A Critical Reflection and Forward Perspective on Empathy and Natural Language Processing

Allison Lahnala[1]**, Charles Welch**[1,3]**, David Jurgens**[2]**, Lucie Flek**[1,3]
[1]Conversational AI and Social Analytics (CAISA) Lab
Department of Mathematics and Computer Science, University of Marburg
[2] School of Information, University of Michigan
[3] The Hessian Center for Artificial Intelligence (Hessian.AI)
{allison.lahnala,welchc,lucie.flek}@uni-marburg.de, jurgens@umich.edu

Abstract

We review the state of research on empathy in natural language processing and identify the following issues: (1) empathy definitions are absent or abstract, which (2) leads to low construct validity and reproducibility. Moreover, (3) emotional empathy is overemphasized, skewing our focus to a narrow subset of simplified tasks. We believe these issues hinder research progress and argue that current directions will benefit from a clear conceptualization that includes operationalizing cognitive empathy components. Our main objectives are to provide insight and guidance on empathy conceptualization for NLP research objectives and to encourage researchers to pursue the overlooked opportunities in this area, highly relevant, e.g., for clinical and educational sectors.

1 Introduction

Interest in empathetic language continues to grow in natural language processing research. Empathy recognition and empathetic response generation tasks have become well-established research directions, especially since the introduction of now fairly mainstream benchmark datasets for each task, EMPATHIC REACTIONS (Buechel et al., 2018) and EMPATHETIC DIALOGUES (Rashkin et al., 2019), and an empathy detection shared task established on the former (Tafreshi et al., 2021; Barriere et al., 2022).

These empathy-focused tasks are highly motivated by myriad benefits, including (i) improved experiences with conversational agents (Shuster et al., 2020; Roller et al., 2021) and satisfaction with customer care dialogue agents (Firdaus et al., 2020; Sanguinetti et al., 2020), (ii) computational social science analyses, e.g., supportive interactions in online forums (Khanpour et al., 2017; Zhou et al., 2020; Sharma et al., 2020), and (iii) tools to assist in training and evaluating health practitioners (Demasi et al., 2019; Pérez-Rosas et al., 2017). With this increased interest in empathy-focused NLP, however, has come little clarity on what empathy is and how it is being operationalized. Papers vary substantially in how they define, annotate, and evaluate empathy, leading to a fractured landscape that hinders progress.

Most NLP work has loosely defined empathy as the *ability to understand another person's feelings and respond appropriately*. As a result, these works have focused primarily on detecting sentiment and emotions in text as proxies for *understanding feelings* and consider empathetic responses to be those that demonstrate success in emotion recognition and express so with a tone consistent with the valence of the target's sentiment. Under this view, systems that primarily recognize or perform emotionally coherent interactions supposedly fulfill the goal of having an empathetic system. However, established findings and theories about human empathy show that this objective is short-sighted—or even misdirected—and misses a critical system known as *cognitive empathy*.

We argue that NLP research has omitted key aspects of empathy through its narrow focus on emotion and, as a result, led us to neglect the cognitive components. Our paper offers three key contributions. First, to show this gap in research on empathy, we provide a theoretical grounding for empathy from psychology (§2). We then survey empathy literature in NLP focusing on ACL[*] venues (§3) and highlight three central problems: (1) computational work has overlooked much of empathy through vague definitions and a focus on emotion (§4), (2) our current underspecification of empathy leads to issues in construct and data validity (§5), and (3) the narrow empathy tasks we choose as a community limit our progress (§6). Finally, we propose a way forward (§7) through a clear conceptualization of empathy that operationally considers processes of cognitive empathy and highlight overlooked research opportunities (§8).

2 What *is* Empathy? A Theoretical Guide

Empathy is a multi-dimensional construct that contains both emotional and cognitive aspects which relate to how an observer reacts to a target (Davis, 1980, 1983). In practice, empathy is diversely defined in terms of these social, emotional, cognitive, and even neurological dimensions (Cuff et al., 2016). Indeed, while folk conceptions of empathy refer to a single construct, multiple studies have pointed to distinct neurological systems for emotional and cognitive empathy (Decety and Jackson, 2006; Shamay-Tsoory, 2011).

In describing empathy in communication, we adopt the standard terminology of (i) a *target* as someone experiencing an emotion or situation and (ii) an *observer* as another person at the disposition for an empathic experience through perceiving the target's emotions and situation.

In psychology, the most discussed aspects of empathy are its affective and cognitive components (Cuff et al., 2016). The affective components, referred to as *emotional empathy*, relate to the observer's emotional reaction to the target (Shamay-Tsoory, 2011). The cognitive components, referred to as *cognitive empathy*, relate to active processes used by the observer to infer the mental state of the target (Blair, 2005). Emotional empathy represents *automatic* (bottom-up) processes whereas cognitive empathy represents *controlled* (top-down) processes (Lamm et al., 2007), and they interact with each other.

2.1 Emotional Empathy

Emotional empathy, or the observer's capacity to experience affective reactions upon perceiving the target, can involve processes such as emotion recognition, contagion, and pain sharing (Shamay-Tsoory, 2011).[1] When these affective processes caused by perceiving the target's emotional state interact with certain contextual factors, the observer's experience is a form of *emotional empathy*. Such forms include the concepts of sympathy, compassion, and tenderness. In NLP, we often use these terms to define empathy without any distinction between them. Accordingly, our review finds we do not investigate them as specific empathy-related phenomena or characterizations of empathetic responses. As we describe in this section, however, the degree to which specific contextual factors interact with the internal affective processes renders

[1] We further discuss contagion and mimicry in Appendix C.

these concepts distinct. We propose differentiable characteristics of sympathy, compassion, and tenderness based on psychology literature. Furthermore, these characteristics are operationalizable for more precise NLP research on emotional empathy.

Distinguishing sympathy, compassion, and tenderness. While each of these concepts relates to having feelings *for* the target, each one is distinct regarding the perceived immediacy of the target's needs, vulnerability, and a desire to help. Sympathy relates to the present situation of the target and involves sorrow and concern for the target arising from perceived suffering, whereas tenderness relates to their long-term needs, and involves warm and fuzzy feelings from perceiving the target as vulnerable, delicate, or defenseless. Compassion is a higher construct that involves feelings based on the target's perceived needs and motivated desires to protect the vulnerable and provide care to those who suffer (Goetz et al., 2010).

Feeling *for* the target means emotions can be different between the target and observer. Emotional congruence is used to describe when an emotion is felt by both target and observer, though some define congruence as a response to a situation that is similar but may result in a different emotion. In this case, the emotion is congruent if it is appropriate for the situation. This idea plays a significant role in how the NLP community views empathetic responses, though we consider it unnecessary for empathy in general.

We can perceive another person as vulnerable or suffering and experience tenderness, sympathy, and compassion for them without thinking hard about it. In other words, cognitive processes are not strongly required for these experiences, distinguishing them as *emotional empathy*. However, empathic cognitive processes can help render these experiences by elevating awareness of those contextual factors (i.e., realizing the target's needs and vulnerability) through active deliberation of the target's situation. In the next section, we describe such processes of *cognitive empathy*.

2.2 Cognitive Empathy

Cognitive empathy centers, in part, on perspective-taking, the process of the observer conceptualizing the target's point of view. This active process is the primary way of achieving cognitive empathy, though other scenarios such as imagined memories or fictional scenarios may also be processes that

help to achieve it (Eisenberg, 2014; Stinson and Ickes, 1992).

Psychologists have proposed a variety of frameworks for what actions or processes constitute perspective-taking. For example, in the *appraisal theory of empathy* (Wondra and Ellsworth, 2015), empathy is considered with respect to different aspects or dimensions of perspective that an observer might have for the target's situation. Six different types of appraisals are proposed: pleasantness, anticipated effort to deal with the situation, anticipated control of the situation, self-other agency for responsibility of the situation, attentional activity (degree of surprise), and certainty of the situation or outcome. With this view, how "empathetic" the observer is depends on the number of appraisals and the degree to which their responses or actions mirror the true feelings of the target.

To be able to empathize successfully, the observer needs to be able to accurately infer the content of the target's thoughts and feelings (Ickes, 2011). The degree of accuracy in the observer's inferences is known as *empathic accuracy*. High empathic accuracy is particularly essential in clinical domains, such as Motivational Interviewing (Miller and Rollnick, 2012), where the therapist (observer) formulates *reflections* based on interpreting what the patient (target) says (see examples in Table 1).

When considering the factors that affect empathic accuracy, the observer is not the only possible source of error. The target must also express and convey the situation accurately. As a result, it is unlikely that both the observer and target will ever perceive situations exactly the same (Stotland et al., 1978). This limitation has implications for empathy annotations, with the inter-annotator agreement being subject to the empathic accuracy and subjective interpretations of the annotators. The *unstructured dyadic interaction* and *standard stimulus* paradigms are two study approaches developed in interpersonal perception research for measuring empathic accuracy that involve comparing observer inferences with the target self-reports of their actual thoughts and feelings (Ickes, 2001); such approaches to study empathic accuracy have (to our knowledge) never been explored in NLP research.

