A Survey of Socio-Emotional Strategies for Generation-Based Conversational Agents

Lorraine Vanel[1,2], Alya Yacoubi[2] and Chloé Clavel[1]
[1]*LTCI, Telecom-Paris, Institut Polytechnique de Paris, France*
[2]*Zaion, Paris, France*

Keywords: Task-Oriented Dialogue Systems, Social Dialogue, Emotional Dialogue, Persona, Empathetic Conversational System, Conversational Dataset, Neural Generation Models, End-to-End Generation.

Abstract: As dialogue systems are expected to display more and more "human" features, a subfield of conversational artificial intelligence (AI) has emerged, aiming to make virtual agents socially competent. As the common outlook firmly places emotion in the domain of chit-chatting, most of the studies tackle the problem of social behaviour in open domain applications. In this paper, we provide an overview of such approaches and think about how they can be applied to a task-oriented setting. The angle we explore is the influence of data on socio-emotional neural generation models. We focus on three aspects: dialogue strategies, emotional strategies and persona.

1 INTRODUCTION

Despite the development of increasingly advanced dialogue generation systems, one main challenge remains for conversational agents: *consistency*. It implies different sub-problems, such as logical consistency (relevance of the generated answer to the conversation history), persona consistency (relevance of the bot's behaviour to its past behaviour in the interaction), social consistency (relevance of the answer to a set of accepted social etiquette and rules) and emotional consistency (relevance to the user's emotional state).

While the logical consistency issue can be solved by the power of recent Natural Language Processing (NLP) engines (Shu et al., 2021), no unique definition or consensus has been reached in literature on the other aspects. However, social components have been proven to improve user experience: (Hardy et al., 2021) show that different dialogue strategies used by a conversational system such as back-channelling, self-disclosure and the use of statements strongly influence users' behaviour, with a positive impact on user engagement.

While several survey papers that study emotionally-aware generation systems exist (Ma et al., 2020; Wardhana et al., 2021; Pamungkas, 2019; Rashkin et al., 2018), they do not discuss more general social behaviours achieved through dialogue strategies or persona. This paper thus constitutes the first survey that takes into account those aspects of social dialogue generation. We look into the recent methods used in the open domain to see how they can be applied to task-oriented bots. To that end, we address the following questions:

Q1. What are the studied socio-emotional strategies and how are they relevant to task-oriented conversational systems? (Section 2)
Q2. How is socio-conversational data currently collected and annotated and what could be improved? (Section 3)

Afterwards, we present the remaining challenges for the field of socio-emotional end-to-end generation models (Section 4).

2 SOCIO-EMOTIONAL STRATEGIES

In the literature of end-to-end generation models for socio-emotional conversational systems, three dimensions clearly emerge: emotional strategies, (*i.e.*, empathy), dialogue strategies and persona.

Emotional Strategies. In literature, one of the most commonly studied emotional strategies is **empathy**.

Vanel, L., Yacoubi, A. and Clavel, C.
A Survey of Socio-Emotional Strategies for Generation-Based Conversational Agents.
DOI: 10.5220/0011632400003393
In *Proceedings of the 15th International Conference on Agents and Artificial Intelligence (ICAART 2023) - Volume 3*, pages 185-192
ISBN: 978-989-758-623-1; ISSN: 2184-433X

Various forms of empathy exist involving different biological processes, but human psychology generally defines empathy as the capacity to understand and feel what another person is experiencing. Being empathetic thus requires the ability to share the other's thoughts, to be aware of what they are going through and to imagine what it would feel like to be in that situation (Cuff et al., 2016). Multiple works in the domain of Artificial Intelligence have been aiming to reproduce this ability in machines (Fung et al., 2018; Lin et al., 2020; Wang et al., 2021; Hosseini and Caragea, 2021). In the case of NLP, it often translates into creating empathetic conversational systems. However, as machines are not yet able to perfectly understand deep, complex emotions or to infer the causes behind them, it is hard to talk about empathy. Hence why the term "Artificial Empathy" has been coined, to refer to the coding of a "simplified" empathy into AI and the different mechanisms that aim to model and mimic human empathy. Empathy is a term that catches people's attention, especially in the domain of conversational systems designed to interact with humans.

