. M. (2018). Towards an artificially empathic conversational agent for mental health applications: System design and user perceptions. *Journal of Medical Internet Research*, *20*, Article e10148.

Newbold, J., Doherty, G., Rintel, S., & Thieme, A. (2019). *Politeness Strategies in the Design of Voice Agents for Mental Health*.

Niu, T., & Bansal, M. (2018). Polite dialogue generation without parallel data. *Transactions of the Association for Computational Linguistics*, *6*, 373–389.

Ohbyung, K., & Sukjae, C. (2009). Context-aware selection of politeness level for polite mobile service in Korea. *Expert Systems with Applications*, *36*, 4198–4206.

Perez-Gaspar, L. -A., Caballero-Morales, S. -O., & Trujillo-Romero, F. (2016). Multimodal emotion recognition with evolutionary computation for human-robot interaction. *Expert Systems with Applications*, *66*, 42–61.

Pérez-Rosas, V., Mihalcea, R., Resnicow, K., Singh, S., & An, L. (2017). Understanding and predicting empathic behavior in counseling therapy. In *Proceedings of the 55th annual meeting of the association for computational linguistics (Vol. 1: Long lowercasePapers)* (pp. 1426–1435).

Pérez-Rosas, V., Wu, X., Resnicow, K., & Mihalcea, R. (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* (pp. 926–935).

Peskov, D., Clarke, N., Krone, J., Fodor, B., Zhang, Y., Youssef, A., et al. (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* (pp. 4526–4536).

Pittermann, J., Pittermann, A., & Minker, W. (2010). Emotion recognition and adaptation in spoken dialogue systems. *International Journal of Speech Technology*, *13*, 49–60.

Pruksachatkun, Y., Pendse, S. R., & Sharma, A. (2019). Moments of change: Analyzing peer-based cognitive support in online mental health forums. In *Proceedings of the 2019 CHI conference on human factors in computing systems* (pp. 1–13).

Qin, L., Che, W., Li, Y., Ni, M., & Liu, T. (2020). Dcr-net: A deep co-interactive relation network for joint dialog act recognition and sentiment classification. In *Proceedings of the AAAI conference on artificial intelligence*: *Vol. 34*, (pp. 8665–8672).

Qin, L., Li, Z., Che, W., Ni, M., & Liu, T. (2021). Co-gat: A co-interactive graph attention network for joint dialog act recognition and sentiment classification. *Vol. 35*, In *Proceedings of the AAAI conference on artificial intelligence* (pp. 13709–13717).

Rashkin, H., Smith, E. M., Li, M., & Boureau, Y. -L. (2018). Towards empathetic open-domain conversation models: A new benchmark and dataset. arXiv preprint arXiv:1811.00207.

Reis, J. C., Correia, A., Murai, F., Veloso, A., & Benevenuto, F. (2019). Supervised learning for fake news detection. *IEEE Intelligent Systems*, *34*, 76–81.

Renner, B. (2020). *(In) directness as an (im) politeness strategy in the contact between German and Brazilian Portuguese as additional languages*.

Ruder, S. (2017). An overview of multi-task learning in deep neural networks. arXiv preprint arXiv:1706.05098.

Sabour, S., Frosst, N., & Hinton, G. E. (2017). Dynamic routing between capsules. *Advances in Neural Information Processing Systems*, *30*.

Saha, T., Chopra, S., Saha, S., Bhattacharyya, P., & Kumar, P. (2021). A large-scale dataset for motivational dialogue system: An application of natural language generation to mental health. In *2021 international joint conference on neural networks* (pp. 1–8). IEEE.

Saha, T., Gakhreja, V., Das, A. S., Chakraborty, S., & Saha, S. (2022). Towards motivational and empathetic response generation in online mental health support. In *Proceedings of the 45th international ACM SIGIR conference on research and development in information retrieval* (pp. 2650–2656).

Saha, T., Priya, N., Saha, S., & Bhattacharyya, P. (2021). A transformer based multi-task model for domain classification, intent detection and slot-filling. In *2021 international joint conference on neural networks* (pp. 1–8). IEEE.