3 Literature Review: Empathy in NLP

Empathy research in NLP primarily focuses on empathy recognition and empathetic response gen-

Client	Well, I know I need to stay active. "Use it or lose it," they say. I want to get my strength back, and they say regular exercise is good for your brain, too.
Interviewer	So that's a puzzle for you – how to be active enough to get your strength back and be healthy, but not so much that would put you in danger of another heart attack.
Client	I think I'm probably being too careful. My last test results were good. It just scares me when I feel pain like that.
Interviewer	It reminds you of your heart attack.
Client	That doesn't make much sense, does it– staying away from activity so I won't have another one?
Interviewer	Like staying away from people so you won't be lonely.
Client	Right. I guess I just need to do it, figure out how to gradually do more so I can stick around for a while.

Table 1: A motivational interviewing interaction demonstrating complex reflections. Example from (Miller and Rollnick, 2012).

eration. There is also work that analyzes empathy-related behaviors, such as supportive intents and counselor strategies. Across this literature, we have found highly varied and often underspecified usage of the term *empathy*.

We summarize our findings from reviewing a collection of computational linguistics and natural language processing papers. Papers were identified by searching for the term "empath" in the ACL anthology. Then we narrowed the resulting sample from 90 to 69 papers whose investigation involves empathy, as opposed to mentioning "empathy", "empathetic", or "empathic" rhetorically. We also included a selection of relevant works outside the ACL anthology from venues such as AAAI.

We focused on papers presenting systems for empathy prediction or empathetic response generation, evaluations of empathy dialogue models, empathy annotation schemes and datasets, and corpus analyses of empathetic language. We labeled a total of 48 papers (excluding 14 WASSA[2] shared task papers) based on their descriptions of empathy and empathy evaluation approaches.

3.1 Findings

We identified six predominant themes of *empathy descriptions* (D0-5), and seven *empathy evaluation approaches* (E0-6). Tables 2 and 3 show the categories we defined for grouping the description

---

[2]Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis

	Definition Themes	Count
D0	Studies do not define or describe empathy.	8
D1	Studies do not describe empathy but an abstract conceptualization could be inferred from the task or system description.	10
D2	Studies provide empathy descriptions that are vague or ambiguous with other concepts.	10
D3	Studies describe relationships between certain behaviors and empathy.	3
D4	Studies provide succinct yet explicitly theory-grounded empathy descriptions and adhere to the theoretical foundation.	9
D5	Studies provide thorough, theory-based descriptions of empathy.	8

Table 2: The number of papers identified for each Definition Theme in our review.

and evaluation themes, together with the number of papers we identified characterized by these themes.

The definition themes are as follows (details and examples are provided in Appendix A):

D0: Studies do not define or describe empathy. Despite the significance of empathy in such studies, a conceptualization is not given nor can it be reasonably inferred. In some cases, empathy is associated with *politeness* and *courtesy*.

D1: Studies do not describe empathy but an abstract conceptualization could be inferred from the task or system description. This is especially the case among empathetic response generation research. In task descriptions, the tendency is to describe empathetic response generation as *the task of understanding the user's emotions and responding appropriately*. System descriptions can be informative about the grounding concept when the design reflects specific dimensions of empathy. For instance, an empathetic response generation system that mainly relies on an emotion recognition module for the purpose of response conditioning suggests the significance of emotion understanding in the work's conceptualization.

D2: Studies provide empathy descriptions that are vague or ambiguous with other concepts. Some studies use *sympathy* and *empathy* interchangeably, and others nearly exchange the term *empathy* with *emotion recognition* by strongly associating them without providing disambiguation. These often regard an appropriate response as one that mimics or mirrors the target's response. This characterization usually accompanies a system that conditions a response on emotions or sentiments predicted by a dedicated module. In other cases, papers reference the conceptualization linked to the dataset they used, as is frequent among WASSA shared task papers (listed in Appendix §B) using the EMPATHIC REACTIONS dataset (Buechel et al., 2018).

D3: Studies describe relationships between certain behaviors and empathy. These works, mainly concerned with narrative and conversation analyses, provide thorough descriptions of other concepts and their relations to empathy that are consistent with multiple aspects of empathy described in §2.

D4: Studies provide succinct yet explicitly theory-grounded empathy descriptions and adhere to the theoretical foundation. This research often involves a counseling/therapy conversations dataset with labels from a scheme developed by experts in those domains.

D5: Studies provide thorough, theory-based descriptions of empathy. This is often the case in works that develop a novel multi-dimensional empathy framework for analysis and annotations. As in D4, they also tend to involve experts in behavioral, social, and health domains.

4 The Definition Problem

The lack of delineation between empathy and related concepts in psychology ultimately leaves us wondering what we are studying. When empathy is not explicitly conceptualized (D0-1), the training data is often implicitly focused on components of emotional empathy and responses based on emotion-matching strategies. However, since the data properties are often under-specified in the NLP papers as well, we only have indirect proxies to infer such emotion-centric conceptualizations. In an ideal case, the training data description would reveal more about the relevant features of emotion matching/mirroring, e.g., separating contextual factors that would define a particular emotional empathy behavior such as sympathy, compassion, or tenderness.

	Evaluation Themes	Count
E6	Multi-item: cognitive & emotional	8
E5	Single label/rating: cognitive & emotional empathy	3
E4	Single label/rating: only emotional empathy	19
E3	Single label/rating: no specification	8
E2	Heuristic empathy labels/ratings	4
E1	Target-observer role labeling	2
E0	No manual evaluation or only automatic	4

Table 3: Themes of evaluations and annotations of empathetic language in NLP literature with counts of papers described the theme. Detailed descriptions of the themes are provided in Appendix B.

While empathy is indeed complex, we demonstrated in Section 2.1 a way to disambiguate these empathetic behaviors by considering the variables of how the observer perceives the target's vulnerability and needs. These are just some possible manifestations of emotional empathy that NLP researchers could investigate using more detailed study designs to control such variables. To alleviate the problem of abstractness and potential inconsistencies, we recommend that researchers disambiguate their objectives from simply "empathy" by considering such concepts as sub-areas of the empathy research direction.

Nearly thirty studies in our review provide vague or no empathy descriptions (D0-2). Without the ability to focus on individual aspects or an understanding of the nature of empathetic interactions and thus what constitutes "appropriate" responses, the issue of what we are measuring arises. Figure 1 displays the interaction between empathy definitions and evaluation designs. There is a clear correspondence between under-specified empathy descriptions and under-specified evaluations. Even when a study provides a multi-dimensional definition of empathy (D4-5), it often is not reflected in the evaluation design. Rather the evaluations default intuitively to the emotional components (E4).

The undefined nature of empathy ultimately manifests in poor operationalization and may contribute to inconsistency in how empathy is measured from observations, resulting in poor construct validity (Coll et al., 2017). With unstable validity, these works may not be deemed reliable for clinical applications, such as counselor training, which has been noted as a critical application of artificial intelligence to the field of psychotherapy (Imel et al., 2017). Aside from such applications, poor validity and lack of definitions lead to reproducibility issues
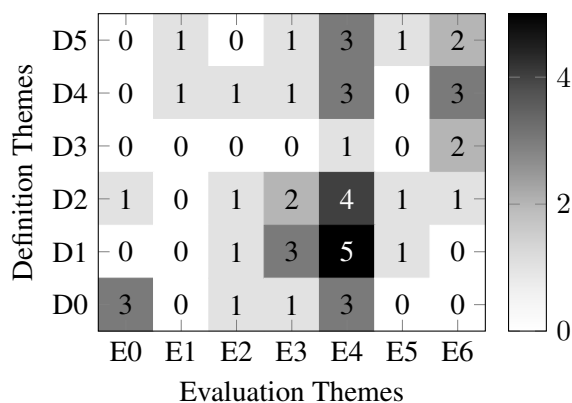

Figure 1: Heatmap of definition and evaluation themes.

(Cuff et al., 2016). Lacking a shared understanding and robust conceptualization of empathy makes it difficult to interpret findings and compare studies.

5 The Measurement Problem

Given the overwhelmingly abstract portrayal of empathy, the effectiveness and validity of our approaches for measuring, annotating, and evaluating it are questionable. Simply put, we do not know if we are investigating the same thing or doing so consistently. Following, we outline where NLP needs to improve in its measurements to move forward.

Construct validity implications for resource construction and evaluation. Yalçın (2019) reviews several scales developed in psychology and suggests potential adaptations for evaluating empathy in NLP. However, issues with measurement, construct, and predictive validity also persist in psychology (Ickes, 2001) and reflect similarly in the limitations of NLP research, e.g., ambiguous evaluation (Coll et al., 2017).

Established psychological scales can help measure empathy, such as the often used Empathic Concern and Personal Distress Scale (Batson et al.,

1987; Buechel et al., 2018). One could, for instance, ask participants to read a passage written by an individual and ask the reader about their emotions. However, various non-empathetic factors can affect how someone feels after reading, not resulting from perspective-taking or understanding the target's emotional state. Psychological studies that use these scales devise experimental controls to promote an other-focus state (Fabi et al., 2019; Batson et al., 1997). One method is to instruct participants to "imagine how the person ... feels about what has happened and how it has affected his or her life" (Toi and Batson, 1982). Another is to control how similar the observer perceives themselves to be to the target (Batson et al., 1981). Self-report measures such as the Davis IRI scale (Davis, 1983) can measure empathic capabilities, but empathy still varies across situations (Litvak et al., 2016; Cuff et al., 2016).