Dialogue Strategies. While empathy certainly does have its merits, as emotionally-aware approaches have proven to significantly improve user experience, one issue rises from this "boom" of empathy: most empathetic approaches seem to critically lack in any real interactive strategy. Even though it is overly represented in literature, empathy is far from being the sole option for efficient social and emotionally-aware interactions. Sometimes it is not even appropriate at all. Strategies can go beyond the line drawn by empathy/emotion, and step into the realm of "action". We call **dialogue strategy** a conversational plan of action to engage and react in a dialogue (Galescu et al., 2018; Santos Teixeira and Dragoni, 2022; Liu et al., 2021). In the following table, you will find a summary of the different dialogue strategies we have seen in the literature. Dialogue strategies as implemented in conversational systems represent the "realisation" of the social and emotional behaviours at an utterance-level. Some of them such as informing and questioning can be linked to the concept of dialogue acts (see upper section of Table 1).

Persona. While the dialogue generation will rely on variable features, like the conversational behaviours described previously, some "fixed" features can also be modelled, such as **persona**, or personality traits. Many works (Wolf et al., 2019; Madotto et al., 2019; Zheng et al., 2020; Wu et al., 2019) have started to explore the idea of granting bots a personality with

Table 1: List of dialogue strategies.

Dialogue Strategy	**Definition**
Agreeing	Showing agreement with the interlocutor *(Rashkin et al., 2018; Welivita et al., 2021)*
Informing	Giving relevant information *(Liu et al., 2021)*
Questioning	Asking a direct question *(Rashkin et al., 2018; Welivita et al., 2021; Liu et al., 2021)*
Back-Channeling	Interjecting to signal attention *(Hardy et al., 2021)*
Consoling	Cheering someone up if they show negative sentiment *(Rashkin et al., 2018; Welivita et al., 2021)*
Encouraging	Showing enthusiasm and support *(Rashkin et al., 2018; Welivita et al., 2021; Liu et al., 2021)*
Providing Suggestions	Giving advice on how to react to a situation *(Rashkin et al., 2018; Welivita et al., 2021; Liu et al., 2021)*
Restating	Reformulating the interlocutor's point *(Liu et al., 2021)*
Self-Disclosing	Revealing information about oneself *(Liu et al., 2021; Soni et al., 2021; Hardy et al., 2021)*
Sympathising	Reflecting the interlocutor's feelings *(Rashkin et al., 2018; Welivita et al., 2021; Liu et al., 2021)*

likes and dislikes. For multi-modal systems, persona can be conveyed through visual and auditory means, but for chatbots it heavily relies on textual content. Response generation will be regulated by the persona features, so that the system will have its own way to react to certain situations in a cohesive manner, using conversational strategies adapted to the interaction's context.

2.1 How Are Socio-Emotional Strategies Relevant to Task-Oriented Dialogue Systems?

Papers are usually not very specific when it comes to the strategies their models adopt. They use adjectives such as "empathetic" or "emotionally-aware" to refer to the ability to detect users' emotions, but are not necessarily very descriptive of the system's socio-emotional approach. Empathy is often used as an umbrella term to designate superficial behaviours where the model mirrors the user's emotion: the system will detect an emotion in the user's utterance using emotion recognition components, then use it to produce a response that matches the tone of the human interlocutor. If a mapping of the emotions is used (*i.e.*, a set of rules that decide what emotion the bot should

express in response to a detected user situation), it is rarely detailed.

Simulating an emotion involves more than only the textual content of the response: it is also displayed by the agent's behaviour. Psychology studies have formed theories on how action and emotion are linked in the brain's behavioural processes (Frijda et al., 1986). For humans, events that can trigger such reactions are numerous. They can be external such as during a social interaction or an interaction with objects, or internal which are provoked by oneself. They usually start a process that involves physiological and behavioural responses leading to one's awareness of their emotional state.

However, for a conversational system, the only stimuli is the external event triggered by the user's situation. In addition, the agent does not actually have feelings that can be influenced. When research tries to grant the bot the ability to "feel", it usually relies on making them "say" the emotion instead of "showing" it by merging the textual content with a fitting behavioural response. This is why we need to give dialogue systems the ability to adopt more general and lasting dialogue strategies, and to teach them how and when to shift their behaviour to better respond to users (Dias and Paiva, 2005; Mascarenhas et al., 2021). Having the bot respond to an emotional input with an emotional output can be sub-optimal, and even intrusive. We need to give them the ability to take certain types of conversational actions as described in Table 1. Those actions must be consistent with both the emotion the agent has to display and the dialogue strategy it has been using during the previous turns of the dialogue.