Saha, T., Reddy, S., Das, A., Saha, S., & Bhattacharyya, P. (2022). A shoulder to cry on: Towards a motivational virtual assistant for assuaging mental agony. In *Proceedings of the 2022 conference of the North American chapter of the association for computational linguistics: Human language technologies* (pp. 2436–2449).

Saha, T., Reddy, S. M., Saha, S., & Bhattacharyya, P. (2022). Mental health disorder identification from motivational conversations. *IEEE Transactions on Computational Social Systems*.

Shao, Z., Chandramouli, R., Subbalakshmi, K., & Boyadjiev, C. T. (2019). An analytical system for user emotion extraction, mental state modeling, and rating. *Expert Systems with Applications*, *124*, 82–96.

Sharma, A., Choudhury, M., Althoff, T., & Sharma, A. (2020). Engagement patterns of peer-to-peer interactions on mental health platforms. *Vol. 14*, In *Proceedings of the international AAAI conference on web and social media* (pp. 614–625).

Sharma, E., & De Choudhury, M. (2018). Mental health support and its relationship to linguistic accommodation in online communities. In *Proceedings of the 2018 CHI conference on human factors in computing systems* (pp. 1–13).

Sharma, A., Lin, I. W., Miner, A. S., Atkins, D. C., & Althoff, T. (2021). Towards facilitating empathic conversations in online mental health support: A reinforcement learning approach. In *Proceedings of the web conference 2021* (pp. 194–205).

Sharma, A., Miner, A. S., Atkins, D. C., & Althoff, T. (2020). A computational approach to understanding empathy expressed in text-based mental health support. arXiv preprint arXiv:2009.08441.

Shi, W., & Yu, Z. (2018). Sentiment adaptive end-to-end dialog systems. arXiv preprint arXiv:1804.10731.

Singh, G. V., Priya, P., Firdaus, M., Ekbal, A., & Bhattacharyya, P. (2022). EmoInHindi: A multi-label emotion and intensity annotated dataset in Hindi for emotion recognition in dialogues. arXiv preprint arXiv:2205.13908.

Srinivasan, V., & Takayama, L. (2016). Help me please: Robot politeness strategies for soliciting help from humans. In *Proceedings of the 2016 CHI conference on human factors in computing systems* (pp. 4945–4955).

Srivastava, N., Hinton, G., Krizhevsky, A., Sutskever, I., & Salakhutdinov, R. (2014). Dropout: A simple way to prevent neural networks from overfitting. *Journal of Machine Learning Research*, *15*, 1929–1958.

Staliūnaitė, I., & Iacobacci, I. (2020). Auxiliary capsules for natural language understanding. In *ICASSP 2020-2020 IEEE international conference on acoustics, speech and signal processing* (pp. 8154–8158). IEEE.

Steinfeld, A., Fong, T., Kaber, D., Lewis, M., Scholtz, J., Schultz, A., et al. (2006). Common metrics for human-robot interaction. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on human-robot interaction* (pp. 33–40).

Sun, K., Zhang, R., Mensah, S., Mao, Y., & Liu, X. (2019). Aspect-level sentiment analysis via convolution over dependency tree. In *Proceedings of the 2019 conference on empirical methods in natural language processing and the 9th international joint conference on natural language processing* (pp. 5679–5688).

Tian, Y., Song, Y., Ao, X., Xia, F., Quan, X., Zhang, T., et al. (2020). Joint Chinese word segmentation and part-of-speech tagging via two-way attentions of auto-analyzed knowledge. In *Proceedings of the 58th annual meeting of the association for computational linguistics* (pp. 8286–8296).

Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., et al. (2017). Attention is all you need. In *Advances in neural information processing systems* (pp. 5998–6008).

Wadden, D., August, T., Li, Q., & Althoff, T. (2021). The effect of moderation on online mental health conversations. In *ICWSM*.

Wang, Y., Meng, X., Liu, Y., Sun, A., Wang, Y., Zheng, Y., et al. (2022). Chat-capsule: A hierarchical capsule for dialog-level emotion analysis. arXiv preprint arXiv:2203.12254.