Empathy involves an accurate understanding of another's mental state. As such, one can compare descriptions given by an observer and a target (Ickes, 2001). These first-person assessments are more accurate than annotations by a third party attempting to judge mental states from language, behaviors, or situational factors.

Empathy and domain shift: What exactly are we trying to transfer? The scarcity and diversity of existing datasets motivate investigating knowledge transfer between them (Wu et al., 2021b; Lahnala et al., 2022). Domains vary across datasets, but datasets also vary in how they capture empathy. These differences make larger-scale efforts that combine datasets more difficult. Some studies leverage heuristics to curate empathy data (see Appendix B), such as bootstrapping with pretrained models, selecting interactions from particular contexts (e.g., particular subreddits), and crowdsourcing conversations grounded on particular emotions. While heuristic approaches can help mitigate curation costs (Hosseini and Caragea, 2021b; Welivita et al., 2021; Wu et al., 2021b), the effectiveness is subject to their validity.

In a study motivated to minimize annotation effort through domain transfer, for instance, Wu et al. (2021b)'s experimental results demonstrated the heuristically curated datasets, EMPATHETIC DIALOGUES and PEC (Zhong et al., 2020), were *insufficient* for predicting empathy in the Motivational Interviewing domain. They suggest that more fine-grained empathy annotation labels would

help smooth the domain gaps by distinguishing the empathy aspects expected to be present. Such efforts are needed for the community to build off of each other's work. Meanwhile, further investigations of knowledge transfer techniques between existing resources could have positive contributions.

6 The Narrow Task Problem

While empathy is widely recognized as a construct NLP should be modeling, we argue that the field has focused on a narrow set of tasks that have held back progress. By bridging current work with a perspective of cognitive empathy, we motivate a series of new and reimagined tasks that would advance the field's ability to model empathy.

Empathy is more than emotions. The predominant empathy conceptualization requires the observer to understand the target's emotions and respond appropriately. We recommended that researchers disambiguate sympathy, compassion, and tenderness and view these as subareas of empathy research. However, we argue that this research would still suffer by ignoring cognitive empathy and its interaction with emotion understanding and empathetic responses. Some perspectives from psychology emphasize that the distinction between emotional and cognitive empathy is less important than the interaction between the two components (Cuff et al., 2016)–and we can draw inspiration from that. Cognitive empathy processes are necessary considerations that can more effectively achieve goals for the emotional empathy subareas.

Cognitive empathy processes can improve methods for *understanding emotions*. In NLP, we often consider emotions first, assuming they can be inferred directly from what the target expresses. However, the target could minimize expressions, making the emotions harder to perceive (refer to the discussion on *empathic accuracy* §2.2). Furthermore, most papers only operate on text (e.g., transcripts), which misses opportunities for empathetic cues in other modalities.

Cognitive processes help with better emotional understanding yielding higher empathic accuracy of inferences about the target's affective state. Cuff et al. (2016) argue that the observer can *infer* emotionality through perspective-taking, imagination, and retrieval of relevant memories. In this way, a cognitive approach first leads to an understanding and more accurate perception of emotions.

We argue that cognitive empathy can benefit

from approaches such as common sense reasoning (Sabour et al., 2021), external knowledge (Li et al., 2020b), and abductive NLI (Bhagavatula et al., 2020). For example, Shen et al. (2021) integrated external knowledge to improve counselor reflection generation. Tu et al. (2022) presented an approach to understanding the target's emotional state using commonsense knowledge that improves over more emotion-focused models. Their approach was inspired by Ma et al. (2020)'s survey of empathetic dialogue systems, which argued that future work should go beyond emotional empathy by pursuing personalization and knowledge (both contextual and general external knowledge. Integrating these types of knowledge supports *reasoning about emotion cause*, as opposed to only recognizing emotions, which itself is insufficient (Gao et al., 2021; Kim et al., 2021). Models that incorporate reasoning about the cause of emotion were shown to outperform emotion recognition and mirroring counterparts, notably including those that included external commonsense and emotional lexical knowledge, by human evaluated empathy scores, i.e. (Lin et al., 2019; Li et al., 2020a; Majumder et al., 2020).

Studies may even be able to specify and investigate what type of information is necessary in order to understand the user's emotions and model empathically accurate responses. This can be done with a controlled selection of samples from knowledge bases. For instance, (Shen et al., 2020, 2022)'s work uses particular aspects of common sense and domain-specific knowledge. Researchers must become more intimate with data and understand the nature of dialogue to design empathetic systems.

Cognitive empathic processes can improve methods to select *appropriate response* strategies. Approaches inspired by cognitive processes could thus result in not only valuable representations of emotion but also enhanced methods for response strategies. Though we critiqued the generally-lacking specificity about response appropriateness earlier, the most salient idea we grasp is that observers should mirror the target's emotion or express a similar sentiment valence. Contrary to this idea, Xie and Pu (2021) find that empathetic listeners often respond to negative emotions such as sadness and anger with questions rather than expressing similar or opposite emotions. In addition, specific empathetic question intents of observers may play more significant roles in regulating the tar-

get's emotions (Svikhnushina et al., 2022). Approaches designed both on emotional and cognitive schemes also had effective results in assisting students write peer-reviews that are perceived more empathetic (Wambsganss et al., 2021, 2022).

7 Refocusing our efforts

We overlook research opportunities that better align with described motivations, and these problems necessitate cognitive empathy. Most empathy research in NLP focuses on emotional chatbots or social media analysis. This focus probably drives the overemphasis on emotional empathy. Neither of the domains, we argue, is helpful for the purposes where empathetic NLP systems would be needed the most; for clinical or educational praxis or other rather formal contexts. Such scarce applications include research on Motivational Interviewing (Pérez-Rosas et al., 2016, 2017, 2018, 2019; Wu et al., 2021b; Shen et al., 2022), assistive writing systems for peer-to-peer mental health support (Sharma et al., 2021, 2022), and other counseling conversations (Althoff et al., 2016; Zhang and Danescu-Niculescu-Mizil, 2020; Zhang et al., 2020; Sun et al., 2021). Work in these areas make the need for cognitive empathy even more apparent.

Empathic capacities vary across individuals, as some experience more intense affective responses to certain target perceptions than others (Eisenberg, 2014). Without techniques to manage affective responses, people can experience significant distress. Cognitive empathy skills strengthen observers' ability to regulate or manage their affective responses. Thus, these skills are critical for individuals in caregiving roles (e.g., counselors and doctors) who interpersonally engage with others who are vulnerable, suffering, or in crisis daily. At the same time, the ability to empathize with the target is essential for their roles in providing effective treatment, making the role of cognitive empathy significant both for self-care and care for others.

NLP research is highly needed for *virtual standardized patient* (VSP) systems. VSPs provide an effective way for learners to develop cognitive empathy and clinical skills (Lok and Foster, 2019). The need for VSPs has grown from the necessity for methods that allow learners to practice rare and realistic crisis scenarios and concern for the safety and expenses of standardized patients. Needs in this area include but are not limited to conversational models for practice in intercultural commu-

nication, end-of-life discussions, and breaking bad news. Development of such systems naturally requires the skills of NLP researchers, and these research efforts would benefit the field.

So far, the task of modeling and simulating the empathy target has received little attention in NLP, leaving a significant research gap that our community is positioned to fill, with exciting challenges that align with the motivations of research on empathy and that can have a broad impact within and beyond our field. One work, for instance, focused on simulating individuals in crisis for the use case of counselor training (Demasi et al., 2019). Similarly, modeling and simulating the target could support broader educational objectives, such as training students' emotionally and cognitively empathic feedback skills (Wambsganss et al., 2021). This direction of simulating a mental health support seeker aligns with the goals of related work for developing tools to assist with evaluating and training counselors (Pérez-Rosas et al., 2018, 2019; Imel et al., 2017; Zhang and Danescu-Niculescu-Mizil, 2020). The ethics section further discusses our perspective on this approach compared to developing empathetic response generation models for support-related motivations.

8   A Forward Perspective

Here we summarize the main obstacles we identified and our recommendations for moving forward.

The Definition Problem: Define research goals and design methodologies based on empathy's more concrete and measurable aspects. The abstractness of empathy has negative implications for the construct validity of measurement approaches and, thereby, the scientific effectiveness and comparability.

The Measurement Problem: Draw inspiration from measurements established in psychology, and in addition, learn from investigations and critiques on their construct validity. As well as referencing psychology literature to aid the developments of measurement techniques (Yalçın, 2019), we should be familiar with existing evaluations and discussions on their construct validity. Future work is needed to investigate the validity of our current approaches methodologically.