In task-oriented applications, the challenge is to be able to provide a response that is both efficient (*i.e.* meets the objective set by the task) and socially and emotionally accurate with regard to the user's situation. The goal of chit-chatting bots is the user's enjoyment and they naturally become the subject of studies that aim to add emotion within generated dialogues. However, many do not consider how more advanced social behaviours could benefit task-oriented chatbots. Chit-chat remains relevant to a task-oriented system (Sun et al., 2020) but its use implies many challenges, such as the timing, safety or relevance to the task.

Though some strategies listed in Table 1 are purely emotional like sympathising or encouraging, they can be applied to a task-oriented conversation for more efficient interaction. Some tasks require the bot to show socio-emotionally-driven approaches. In such cases, social (including emotions) and task should not be considered independently.

We can take a concrete example of a virtual agent tasked to help people calling a towing service. This conversational agent needs to converse with customers in distress or shock after a car accident. Having a bot able to recognise such situations, act accordingly and regulate the intensity of its emotional response depending on the context (Ochs et al., 2008) would be a great improvement to an inconsistent, systematic use of empathy - even if it is to transfer the conversation to a human agent better equipped to deal with the situation.

3 DATA

After establishing the importance of those strategies, our next step is to understand how they are illustrated in existing conversational corpora. Table 2 presents the different "socio-emotional" conversational datasets, along with a summary of some of their features. Our goal is to study how to effectively integrate the different socio-emotional dimensions and labels in conversational data.

3.1 Data Collection

Firstly, we go over the three main collection approaches we have observed in the literature of conversational systems.

3.1.1 Crowd-Sourcing

When applied to data collection, crowd-sourcing is a participatory method where a group of people contributes to creating data samples. The crowd-sourced data is typically Human-Human interaction (H-H). The data is collected by having two crowd-workers interact by following set guidelines: the speaker (or seeker) sets up a situation and the listener (or helper) has to answer accordingly. Dialogue systems are trained to perform in the role of Listeners. In (Rashkin et al., 2018; Liu et al., 2021), workers taking the role of speakers are asked to start the conversation following an emotional prompt. The listeners have to adapt their replies to the context presented by their interlocutor without being aware of the prompt.

3.1.2 Crawled from Online Sources

Another common way to collect data is by extracting information from online sources. In the case of textual data, it is often posts and comments crawled from social media and is thus natural Human-Human speech (Zhong et al., 2020; Mazaré et al., 2018). It

can also come from other sources, such as OpenSubtitles (Welivita et al., 2021) where the data is scripted. The data extracted from those websites is usually not labelled and annotation processes must be designed to label the corpora.

3.1.3 Retrieved from Conversation Logs

Human-Agent. Most of the time, this approach applies to human-machine interactions, through conversational models deployed for data-collection purposes or as services. When such models are already deployed, it is possible to retrieve the logs to compile them into corpora. One such dataset is *EmoContext* (Chatterjee et al., 2019), which can only be accessed by directly requesting it from its authors.

Human-Human. This approach can also be used to retrieve human-human conversation logs from call center data, like in (Clavel et al., 2013).

Such data is less accessible, as this practice requires having the means to deploy such services or requesting data from a company with such resources. Even then, the data is usually confidential and thus cannot be shared as public datasets, unless user consent was given and/or data has been properly anonymised.

3.2 Data Labelling

There are different approaches to data annotation: they differ depending on the point of view of the annotator, the tools and resources required to carry out the labelling tasks, or the information available directly from the source of the data.

Manual Annotations. This approach can be used on any kind of data as long as the dataset is small or if the research team has sufficient material means. Data can be entirely annotated by human experts or annotators that have been trained on that specific annotation task. (Li et al., 2017) introduce a 13K dataset, *DailyDialog*, that was annotated by 3 experts holding a good understanding of dialogue and communication theory, and were taught the guidelines of the particular task (*i.e.* emotion and dialogue act annotation).