Wang, Y. -C., Papangelis, A., Wang, R., Feizollahi, Z., Tur, G., & Kraut, R. (2020). Can you be more social? Injecting politeness and positivity into task-oriented conversational agents. arXiv preprint arXiv:2012.14653.

Wang, Y., Sun, A., Huang, M., & Zhu, X. (2019). Aspect-level sentiment analysis using as-capsules. In *The world wide web conference* (pp. 2033–2044).

Watts, R. J. (2003). *Politeness*. Cambridge University Press.

Welch, B. L. (1947). The generalization of 'student's'problem when several different population variances are involved. *Biometrika*, *34*, 28–35.

White, M., & Dorman, S. M. (2001). Receiving social support online: Implications for health education. *Health Education Research*, *16*, 693–707.

Xiao, L., Zhang, H., Chen, W., Wang, Y., & Jin, Y. (2018). Mcapsnet: Capsule network for text with multi-task learning. In *Proceedings of the 2018 conference on empirical methods in natural language processing* (pp. 4565–4574).

Xiao, C., Zhong, H., Guo, Z., Tu, C., Liu, Z., Sun, M., et al. (2018). Cail2018: A large-scale legal dataset for judgment prediction. arXiv preprint arXiv:1807.02478.

Xu, Y., Yao, E., Liu, C., Liu, Q., & Xu, M. (2023). A novel ensemble model with two-stage learning for joint dialog act recognition and sentiment classification. *Pattern Recognition Letters*, *165*, 77–83.

Yang, Y., Wu, B., Li, L., & Wang, S. (2020). A joint model for aspect-category sentiment analysis with textgcn and bi-GRU. In *2020 IEEE fifth international conference on data science in cyberspace* (pp. 156–163). IEEE.

Yang, D., Yao, Z., Seering, J., & Kraut, R. (2019). The channel matters: Self-disclosure, reciprocity and social support in online cancer support groups. In *Proceedings of the 2019 chi conference on human factors in computing systems* (pp. 1–15).

Yazdavar, A. H., Al-Olimat, H. S., Ebrahimi, M., Bajaj, G., Banerjee, T., Thirunarayan, K., et al. (2017). Semi-supervised approach to monitoring clinical depressive symptoms in social media. In *Proceedings of the 2017 IEEE/ACM international conference on advances in social networks analysis and mining 2017* (pp. 1191–1198).

Yazdavar, A. H., Mahdavinejad, M. S., Bajaj, G., Romine, W., Sheth, A., Monadjemi, A. H., et al. (2020). Multimodal mental health analysis in social media. *Plos One*, *15*, Article e0226248.

Yazdavar, A. H., Mahdavinejad, M. S., Bajaj, G., Thirunarayan, K., Pathak, J., & Sheth, A. (2018). Mental health analysis via social media data. In *2018 IEEE international conference on healthcare informatics* (pp. 459–460). IEEE.

Zhang, J., & Danescu-Niculescu-Mizil, C. (2020). Balancing objectives in counseling conversations: Advancing forwards or looking backwards. arXiv preprint arXiv:2005.04245.

Zhang, J., Filbin, R., Morrison, C., Weiser, J., & Danescu-Niculescu-Mizil, C. (2019). Finding your voice: The linguistic development of mental health counselors. arXiv preprint arXiv:1906.07194.

Zhang, C., Li, Y., Du, N., Fan, W., & Yu, P. S. (2018). Joint slot filling and intent detection via capsule neural networks. arXiv preprint arXiv:1812.09471.

Zhong, H., Xiao, C., Tu, C., Zhang, T., Liu, Z., & Sun, M. (2020). JEC-QA: A legal-domain question answering dataset. In *Proceedings of the AAAI conference on artificial intelligence*: *Vol. 34*, (pp. 9701–9708).

Zhou, H., Huang, M., Zhang, T., Zhu, X., & Liu, B. (2018). Emotional chatting machine: Emotional conversation generation with internal and external memory. In *Proceedings of the AAAI conference on artificial intelligence*: *Vol. 32*.

Zhou, X., & Wang, W. Y. (2017). Mojitalk: Generating emotional responses at scale. arXiv preprint arXiv:1711.04090.