The Narrow Task Problem: Lessons from our own field. The interaction between cognitive and emotional processes is significant. Cognitive processes can enable higher empathic accuracy. Perspective-taking is a method of cognitive empathy that enables better empathic understanding, and the appraisal theory of empathy provides a framework specific to situational aspects an observer can consider. These processes relate clearly to reasoning processes explored by NLP tasks. We recommend future work that intersects with the newer area of *abductive commonsense reasoning* (Bhagavatula et al., 2020).

Refocusing our efforts: Supporting clinical and educational domains aligns better with the motivations of the empathy research area. Much of what motivates the empathy research area is a desire to support those in need (i.e., *compassionate* motivations, to employ the new terminology). Our position is that current efforts in empathetic dialogue generation should be reallocated to support the needs of clinical domains, e.g., systems that support communicative skill training for people in care-providing roles and systems for supportive cognitive empathic skills development in-general.[3] The need for cognitive empathy is even more pressing for these applications, and therefore NLP research can best support these needs by emphasizing the cognitive aspects in our empathy conceptualizations going forward.

9   Conclusion

Language technology is deployed increasingly in interpersonal settings that require an empathetic understanding of the speaker. Our field's construct represents a significant obstacle to advancing our ability to develop empathetic language technologies. While multiple NLP works have attempted to incorporate empathy into their models and applications, we argue that these works have underspecified empathy—focusing primarily on emotion—and overlooked a major component: cognitive empathy. However, we argue that this gap represents a significant opportunity for developing new models that better reflect cognitive processes and theory of mind while supporting much-needed human-centered applications, such as those in clinical settings.

Ethical Perspectives

While our paper is largely an argument for increased depth and specificity in the study of NLP,

---

[3]Ethical perspective is provided in the ethics section.

our call to overcome these obstacles still comes with ethical considerations. Following, we outline two main ethical points.

The conversational settings used for empathetic communication often feature highly personal dialog from the target that require special care, e.g., those in the medical domain. Advocating for more work in empathy comes at a potential risk to these targets in keeping their potentially-sensitive data private outside of valid uses. The sensitive nature of the data also likely increases the costs and difficulty of developing empathy resources, rendering such efforts rather infeasible for many (Hosseini and Caragea, 2021b; Welivita et al., 2021). Furthermore, some datasets cannot be made public to uphold license agreements to protect the rights and privacy of the stakeholders, which is especially the case in counseling domains (Althoff et al., 2016; Pérez-Rosas et al., 2017; Sharma et al., 2020; Zhang and Danescu-Niculescu-Mizil, 2020; Zhang et al., 2020). While these efforts to preserve privacy are rightly stringent, they also create a two-tier system where only researchers who have access to subjects can use the data to participate in such research. Thus, we call for consideration of possible initiatives to enable researchers to utilize datasets crafted for domains beyond publishable social media data. For instance, shared tasks (such as WASSA 2021 and 2022 (Tafreshi et al., 2021; Barriere et al., 2022)) may enable many to experiment on carefully crafted datasets. Initiatives may draw inspiration from recent CLPsych shared task models that enabled approved researchers to work on sensitive data under signed data use and ethical practice agreements.[4]

Prior work has, for the most part, assumed that empathy must be prosocial and, therefore, improved empathy models would offer societal benefits. However, emotional and cognitive understanding is not always employed for prosocial purposes. For instance, understanding emotions can be used for abuse, and manipulation (Hart et al., 1995; Hodges and Biswas-Diener, 2007). Empathy is also capable of motivating hostile acts and fueling hostility toward out-groups (Breithaupt, 2012, 2018); or, if an observer engages with a target with whom they have a bad relationship, empathy for the target's negative experiences may lead to "malicious gloating" on the observer's behalf (Bischof-Köhler, 1991, p. 259). Thus, the development of new em-

pathetic systems and data could lead to uses that are decidedly antisocial. Consider also the role of empathy in persuasion. In a prosocial context, one study on advice-giving robots showed that students were more likely to be persuaded by the advice of the robot when it used empathetic strategies (Langedijk and Ham, 2021). Emotional appeals are an effective rhetorical tool for persuasion, such as through personal narratives that can increase understanding of other perspectives (Wachsmuth et al., 2017; Vecchi et al., 2021). But on that end, emotional understanding can also be used for manipulative appeals to emotion. Huffaker et al. (2020) introduced a task of detecting emotionally manipulative language, language intended to induce an emotional reaction in the reader (e.g., fear-mongering rhetoric to spread). Their study reviews how adversarial actors have strategically used emotionally manipulative rhetoric in media manipulation. NLP attention on abstract ideas of empathy could be dedicated to such problems. For instance, a new setup could be imagined to anticipate emotional empathy to emotionally manipulative stimuli or focus more on the target's language, which leads to pathogenic empathy in the observers.

Limitations

We presented theoretical groundwork on empathy from psychology and neuroscience. We provide distinct descriptions of empathy-related concepts in a way we view practical for aligning research methodologies with more precise conceptualizations for the NLP community. The disambiguation approach we presented is informed by our efforts to review and understand multiple perspectives from psychology and neuroscience thoroughly. We ultimately constructed this approach with references to selected studies from fields that helped us to identify meaningful delineations between the concepts. However, there is no single conceptualization of empathy or related concepts from those areas, so our construct will differ from some. We hope our paper as a whole inspires collective efforts from our community to scrutinize the empathy frameworks that ground the empathy research direction with more informed perspectives.

This work presented themes about empathy descriptions and evaluations based on a systematic literature review. The search methodology is limited in that it focuses primarily on papers in the ACL Anthology. Our review of related works in-

---

[4]https://clpsych.org/sharedtask2022/

cludes literature outside of ACL venues, and they are consistent with the themes we identify. However, a more extensive study, including works from a broader set of publication venues using standard survey methodology, would provide a more comprehensive outlook on empathetic technology research.

Acknowledgements

References

Muhammad Abdul-Mageed, Anneke Buffone, Hao Peng, Johannes Eichstaedt, and Lyle Ungar. 2017. Recognizing pathogenic empathy in social media. In *Proceedings of the International AAAI Conference on Web and Social Media*, volume 11, pages 448–451. 19

Aseel Addawood, Rezvaneh Rezapour, Omid Abdar, and Jana Diesner. 2017. Telling apart tweets associated with controversial versus non-controversial topics. In *Proceedings of the Second Workshop on NLP and Computational Social Science*, pages 32–41, Vancouver, Canada. Association for Computational Linguistics. 20

Firoj Alam, Shammur Absar Chowdhury, Morena Danieli, and Giuseppe Riccardi. 2016a. How interlocutors coordinate with each other within emotional segments? In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: Technical Papers*, pages 728–738, Osaka, Japan. The COLING 2016 Organizing Committee. 19

Firoj Alam, Morena Danieli, and Giuseppe Riccardi. 2016b. Can we detect speakers' empathy?: A real-life case study. In *2016 7th IEEE International Conference on Cognitive Infocommunications (CogInfoCom)*. IEEE. 19

Firoj Alam, Morena Danieli, and Giuseppe Riccardi. 2018. Annotating and modeling empathy in spoken conversations. *Computer Speech & Language*, 50. 19, 20

Tim Althoff, Kevin Clark, and Jure Leskovec. 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. 7, 9

Valentin Barriere, Shabnam Tafreshi, João Sedoc, and Sawsan Alqahtani. 2022. WASSA 2022 shared task: Predicting empathy, emotion and personality in reaction to news stories. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, Dublin, Ireland. Association for Computational Linguistics. 1, 9, 20

C Daniel Batson, Bruce D Duncan, Paula Ackerman, Terese Buckley, and Kimberly Birch. 1981. Is empathic emotion a source of altruistic motivation? *Journal of personality and Social Psychology*, 40(2):290. 6

C Daniel Batson, Shannon Early, and Giovanni Salvarani. 1997. Perspective taking: Imagining how another feels versus imaging how you would feel. *Personality and social psychology bulletin*, 23(7):751–758. 6

C Daniel Batson, Jim Fultz, and Patricia A Schoenrade. 1987. Distress and empathy: Two qualitatively distinct vicarious emotions with different motivational consequences. *Journal of personality*, 55(1). 5

Chandra Bhagavatula, Ronan Le Bras, Chaitanya Malaviya, Keisuke Sakaguchi, Ari Holtzman, Hannah Rashkin, Doug Downey, Wen-tau Yih, and Yejin Choi. 2020. Abductive commonsense reasoning. In *8th International Conference on Learning Representations, ICLR 2020, Addis Ababa, Ethiopia, April 26-30, 2020*. OpenReview.net. 7, 8

Vitthal Bhandari and Poonam Goyal. 2022. bitsa_nlp@LT-EDI-ACL2022: Leveraging pretrained language models for detecting homophobia and transphobia in social media comments. In *Proceedings of the Second Workshop on Language Technology for Equality, Diversity and Inclusion*, pages 149–154, Dublin, Ireland. Association for Computational Linguistics. 20