Semi-Automatic Annotations. Manual annotation can be paired up with algorithms to accelerate the work and lighten the human judges' workload. This hybrid approach is called semi-automatic annotation (Lu et al., 2021; Welivita et al., 2021). In general, the first step is to have human judges annotate a small fraction of the collected dialogues. To provide further support to the human helpers, (Welivita et al., 2021) train a Bert-based model on the *empathetic dialogues* dataset (Rashkin et al., 2018) to output the top 3 most likely label for each dialogue. Instead of considering all 42 available labels, the human judges can pick between 3, with the possibility of selecting one of the others if needed. The second part of the process is to use this manually annotated set as training data for a classifier that will automatically annotate the rest of the collected data.

This can also be used for data retrieved from conversation logs. For their *EmoContext* dataset, (Chatterjee et al., 2019) have 50 human annotators manually annotate 300 dialogues for each of the 4 emotional classes. Each dialogue is looked over by 7 judges. These annotated dialogues were embedded as vectors, and used with cosine similarity thresholds to find similar occurrences in the non-annotated pool of data. Results would then be checked by human judges who make the last ruling.

Derived Annotations. It is the main method of annotation for crowd-sourced data, where the emotions and dialogue strategies labels associated with the data can be directly derived from the instructions given to the annotators in (Rashkin et al., 2018; Liu et al., 2021). Additionally, (Liu et al., 2021) collect answers to surveys submitted to the workers during the collection process on both the listener and the speaker side. This allows for the collection of more data such as empathy grading and utterance-level dialogue strategies.

(Zhong et al., 2020) use the context in which web data was posted and extracts posts and comments on two subreddits: happy and offmychest. The original Reddit environment thus provides a label and what is left to do is to direct a quality check by asking human annotators to annotate a small set of the conversations: 100 from the happy reddit, 100 from the offmychest reddit and for control, 100 from the casualconversations reddit.

For conversation logs, derived annotation can be user feedback. This is because some deployed bots ask for customer satisfaction feedback either directly or through surveys. This information can be used to annotate certain conversations (Maslowski et al., 2017).

In the case of derived annotation, the label comes directly from the source of the data. The point of view of the annotator is internal, which differs greatly from the external viewpoint of the annotators in the other methods described above, where they have to analyse data that comes from an outside source. This difference is even more prominent when it comes to labels

Table 2: List of Conversational Emotional Corpora in Literature. *Volume refers to the number of conversations. The Dimension Labels column presents the different dimension presented by the datasets: ES for Emotional Strategy, DS for Dialogue Strategies and P for Persona; $_U$ refers to utterance-level annotation, and $_C$ refers to conversation-level annotation; the following number refers to the number of labels in the category. For collection, CS refers to Crowd-Sourced. N stands for "Natural", where S stands for "Scripted".*

Name	Interaction	Volume	Dimension Labels	Collection	Scripted-Natural	Annotation
Emotional Support Conversation Dataset (ESConv), (Liu et al., 2021)	H-H	1300	$ES_C(7)$, $DS_C(8)$	CS	S	Emotion Prompt
Empathetic dialogues, (Rashkin et al., 2018)	H-H	25k	$ES_C(32)$, $DS_U(8)$	CS	S	Emotion Prompt
EmoWOZ, (Feng et al., 2021)	H-H	11K	$ES_U(7)$	CS	S	Manual annotation (Amazon Mechanical Turk)
Persona Chat (PC), (Zhang et al., 2018)	H-H	10K	P(1155)	CS	S	No Annotation
EmotionLines, (Chen et al., 2018)	H-H	2K	$ES_U(7)$	TV Script	S	Manual annotation with 5 Human Annotators
DailyDialog, (Li et al., 2017)	H-H	13K	$ES_U(6)$, $DS_U(4)$	Internet	S	Manual Annotation by 3 Human Annotators
Emotional dialogue in OpenSubtitles (EDOS), (Welivita et al., 2021)	H-H	1M	$ES_U(32)$, $DS_U(8)$	OpenSubtitles	S	Semi-Automatic annotation with 3 Human Annotators
Persona-based Empathetic Conversation (PEC), (Zhong et al., 2020)	H-H	11k	$ES_C(2)$, P(100)	Reddit	N	Automatic annotation with Human Control Check
PCR, (Mazaré et al., 2018)	H-H	5M	P(7.3M)	Reddit	N	Semi-Automatic annotation with Human Annotators
Sentimental Douban Conversational Corpus, (Lu et al., 2021)	H-H	1M	ES_U(Sentiment)	Douban	N	Semi-Automatic annotation with 3 Human Annotators
EmoContext, (Chatterjee et al., 2019)	H-M	38K	$ES_C(4)$	Bot Logs	N	Semi-Automatic annotation with 50 Human Annotators