Shweta Bhargava, Srinivasan Janarthanam, Helen Hastie, Amol Deshmukh, Ruth Aylett, Lee Corrigan, and Ginevra Castellano. 2013. Demonstration of the EmoteWizard of Oz interface for empathic robotic tutors. In *Proceedings of the SIGDIAL 2013 Conference*, pages 363–365, Metz, France. Association for Computational Linguistics. 20

Doris Bischof-Köhler. 1991. The development of empathy in infants. *Infant development: Perspectives from German-speaking countries*, pages 245–273. 9

Robert James R Blair. 2005. Responding to the emotions of others: Dissociating forms of empathy through the study of typical and psychiatric populations. *Consciousness and cognition*, 14(4):698–718. 2

Fritz Breithaupt. 2012. A three-person model of empathy. *Emotion Review*, 4(1). 9

Fritz Breithaupt. 2018. The bad things we do because of empathy. *Interdisciplinary Science Reviews*, 43(2). 9

Sven Buechel, Anneke Buffone, Barry Slaff, Lyle Ungar, and João Sedoc. 2018. Modeling empathy and distress in reaction to news stories. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 4758–4765, Brussels, Belgium. Association for Computational Linguistics. 1, 4, 6, 19

Yash Butala, Kanishk Singh, Adarsh Kumar, and Shrey Shrivastava. 2021. Team phoenix at WASSA 2021: Emotion analysis on news stories with pre-trained language models. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 274–280, Online. Association for Computational Linguistics. 20

Laurie Carr, Marco Iacoboni, Marie-Charlotte Dubeau, John C. Mazziotta, and Gian Luigi Lenzi. 2003. Neural mechanisms of empathy in humans: A relay from neural systems for imitation to limbic areas. *Proceedings of the National Academy of Sciences*, 100(9):5497–5502. 20

Ginevra Castellano, Ana Paiva, Arvid Kappas, Ruth Aylett, Helen Hastie, Wolmet Barendregt, Fernando Nabais, and Susan Bull. 2013. Towards empathic virtual and robotic tutors. In *International conference on artificial intelligence in education*. Springer. 20

Kezhen Chen, Qiuyuan Huang, Daniel McDuff, Xiang Gao, Hamid Palangi, Jianfeng Wang, Kenneth Forbus, and Jianfeng Gao. 2021. NICE: Neural image commenting with empathy. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, Punta Cana, Dominican Republic. Association for Computational Linguistics. 19

Yue Chen, Yingnan Ju, and Sandra Kübler. 2022. IUCL at WASSA 2022 shared task: A text-only approach to empathy and emotion detection. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, Dublin, Ireland. Association for Computational Linguistics. 20

Michel-Pierre Coll, Essi Viding, Markus Rütgen, Giorgia Silani, Claus Lamm, Caroline Catmur, and Geoffrey Bird. 2017. Are we really measuring empathy? proposal for a new measurement framework. *Neuroscience & Biobehavioral Reviews*, 83:132–139. 5

Benjamin MP Cuff, Sarah J Brown, Laura Taylor, and Douglas J Howat. 2016. Empathy: A review of the concept. *Emotion review*, 8(2). 2, 5, 6

Mark Davis. 1980. A multidimensional approach to individual differences in empathy. *JSAS Catalog Sel. Doc. Psychol.*, 10. 2, 19

Mark H Davis. 1983. Measuring individual differences in empathy: evidence for a multidimensional approach. *Journal of personality and social psychology*, 44(1). 2, 6, 19

Jean Decety and Philip L Jackson. 2006. A social-neuroscience perspective on empathy. *Current directions in psychological science*, 15(2):54–58. 2

Jean Decety and Claus Lamm. 2006. Human empathy through the lens of social neuroscience. *TheScientificWorldJOURNAL*, 6:1146–1163. 20

Flor Miriam Del Arco, Jaime Collado-Montañez, L. Alfonso Ureña, and María-Teresa Martín-Valdivia. 2022. Empathy and distress prediction using transformer multi-output regression and emotion analysis with an ensemble of supervised and zero-shot learning models. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, Dublin, Ireland. Association for Computational Linguistics. 20

Orianna Demasi, Marti A. Hearst, and Benjamin Recht. 2019. Towards augmenting crisis counselor training by improving message retrieval. In *Proceedings of the Sixth Workshop on Computational Linguistics and Clinical Psychology*, pages 1–11, Minneapolis, Minnesota. Association for Computational Linguistics. 1, 8, 20

Dorottya Demszky, Dana Movshovitz-Attias, Jeongwoo Ko, Alan Cowen, Gaurav Nemade, and Sujith Ravi. 2020. GoEmotions: A dataset of fine-grained emotions. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 4040–4054, Online. Association for Computational Linguistics. 20

Alexandre Denis, Samuel Cruz-Lara, Nadia Bellalem, and Lotfi Bellalem. 2014. Synalp-empathic: A valence shifting hybrid system for sentiment analysis. In *Proceedings of the 8th International Workshop on Semantic Evaluation (SemEval 2014)*, pages 605–609, Dublin, Ireland. Association for Computational Linguistics. 20

Nancy Eisenberg. 2014. *Altruistic emotion, cognition, and behavior (PLE: Emotion)*. Psychology Press. 3, 7

Sarah Fabi, Lydia Anna Weber, and Hartmut Leuthold. 2019. Empathic concern and personal distress depend on situational but not dispositional factors. *PloS one*, 14(11):e0225102. 6

Neele Falk and Gabriella Lapesa. 2022. Reports of personal experiences and stories in argumentation: datasets and analysis. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 5530–5553, Dublin, Ireland. Association for Computational Linguistics. 20

Mauajama Firdaus, Asif Ekbal, and Pushpak Bhattacharyya. 2020. Incorporating politeness across languages in customer care responses: Towards building a multi-lingual empathetic dialogue agent. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 4172–4182, Marseille, France. European Language Resources Association. 1, 19, 20

Tommaso Fornaciari, Federico Bianchi, Debora Nozza, and Dirk Hovy. 2021. MilaNLP @ WASSA: Does BERT feel sad when you cry? In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 269–273, Online. Association for Computational Linguistics. 20

Pascale Fung, Anik Dey, Farhad Bin Siddique, Ruixi Lin, Yang Yang, Dario Bertero, Yan Wan, Ricky Ho Yin Chan, and Chien-Sheng Wu. 2016a. Zara: A virtual interactive dialogue system incorporating emotion, sentiment and personality recognition. In *Proceedings of COLING 2016, the 26th International Conference on Computational Linguistics: System Demonstrations*, pages 278–281, Osaka, Japan. The COLING 2016 Organizing Committee. 20

Pascale Fung, Anik Dey, Farhad Bin Siddique, Ruixi Lin, Yang Yang, Yan Wan, and Ho Yin Ricky Chan. 2016b. Zara the Supergirl: An empathetic personality recognition system. In *Proceedings of the 2016 Conference of the North American Chapter of the Association for Computational Linguistics: Demonstrations*, pages 87–91, San Diego, California. Association for Computational Linguistics. 20

Jun Gao, Yuhan Liu, Haolin Deng, Wei Wang, Yu Cao, Jiachen Du, and Ruifeng Xu. 2021. Improving empathetic response generation by recognizing emotion cause in conversations. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, Punta Cana, Dominican Republic. Association for Computational Linguistics. 7, 18, 19

Soumitra Ghosh, Dhirendra Maurya, Asif Ekbal, and Pushpak Bhattacharyya. 2022. Team IITP-AINLPML at WASSA 2022: Empathy detection, emotion classification and personality detection. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, Dublin, Ireland. Association for Computational Linguistics. 20

Jennifer L Goetz, Dacher Keltner, and Emiliana Simon-Thomas. 2010. Compassion: an evolutionary analysis and empirical review. *Psychological bulletin*, 136(3):351. 2

Alvin I. Goldman. 1993. Ethics and cognitive science. *Ethics*, 103(2):337–360. 20

Bhanu Prakash Reddy Guda, Aparna Garimella, and Niyati Chhaya. 2021. EmpathBERT: A BERT-based framework for demographic-aware empathy prediction. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 3072–3079, Online. Association for Computational Linguistics. 20

Marco Guerini, Sara Falcone, and Bernardo Magnini. 2018. A methodology for evaluating interaction strategies of task-oriented conversational agents. In *Proceedings of the 2018 EMNLP Workshop SCAI: The 2nd International Workshop on Search-Oriented Conversational AI*, pages 24–32, Brussels, Belgium. Association for Computational Linguistics. 20

Yuting Guo and Jinho D. Choi. 2021. Enhancing cognitive models of emotions with representation learning. In *Proceedings of the Workshop on Cognitive Modeling and Computational Linguistics*, pages 141–148, Online. Association for Computational Linguistics. 20

Stephen David Hart, David Neil Cox, and Robert D Hare. 1995. *Hare psychopathy checklist: Screening version (PCL: SV)*. Multi-Heath Systems. 9

Helen Hastie, Mei Yii Lim, Srini Janarthanam, Amol Deshmukh, Ruth Aylett, Mary Ellen Foster, and Lynne Hall. 2016. I remember you! interaction with memory for an empathic virtual robotic tutor. 20

Devamanyu Hazarika, Soujanya Poria, Rada Mihalcea, Erik Cambria, and Roger Zimmermann. 2018a. ICON: Interactive conversational memory network for multimodal emotion detection. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*, pages 2594–2604, Brussels, Belgium. Association for Computational Linguistics. 20

Devamanyu Hazarika, Soujanya Poria, Amir Zadeh, Erik Cambria, Louis-Philippe Morency, and Roger Zimmermann. 2018b. Conversational memory network for emotion recognition in dyadic dialogue videos. In *Proceedings of the 2018 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies, Volume 1 (Long Papers)*, pages 2122–2132, New Orleans, Louisiana. Association for Computational Linguistics. 20

Sara D Hodges and Robert Biswas-Diener. 2007. Balancing the empathy expense account: Strategies for regulating empathic response. *Empathy in mental illness*, pages 389–407. 9

Mahshid Hosseini and Cornelia Caragea. 2021a. Distilling knowledge for empathy detection. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, Punta Cana, Dominican Republic. Association for Computational Linguistics. 20

Mahshid Hosseini and Cornelia Caragea. 2021b. It takes two to empathize: One to seek and one to provide. *Proceedings of the AAAI Conference on Artificial Intelligence*, 35(14):13018–13026. 6, 9, 20

Dou Hu, Lingwei Wei, and Xiaoyong Huai. 2021a. DialogueCRN: Contextual reasoning networks for emotion recognition in conversations. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 7042–7052, Online. Association for Computational Linguistics. 20

Jingwen Hu, Yuchen Liu, Jinming Zhao, and Qin Jin. 2021b. MMGCN: Multimodal fusion via deep graph convolution network for emotion recognition in conversation. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 5666–5675, Online. Association for Computational Linguistics. 20

Jordan S. Huffaker, Jonathan K. Kummerfeld, Walter S. Lasecki, and Mark S. Ackerman. 2020. Crowd-sourced detection of emotionally manipulative language. In *CHI '20: CHI Conference on Human Factors in Computing Systems, Honolulu, HI, USA, April 25-30, 2020*, pages 1–14. ACM. 9

William Ickes. 2001. Measuring empathic accuracy. *Interpersonal sensitivity: Theory and measurement*, 1:219–241. 3, 5, 6

William Ickes. 2011. Everyday mind reading is driven by motives and goals. *Psychological Inquiry*, 22(3):200–206. 3, 20

Tatsuya Ide and Daisuke Kawahara. 2022. Building a dialogue corpus annotated with expressed and experienced emotions. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics: Student Research Workshop*, pages 21–30, Dublin, Ireland. Association for Computational Linguistics. 20

Zac E Imel, Derek D Caperton, Michael Tanana, and David C Atkins. 2017. Technology-enhanced human interaction in psychotherapy. *Journal of counseling psychology*, 64(4):385. 5, 8

Koji Inoue, Divesh Lala, Kenta Yamamoto, Shizuka Nakamura, Katsuya Takanashi, and Tatsuya Kawahara. 2020. An attentive listening system with android ERICA: Comparison of autonomous and WOZ interactions. In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 118–127, 1st virtual meeting. Association for Computational Linguistics. 20

Koji Inoue, Hiromi Sakamoto, Kenta Yamamoto, Divesh Lala, and Tatsuya Kawahara. 2021. A multi-party attentive listening robot which stimulates involvement from side participants. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 261–264, Singapore and Online. Association for Computational Linguistics. 20

Micah Iserman and Molly Ireland. 2017. A dictionary-based comparison of autobiographies by people and murderous monsters. In *Proceedings of the Fourth Workshop on Computational Linguistics and Clinical Psychology — From Linguistic Signal to Clinical Reality*, pages 74–84, Vancouver, BC. Association for Computational Linguistics. 20

Etsuko Ishii, Genta Indra Winata, Samuel Cahyawijaya, Divesh Lala, Tatsuya Kawahara, and Pascale Fung. 2021. ERICA: An empathetic android companion for covid-19 quarantine. In *Proceedings of the 22nd Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 257–260, Singapore and Online. Association for Computational Linguistics. 20

Koichiro Ito, Masaki Murata, Tomohiro Ohno, and Shigeki Matsubara. 2020. Relation between degree of empathy for narrative speech and type of responsive utterance in attentive listening. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 696–701, Marseille, France. European Language Resources Association. 18, 19, 20

Jin Yea Jang, San Kim, Minyoung Jung, Saim Shin, and Gahgene Gweon. 2021. BPM_MT: Enhanced backchannel prediction model using multi-task learning. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 3447–3452, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics. 20

Hamed Khanpour, Cornelia Caragea, and Prakhar Biyani. 2017. Identifying empathetic messages in online health communities. In *Proceedings of the Eighth International Joint Conference on Natural Language Processing (Volume 2: Short Papers)*, pages 246–251, Taipei, Taiwan. Asian Federation of Natural Language Processing. 1, 18, 19

Hyunwoo Kim, Byeongchang Kim, and Gunhee Kim. 2021. Perspective-taking and pragmatics for generating empathetic responses focused on emotion causes. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics. 7, 19

Atharva Kulkarni, Sunanda Somwase, Shivam Rajput, and Manisha Marathe. 2021. PVG at WASSA 2021: A multi-input, multi-task, transformer-based architecture for empathy and distress prediction. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 105–111, Online. Association for Computational Linguistics. 20

Allison Lahnala, Charles Welch, and Lucie Flek. 2022. CAISA at WASSA 2022: Adapter-tuning for empathy prediction. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, Dublin, Ireland. Association for Computational Linguistics. 6, 20

Claus Lamm, C. Daniel Batson, and Jean Decety. 2007. The Neural Substrate of Human Empathy: Effects of Perspective-taking and Cognitive Appraisal. *Journal of Cognitive Neuroscience*, 19(1). 2

Rosalyn M Langedijk and Jaap Ham. 2021. More than advice: The influence of adding references to prior discourse and signals of empathy on the persuasiveness of an advice-giving robot. *Interaction Studies*, 22(3). 9, 20

Bin Li, Yixuan Weng, Qiya Song, Bin Sun, and Shutao Li. 2022. Continuing pre-trained model with multiple training strategies for emotional classification. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, pages 233–238, Dublin, Ireland. Association for Computational Linguistics. 20

Qintong Li, Hongshen Chen, Zhaochun Ren, Pengjie Ren, Zhaopeng Tu, and Zhumin Chen. 2020a. EmpDG: Multi-resolution interactive empathetic dialogue generation. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4454–4466, Barcelona, Spain (Online). International Committee on Computational Linguistics. 7, 19

Qintong Li, Piji Li, Zhumin Chen, and Zhaochun Ren. 2020b. Towards empathetic dialogue generation over multi-type knowledge. *arXiv preprint arXiv:2009.09708*. 7

Zhaojiang Lin, Andrea Madotto, Jamin Shin, Peng Xu, and Pascale Fung. 2019. MoEL: Mixture of empathetic listeners. In *Proceedings of the 2019 Conference on Empirical Methods in Natural Language Processing and the 9th International Joint Conference on Natural Language Processing (EMNLP-IJCNLP)*, pages 121–132, Hong Kong, China. Association for Computational Linguistics. 7, 18, 19

Marina Litvak, Jahna Otterbacher, Chee Siang Ang, and David Atkins. 2016. Social and linguistic behavior and its correlation to trait empathy. In *Proceedings of the Workshop on Computational Modeling of People's Opinions, Personality, and Emotions in Social Media (PEOPLES)*, pages 128–137, Osaka, Japan. The COLING 2016 Organizing Committee. 6, 19

Benjamin Lok and Adriana E. Foster. 2019. *Can Virtual Humans Teach Empathy?*, pages 143–163. Springer International Publishing, Cham. 7

Xin Lu, Yijian Tian, Yanyan Zhao, and Bing Qin. 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, Punta Cana, Dominican Republic. Association for Computational Linguistics. 20

Yukun Ma, Khanh Linh Nguyen, Frank Z. Xing, and Erik Cambria. 2020. A survey on empathetic dialogue systems. *Information Fusion*, 64. 7

Khyati Mahajan and Samira Shaikh. 2019. Emoji usage across platforms: A case study for the charlottesville event. In *Proceedings of the 2019 Workshop on Widening NLP*, pages 160–162, Florence, Italy. Association for Computational Linguistics. 20

Himanshu Maheshwari and Vasudeva Varma. 2022. An ensemble approach to detect emotions at an essay level. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, pages 276–279, Dublin, Ireland. Association for Computational Linguistics. 20

Navonil Majumder, Pengfei Hong, Shanshan Peng, Jiankun Lu, Deepanway Ghosal, Alexander Gelbukh, Rada Mihalcea, and Soujanya Poria. 2020. MIME: MIMicking emotions for empathetic response generation. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 8968–8979, Online. Association for Computational Linguistics. 7, 19, 20

William R Miller and Stephen Rollnick. 2012. *Motivational interviewing: Helping people change*. Guilford press. 3

Jay Mundra, Rohan Gupta, and Sagnik Mukherjee. 2021. WASSA@IITK at WASSA 2021: Multi-task learning and transformer finetuning for emotion classification and empathy prediction. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 112–116, Online. Association for Computational Linguistics. 20