as subjective as emotion. The internal annotator can label accurately what emotion they are feeling and expressing in the data, but an external annotator has to make a guess as to what emotion is being expressed. This can be an easy guess, or be quite ambiguous and perceived differently depending on one's personal experiences and sensibilities.

3.3 Annotation Dimensions

We now present three "dimensions" (Column 4 in Table 2) that are annotated in data: emotional labels, dialogue strategies and persona features. Another aspect we comment on is the annotation level: labels can either be given on a conversation-level (one label for the entire dialogue) or utterance-level (one label per speaker turns).

Emotional States. As we have mentioned in the first section on dialogue strategies, empathy and emotions are the most represented socio-emotional strategy. As a direct consequence of this, most social corpora are annotated with emotional states.

(Lu et al., 2021) use sentiment instead of emotion. Sentiment analysis is an NLP task that consists of understanding the polarity of an utterance and classifying it into the *positive* sentiment or the *negative* sentiment. Sometimes there is also the *neutral* class, but (Lu et al., 2021) only use binary labels. (Zhong et al., 2020) do something similar to sentiment labelling, using the reddit context the data was extracted from to annotate whole conversations: *happy* for a positive polarity, *offmychest* for a negative polarity and *casualConversation* for a neutral control sample.

Other studies have referred to various psychology theories for emotion classification, but there is no consensus on the way to define and classify emotions in conversation analysis (Clavel and Callejas, 2016). (Chatterjee et al., 2019) use a simplified classification of emotions, *angry, happy, sad* and *others*. (Li et al., 2017) base their annotations on Ekman's model (Ekman, 1999). On the other hand, (Liu et al., 2021; Rashkin et al., 2018; Welivita et al., 2021) use classical theories derived from biological responses (Ekman, 1999; Plutchik, 1984) as well as studies that focused on larger sets of subtle, context-dependant emotions (Skerry et al., 2015) reaching as many as 32 emotion labels.

(Feng et al., 2021) use a classification tailored to task-oriented settings: *neutral, fearful, dissatisfied, apologetic, abusive, excited* and *satisfied*.

Dialogue Strategies. Another dimension of annotation is dialogue strategies, as some papers have led a dual annotation, taking into account both emotions and dialogue strategies. Table 1 presented in the first section summarises the ones we have seen being used in literature. As seen in Section 2, dialogue strate-

gies can be associated with dialogue acts and a large number of existing corpora are annotated in dialogue acts. For example, (Li et al., 2017) use *inform, question, directive, commissive* as labels to develop socio-emotional conversational systems. Such strategies are included in the works of (Bunt, 2006) on the classification and the dimensions of dialogue acts.

Persona. The last aspect we want to introduce in this section is persona features. Corpora that introduce persona characteristics (Zhong et al., 2020; Mazaré et al., 2018) usually define a "persona" as a set of sentences representing a given personality that influences the agent's wording and general behaviour. Persona sentences are often collected from online user profiles found in conversational data, with one user representing one persona. It may be independent from the conversation that can be annotated with other labels such as emotion.

3.4 Guidelines for the Collection and Annotation of Socio-Emotional Data

For a model to provide good performance, the training data, as well as the label types, must be similar to the real-time data the system will face when deployed to users. Public corpora often compile conversations from the open domain that exhibit different strategies, formulations and goals. For good task-oriented results, it is a challenge to gather data that is the closest possible to the task's conversational context.

For the means of collection, crowdsourcing allows for precise control on the content of the dialogues, the gathering of numerous and diverse samples that answer specific problems as defined by the research team, and a label given directly by the source of the data (the *speaker*). Nevertheless, the scripted nature of the interactions induces a significant bias in the corpus, as it differs from how one might spontaneously express themselves in a similar situation. Open-domain data crawled from online sources is another issue, as it is not structured in the same way task-oriented dialogues are. Ultimately, the optimal solution for service providers that already have agents deployed in production is to retrieve and use conversation logs. Data retrieved from deployed conversational agents falls squarely into the Human-Machine interaction scope. While the acceptability of such bots can influence the conversation and the way the user conveys their intent, the interaction remains spontaneous and authentic. However, biases that are intrinsic to the machine's behaviour and development will be preserved in the corpus. Even so, it offers the best compromise between task-oriented utterances that match the intended use of the bot and the spontaneous nature of the conversations and the genuine reactions, even if they might be influenced by their acceptability of technology and AI in customer service.

Once the data has been collected, we believe that a very important step must precede the annotation phase. A labelling process must be designed that establishes what strategies and emotions need to be annotated. This can be done by studying patterns in the data and extracting the features that will be taught to the agent. Once this process is clearly defined, a semi-automatic annotation allows a good optimisation of resources, while guaranteeing good consistency with quality control by human annotators. It is important to note that annotating emotional states is a complex task, even for humans, because of how subjective it is. Two people might not agree on an emotion expressed in a same sample, because of the personal and internal biases that colour our perception of others' emotional states.

If we consider our towing service example again: how can we build a socio-emotional dataset relevant to the response generation task? The towing company provided us with conversation files between their human agents and their customers. For the collection of relevant annotations in dialogue strategies, we look at the different sets of socio-emotional labels used in literature (see Section 3.3). By observing conversations from the company's logs, we are able to narrow down a list of relevant strategies and emotion labels. We noticed that some might not be as common in the context of our task, like 'disgust' for emotions. Furthermore, some new labels emerged and were added to our list, such as 'Repair/Self-Repair' for the act of correcting oneself or one's interlocutor.

4 REMAINING CHALLENGES

The growing availability of efficient and pre-trained language models (Devlin et al., 2019; Radford et al., 2019) has spurred the field of end-to-end NLP models. These data-driven approaches applied to socio-emotional dialogue have raised many challenges.

The first one is related to the architecture: the absence of causality modelling creates a lack of control and explainability of the process (Yacoubi and Sabouret, 2018). Nonetheless, some approaches have attempted to bridge this issue, offering better visibility on the problem decomposition while making the most of the computational benefits of end-to-end training (Gao et al., 2021; Soni et al., 2021).

Secondly, these approaches also need a huge

amount of annotated data, which implies specific data collection and annotation processes.

Finally, papers tend to contrast automatic metrics designed to evaluate text generation systems (Papineni et al., 2002; Lin, 2004) to human evaluation (Finch and Choi, 2020). Empathetic and social behaviour is better appraised by humans who are both the target-users and have the social intelligence that AI systems try to replicate.

5 CONCLUSION

This survey paper provides an overview of the recent works that tackle the challenging goal of adding a socio-emotional dimension to end-to-end dialogue generation. As most of those studies focus on open-domain applications, we think about how to apply these methods to a task-oriented context. In order to create conversational systems with efficient social behaviour, it is important to combine dialogue strategies and emotions, as well as eventually adding persona features for better consistency in the responses.

Our future work involves designing a general, efficient and comprehensive labelling process based on the observations made in this paper. This also includes verifying the annotation process and the quality of the resulting labelling.

REFERENCES

Bunt, H. (2006). Dimensions in dialogue act annotation. In *Proceedings of the Fifth International Conference on Language Resources and Evaluation (LREC'06)*, Genoa, Italy. European Language Resources Association (ELRA).

Chatterjee, A., Narahari, K. N., Joshi, M., and Agrawal, P. (2019). SemEval-2019 task 3: EmoContext contextual emotion detection in text. In *Proceedings of the 13th International Workshop on Semantic Evaluation*, pages 39–48, Minneapolis, Minnesota, USA. Association for Computational Linguistics.

Chen, S.-Y., Hsu, C.-C., Kuo, C.-C., Ting-Hao, Huang, and Ku, L.-W. (2018). Emotionlines: An emotion corpus of multi-party conversations.