Tarek Naous, Wissam Antoun, Reem Mahmoud, and Hazem Hajj. 2021. Empathetic BERT2BERT conversational model: Learning Arabic language generation with little data. In *Proceedings of the Sixth Arabic Natural Language Processing Workshop*, pages 164–172, Kyiv, Ukraine (Virtual). Association for Computational Linguistics. 19, 20

Tarek Naous, Christian Hokayem, and Hazem Hajj. 2020. Empathy-driven Arabic conversational chatbot. In *Proceedings of the Fifth Arabic Natural Language Processing Workshop*, pages 58–68, Barcelona, Spain (Online). Association for Computational Linguistics. 19

Verónica Pérez-Rosas, Rada Mihalcea, Kenneth Resnicow, Satinder Singh, and Lawrence An. 2016. Building a motivational interviewing dataset. In *Proceedings of the Third Workshop on Computational Linguistics and Clinical Psychology*, pages 42–51, San Diego, CA, USA. Association for Computational Linguistics. 7

Verónica Pérez-Rosas, Rada Mihalcea, Kenneth Resnicow, Satinder Singh, and Lawrence An. 2017. Understanding and predicting empathic behavior in counseling therapy. In *Proceedings of the 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1426–1435, Vancouver, Canada. Association for Computational Linguistics. 1, 7, 9, 18, 19

Verónica Pérez-Rosas, Xuetong Sun, Christy Li, Yuchen Wang, Kenneth Resnicow, and Rada Mihalcea. 2018. Analyzing the quality of counseling conversations: the tell-tale signs of high-quality counseling. In *Proceedings of the Eleventh International Conference on Language Resources and Evaluation (LREC 2018)*, Miyazaki, Japan. European Language Resources Association (ELRA). 7, 8

Verónica Pérez-Rosas, Xinyi Wu, Kenneth Resnicow, and Rada Mihalcea. 2019. What makes a good counselor? learning to distinguish between high-quality and low-quality counseling conversations. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 926–935, Florence, Italy. Association for Computational Linguistics. 7, 8

Vitou Phy, Yang Zhao, and Akiko Aizawa. 2020. Deconstruct to reconstruct a configurable evaluation metric for open-domain dialogue systems. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4164–4178, Barcelona, Spain (Online). International Committee on Computational Linguistics. 19

Yada Pruksachatkun, Sachin R. Pendse, and Amit Sharma. 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, CHI 2019, Glasgow, Scotland, UK, May 04-09, 2019*, page 64. ACM. 20

Shenbin Qian, Constantin Orasan, Diptesh Kanojia, Hadeel Saadany, and Félix Do Carmo. 2022. SURREY-CTS-NLP at WASSA2022: An experiment of discourse and sentiment analysis for the prediction of empathy, distress and emotion. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, Dublin, Ireland. Association for Computational Linguistics. 20

Hannah Rashkin, Eric Michael Smith, Margaret Li, and Y-Lan Boureau. 2019. Towards empathetic open-domain conversation models: A new benchmark and dataset. In *Proceedings of the 57th Annual Meeting of the Association for Computational Linguistics*, pages 5370–5381, Florence, Italy. Association for Computational Linguistics. 1, 18, 19, 20

Stephen Roller, Emily Dinan, Naman Goyal, Da Ju, Mary Williamson, Yinhan Liu, Jing Xu, Myle Ott, Eric Michael Smith, Y-Lan Boureau, and Jason Weston. 2021. Recipes for building an open-domain chatbot. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 300–325, Online. Association for Computational Linguistics. 1, 20

Sahand Sabour, Chujie Zheng, and Minlie Huang. 2021. Cem: Commonsense-aware empathetic response generation. *arXiv preprint arXiv:2109.05739*. 7, 18, 19, 20

Manuela Sanguinetti, Alessandro Mazzei, Viviana Patti, Marco Scalerandi, Dario Mana, and Rossana Simeoni. 2020. Annotating errors and emotions in human-chatbot interactions in Italian. In *Proceedings of the 14th Linguistic Annotation Workshop*, pages 148–159, Barcelona, Spain. Association for Computational Linguistics. 1, 19, 20

João Sedoc, Sven Buechel, Yehonathan Nachmany, Anneke Buffone, and Lyle Ungar. 2020. Learning word ratings for empathy and distress from document-level user responses. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1664–1673, Marseille, France. European Language Resources Association. 20

Simone G. Shamay-Tsoory. 2011. The neural bases for empathy. *The Neuroscientist*, 17(1):18–24. PMID: 21071616. 2, 20

Ashish Sharma, Inna W Lin, Adam S Miner, David C Atkins, and Tim Althoff. 2021. Towards facilitating empathic conversations in online mental health support: A reinforcement learning approach. In *Proceedings of the Web Conference 2021*. 7

Ashish Sharma, Inna W. Lin, Adam S. Miner, David C. Atkins, and Tim Althoff. 2022. Human-ai collaboration enables more empathic conversations in text-based peer-to-peer mental health support. 7

Ashish Sharma, Adam Miner, David Atkins, and Tim Althoff. 2020. A computational approach to understanding empathy expressed in text-based mental health support. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 5263–5276, Online. Association for Computational Linguistics. 1, 9, 18, 19

Lei Shen and Yang Feng. 2020. CDL: Curriculum dual learning for emotion-controllable response generation. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 556–566, Online. Association for Computational Linguistics. 20

Lei Shen, Jinchao Zhang, Jiao Ou, Xiaofang Zhao, and Jie Zhou. 2021. Constructing emotional consensus and utilizing unpaired data for empathetic dialogue generation. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, Punta Cana, Dominican Republic. Association for Computational Linguistics. 7, 18, 19

Siqi Shen, Veronica Perez-Rosas, Charles Welch, Soujanya Poria, and Rada Mihalcea. 2022. Knowledge enhanced reflection generation for counseling dialogues. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 3096–3107, Dublin, Ireland. Association for Computational Linguistics. 7

Siqi Shen, Charles Welch, Rada Mihalcea, and Verónica Pérez-Rosas. 2020. Counseling-style reflection generation using generative pretrained transformers with

augmented context. In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 10–20, 1st virtual meeting. Association for Computational Linguistics. 7

Kurt Shuster, Da Ju, Stephen Roller, Emily Dinan, Y-Lan Boureau, and Jason Weston. 2020. The dialogue dodecathlon: Open-domain knowledge and image grounded conversational agents. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2453–2470, Online. Association for Computational Linguistics. 1, 20

Farhad Bin Siddique, Onno Kampman, Yang Yang, Anik Dey, and Pascale Fung. 2017. Zara returns: Improved personality induction and adaptation by an empathetic virtual agent. In *Proceedings of ACL 2017, System Demonstrations*, pages 121–126, Vancouver, Canada. Association for Computational Linguistics. 20

Eric Michael Smith, Mary Williamson, Kurt Shuster, Jason Weston, and Y-Lan Boureau. 2020. Can you put it all together: Evaluating conversational agents' ability to blend skills. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 2021–2030, Online. Association for Computational Linguistics. 19

Achim Stephan. 2015. Empathy for artificial agents. *International Journal of Social Robotics*, 7(1). 20

Linda Stinson and William Ickes. 1992. Empathic accuracy in the interactions of male friends versus male strangers. *Journal of personality and social psychology*, 62(5):787. 3

E Stotland, KE Mathews Jr, S Sherman, , R.O. Hansson, and B.Z. Richardson. 1978. Empathy, fantasy and helping. *Empathy, fantasy, and helping. Beverly Hills: Sage*. 3

Merlin Teodosia Suarez, Jocelynn Cu, and Madelene Sta. Maria. 2012. Building a multimodal laughter database for emotion recognition. In *Proceedings of the Eighth International Conference on Language Resources and Evaluation (LREC'12)*, pages 2347–2350, Istanbul, Turkey. European Language Resources Association (ELRA). 20

Hao Sun, Zhenru Lin, Chujie Zheng, Siyang Liu, and Minlie Huang. 2021. PsyQA: A Chinese dataset for generating long counseling text for mental health support. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 1489–1503, Online. Association for Computational Linguistics. 7, 20

Ekaterina Svikhnushina, Iuliana Voinea, Anuradha Welivita, and Pearl Pu. 2022. A taxonomy of empathetic questions in social dialogs. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Dublin, Ireland. Association for Computational Linguistics. 7, 19

Shabnam Tafreshi, Orphee De Clercq, Valentin Barriere, Sven Buechel, João Sedoc, and Alexandra Balahur. 2021. WASSA 2021 shared task: Predicting empathy and emotion in reaction to news stories. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 92–104, Online. Association for Computational Linguistics. 1, 9, 20

Miho Toi and C Daniel Batson. 1982. More evidence that empathy is a source of altruistic motivation. *Journal of personality and social psychology*, 43(2):281. 6