Clavel, C., Adda, G., Cailliau, F., Garnier-Rizet, M., Cavet, A., Chapuis, G., Courcinous, S., Danesi, C., Daquo, A.-L., Deldossi, M., Guillemin-Lanne, S., Seizou, M., and Suignard, P. (2013). Spontaneous speech and opinion detection: mining call-centre transcripts. *Language Resources and Evaluation*, 47(4):1089–1125.

Clavel, C. and Callejas, Z. (2016). Sentiment analysis: From opinion mining to human-agent interaction. *IEEE Transactions on Affective Computing*, 7(1):74–93.

Cuff, B., Brown, S., Taylor, L., and Howat, D. (2016). Empathy: A review of the concept. *Emotion Review*, 8:144–153.

Devlin, J., Chang, M.-W., Lee, K., and Toutanova, K. (2019). BERT: Pre-training of deep bidirectional transformers for language understanding. In *Proceedings of the 2019 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long and Short Papers)*, pages 4171–4186, Minneapolis, Minnesota. Association for Computational Linguistics.

Dias, J. and Paiva, A. (2005). Feeling and reasoning: A computational model for emotional characters. In *Portuguese conference on artificial intelligence*, pages 127–140. Springer.

Ekman, P. (1999). Basic emotions. *Handbook of cognition and emotion*, 98(45-60):16.

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

Finch, S. E. and Choi, J. D. (2020). Towards unified dialogue system evaluation: A comprehensive analysis of current evaluation protocols. *arXiv preprint arXiv:2006.06110*.

Frijda, N., FRIDJA, N., Manstead, A., and Oatley, K. (1986). *The Emotions*. Studies in Emotion and Social Interaction. Cambridge University Press.

Fung, P., Bertero, D., Xu, P., Park, J. H., Wu, C.-S., and Madotto, A. (2018). Empathetic dialog systems. In *The international conference on language resources and evaluation. European Language Resources Association*.

Galescu, L., Teng, C. M., Allen, J., and Perera, I. (2018). Cogent: A generic dialogue system shell based on a collaborative problem solving model. In *Proceedings of the 19th Annual SIGdial Meeting on Discourse and Dialogue*, pages 400–409.

Gao, J., Liu, Y., Deng, H., Wang, W., Cao, Y., Du, J., and Xu, R. (2021). Improving empathetic response generation by recognizing emotion cause in conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 807–819.

Hardy, A., Paranjape, A., and Manning, C. D. (2021). Effective social chatbot strategies for increasing user initiative. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 99–110.

Hosseini, M. and Caragea, C. (2021). It takes two to empathize: One to seek and one to provide. In *Proceedings of the AAAI Conference on Artificial Intelligence. To appear*.

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

Lin, C.-Y. (2004). ROUGE: A package for automatic evaluation of summaries. In *Text Summarization Branches Out*, pages 74–81, Barcelona, Spain. Association for Computational Linguistics.

Lin, Z., Xu, P., Winata, G. I., Siddique, F. B., Liu, Z., Shin, J., and Fung, P. (2020). Caire: An end-to-end empathetic chatbot. In *Proceedings of the AAAI Conference on Artificial Intelligence*, volume 34, pages 13622–13623.

Liu, S., Zheng, C., Demasi, O., Sabour, S., Li, Y., Yu, Z., Jiang, Y., and Huang, M. (2021). Towards emotional support dialog systems. *ArXiv*, abs/2106.01144.

Lu, X., Tian, Y., Zhao, Y., and Qin, B. (2021). Retrieve, discriminate and rewrite: A simple and effective framework for obtaining affective response in retrieval-based chatbots. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 1956–1969.

Ma, Y., Nguyen, K. L., Xing, F. Z., and Cambria, E. (2020). A survey on empathetic dialogue systems. *Information Fusion*, 64:50–70.

Madotto, A., Lin, Z., Wu, C.-S., and Fung, P. (2019). Personalizing dialogue agents via meta-learning. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5454–5459.

Mascarenhas, S., Guimarães, M., Santos, P. A., Dias, J., Prada, R., and Paiva, A. (2021). Fatima toolkit–toward an effective and accessible tool for the development of intelligent virtual agents and social robots. *arXiv preprint arXiv:2103.03020*.

Maslowski, I., Lagarde, D., and Clavel, C. (2017). In-the-wild chatbot corpus: from opinion analysis to interaction problem detection. In *ICNLSSP 2017*, pages 115–120, Casablanca, Morocco. ISGA, Institut Supérieur d'inGénierie et des Affaires.

Mazaré, P.-E., Humeau, S., Raison, M., and Bordes, A. (2018). Training millions of personalized dialogue agents. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2775–2779, Brussels, Belgium. Association for Computational Linguistics.

Ochs, M., Sabouret, N., and Corruble, V. (2008). Modeling the dynamics of non-player characters' social relations in video game. In *Proceedings of the AAAI Conference on Artificial Intelligence and Interactive Digital Entertainment*, volume 4, pages 90–95.

Pamungkas, E. W. (2019). Emotionally-aware chatbots: A survey. *arXiv preprint arXiv:1906.09774*.

Papineni, K., Roukos, S., Ward, T., and Zhu, W.-J. (2002). Bleu: a method for automatic evaluation of machine translation. In *Proceedings of the 40th Annual Meeting of the Association for Computational Linguistics*, pages 311–318, Philadelphia, Pennsylvania, USA. Association for Computational Linguistics.

Plutchik, R. (1984). Emotions : a general psychoevolutionary theory.

Radford, A., Wu, J., Child, R., Luan, D., Amodei, D., and Sutskever, I. (2019). Language models are unsupervised multitask learners.

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

Santos Teixeira, M. and Dragoni, M. (2022). A review of plan-based approaches for dialogue management. *Cognitive Computation*, pages 1–20.

Shu, C., Zhang, Y., Dong, X., Shi, P., Yu, T., and Zhang, R. (2021). Logic-consistency text generation from semantic parses. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 4414–4426, Online. Association for Computational Linguistics.

Skerry, A. E., Saxe, R., Skerry, A. E., and Saxe, R. (2015). Neural representations of emotion are organized around abstract event features. *Curr. Biol.*

Soni, M., Cowan, B., and Wade, V. (2021). Enhancing self-disclosure in neural dialog models by candidate re-ranking. *arXiv preprint arXiv:2109.05090*.

Sun, K., Moon, S., Crook, P., Roller, S., Silvert, B., Liu, B., Wang, Z., Liu, H., Cho, E., and Cardie, C. (2020). Adding chit-chat to enhance task-oriented dialogues.

Wang, Y.-H., Hsu, J.-H., Wu, C.-H., and Yang, T.-H. (2021). Transformer-based empathetic response generation using dialogue situation and advanced-level definition of empathy. In *2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP)*, pages 1–5.

Wardhana, A. K., Ferdiana, R., and Hidayah, I. (2021). Empathetic chatbot enhancement and development: A literature review. In *2021 International Conference on Artificial Intelligence and Mechatronics Systems (AIMS)*, pages 1–6.

Welivita, A., Xie, Y., and Pu, P. (2021). A large-scale dataset for empathetic response generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 1251–1264.

Wolf, T., Sanh, V., Chaumond, J., and Delangue, C. (2019). Transfertransfo: A transfer learning approach for neural network based conversational agents. *arXiv preprint arXiv:1901.08149*.

Wu, B., Li, M., Wang, Z., Chen, Y., Wong, D., Feng, Q., Huang, J., and Wang, B. (2019). Guiding variational response generator to exploit persona.

Yacoubi, A. and Sabouret, N. (2018). Teatime: A formal model of action tendencies in conversational agents. In *ICAART (2)*, pages 143–153.

Zhang, S., Dinan, E., Urbanek, J., Szlam, A., Kiela, D., and Weston, J. (2018). Personalizing dialogue agents: I have a dog, do you have pets too? In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 2204–2213, Melbourne, Australia. Association for Computational Linguistics.

Zheng, Y., Zhang, R., Huang, M., and Mao, X. (2020). A pre-training based personalized dialogue generation model with persona-sparse data. In *Proceedings of the AAAI Conferencezh on Artificial Intelligence*, volume 34, pages 9693–9700.

Zhong, P., Zhang, C., Wang, H., Liu, Y., and Miao, C. (2020). Towards persona-based empathetic conversational models. *arXiv preprint arXiv:2004.12316*.