Alicia Tsai, Shereen Oraby, Vittorio Perera, Jiun-Yu Kao, Yuheng Du, Anjali Narayan-Chen, Tagyoung Chung, and Dilek Hakkani-Tur. 2021. Style control for schema-guided natural language generation. In *Proceedings of the 3rd Workshop on Natural Language Processing for Conversational AI*, Online. Association for Computational Linguistics. 20

Quan Tu, Yanran Li, Jianwei Cui, Bin Wang, Ji-Rong Wen, and Rui Yan. 2022. MISC: A mixed strategy-aware model integrating COMET for emotional support conversation. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Dublin, Ireland. Association for Computational Linguistics. 7, 19, 20

Lindsey Vanderlyn, Gianna Weber, Michael Neumann, Dirk Väth, Sarina Meyer, and Ngoc Thang Vu. 2021. "it seemed like an annoying woman": On the perception and ethical considerations of affective language in text-based conversational agents. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, Online. Association for Computational Linguistics. 20

Deeksha Varshney, Asif Ekbal, and Pushpak Bhattacharyya. 2021. Modelling context emotions using multi-task learning for emotion controlled dialog generation. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 2919–2931, Online. Association for Computational Linguistics. 20

Himil Vasava, Pramegh Uikey, Gaurav Wasnik, and Raksha Sharma. 2022. Transformer-based architecture for empathy prediction and emotion classification. In *Proceedings of the 12th Workshop on Computational Approaches to Subjectivity, Sentiment & Social Media Analysis*, Dublin, Ireland. Association for Computational Linguistics. 20

Eva Maria Vecchi, Neele Falk, Iman Jundi, and Gabriella Lapesa. 2021. Towards argument mining for social good: A survey. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 1338–1352, Online. Association for Computational Linguistics. 9

Giuseppe Vettigli and Antonio Sorgente. 2021. EmpNa at WASSA 2021: A lightweight model for the prediction of empathy, distress and emotions from reactions to news stories. In *Proceedings of the Eleventh Workshop on Computational Approaches to Subjectivity, Sentiment and Social Media Analysis*, pages 264–268, Online. Association for Computational Linguistics. 20

Henning Wachsmuth, Nona Naderi, Yufang Hou, Yonatan Bilu, Vinodkumar Prabhakaran, Tim Alberdingk Thijm, Graeme Hirst, and Benno Stein. 2017. Computational argumentation quality assessment in natural language. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*, pages 176–187, Valencia, Spain. Association for Computational Linguistics. 9

Thiemo Wambsganss, Christina Niklaus, Matthias Söllner, Siegfried Handschuh, and Jan Marco Leimeister. 2021. Supporting cognitive and emotional empathic writing of students. In *Proceedings of the 59th Annual Meeting of the Association for Computational Linguistics and the 11th International Joint Conference on Natural Language Processing (Volume 1: Long Papers)*, pages 4063–4077, Online. Association for Computational Linguistics. 7, 8, 18, 19, 20

Thiemo Wambsganss, Matthias Soellner, Kenneth R Koedinger, and Jan Marco Leimeister. 2022. Adaptive empathy learning support in peer review scenarios. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22, New York, NY, USA. Association for Computing Machinery. 7

Yan Wang, Jiayu Zhang, Jun Ma, Shaojun Wang, and Jing Xiao. 2020. Contextualized emotion recognition in conversation as sequence tagging. In *Proceedings of the 21th Annual Meeting of the Special Interest Group on Discourse and Dialogue*, pages 186–195, 1st virtual meeting. Association for Computational Linguistics. 20

Anuradha Welivita and Pearl Pu. 2020. A taxonomy of empathetic response intents in human social conversations. In *Proceedings of the 28th International Conference on Computational Linguistics*, pages 4886–4899, Barcelona, Spain (Online). International Committee on Computational Linguistics. 19

Anuradha Welivita, Yubo Xie, and Pearl Pu. 2021. A large-scale dataset for empathetic response generation. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics. 6, 9, 20

Joshua D Wondra and Phoebe C Ellsworth. 2015. An appraisal theory of empathy and other vicarious emotional experiences. *Psychological review*, 122(3):411. 3

Chen Henry Wu, Yinhe Zheng, Xiaoxi Mao, and Minlie Huang. 2021a. Transferable persona-grounded dialogues via grounded minimal edits. In *Proceedings of the 2021 Conference on Empirical Methods in Natural Language Processing*, pages 2368–2382, Online and Punta Cana, Dominican Republic. Association for Computational Linguistics. 19

Zixiu Wu, Rim Helaoui, Diego Reforgiato Recupero, and Daniele Riboni. 2021b. Towards low-resource real-time assessment of empathy in counselling. In *Proceedings of the Seventh Workshop on Computational Linguistics and Clinical Psychology: Improving Access*, pages 204–216, Online. Association for Computational Linguistics. 6, 7, 18, 19

Yubo Xie and Pearl Pu. 2021. Empathetic dialog generation with fine-grained intents. In *Proceedings of the 25th Conference on Computational Natural Language Learning*, Online. Association for Computational Linguistics. 7, 19

Özge Nilay Yalçın. 2019. Evaluating empathy in artificial agents. In *2019 8th International Conference on Affective Computing and Intelligent Interaction (ACII)*, pages 1–7. 5, 8

Clay H. Yoo, Shriphani Palakodety, Rupak Sarkar, and Ashiqur KhudaBukhsh. 2021. Empathy and hope: Resource transfer to model inter-country social media dynamics. In *Proceedings of the 1st Workshop on NLP for Positive Impact*, pages 125–134, Online. Association for Computational Linguistics. 20

Chengkun Zeng, Guanyi Chen, Chenghua Lin, Ruizhe Li, and Zhi Chen. 2021. Affective decoding for empathetic response generation. In *Proceedings of the 14th International Conference on Natural Language Generation*, pages 331–340, Aberdeen, Scotland, UK. Association for Computational Linguistics. 19

Justine Zhang and Cristian Danescu-Niculescu-Mizil. 2020. Balancing objectives in counseling conversations: Advancing forwards or looking backwards. In *Proceedings of the 58th Annual Meeting of the Association for Computational Linguistics*, pages 5276–5289, Online. Association for Computational Linguistics. 7, 8, 9, 18, 19

Justine Zhang, Sendhil Mullainathan, and Cristian Danescu-Niculescu-Mizil. 2020. Quantifying the causal effects of conversational tendencies. *Proceedings of the ACM on Human-Computer Interaction*, 4(CSCW2). 7, 9

Chujie Zheng, Yong Liu, Wei Chen, Yongcai Leng, and Minlie Huang. 2021. CoMAE: A multi-factor hierarchical framework for empathetic response generation. In *Findings of the Association for Computational Linguistics: ACL-IJCNLP 2021*, pages 813–824, Online. Association for Computational Linguistics. 19

Peixiang Zhong, Chen Zhang, Hao Wang, Yong Liu, and Chunyan Miao. 2020. Towards persona-based empathetic conversational models. In *Proceedings of the*

2020 Conference on Empirical Methods in Natural Language Processing (EMNLP), pages 6556–6566, Online. Association for Computational Linguistics. 6, 20

Li Zhou, Jianfeng Gao, Di Li, and Heung-Yeung Shum. 2020. The design and implementation of XiaoIce, an empathetic social chatbot. *Computational Linguistics*, 46(1):53–93. 1, 20

Naitian Zhou and David Jurgens. 2020. Condolence and empathy in online communities. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 609–626, Online. Association for Computational Linguistics. 19

Xianda Zhou and William Yang Wang. 2018. MojiTalk: Generating emotional responses at scale. In *Proceedings of the 56th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1128–1137, Melbourne, Australia. Association for Computational Linguistics. 20

Ling.Yu Zhu, Zhengkun Zhang, Jun Wang, Hongbin Wang, Haiying Wu, and Zhenglu Yang. 2022. Multiparty empathetic dialogue generation: A new task for dialog systems. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, Dublin, Ireland. Association for Computational Linguistics. 18, 19

A Definition Themes

(5) Thorough, theory-based description of empathy. Empathy is extensively described with the inclusion of perceptive insights from psychology, neuroscience, or other related fields. There are clear distinctions between empathy and concepts such as emotion recognition and mirroring, sympathy, and compassion. There may be explicit efforts to describe and distinguish between emotional and cognitive empathy and the significance of the distinction. **Examples:** Sharma et al. (2020); Wambsganss et al. (2021); Sabour et al. (2021).

(4) Description is succinct yet explicitly theorygrounded and adheres to the theoretical foundation. Empathy is explicitly defined or described in a way grounded in a complex psychological perspective. As opposed to **(5)**, longer, more specified descriptions of empathy and its dimensions are not provided. Nevertheless, these works remain consistent with the selected perspective through their analyses and interpretations of their findings. **Example:** Pérez-Rosas et al. (2017).

(3) Thorough descriptions of other concepts and their relations to empathy are consistent with multi-dimensional theories. Empathy itself is not explicitly defined or described. It may be unclear whether the idea is rooted in psychology with no specific references. However, behaviors associated with empathy are described that are consistent with multiple aspects of empathy described in the Theoretical Guide §2. This category can also include studies that do not specifically describe (or explore) the concept of empathy, but relate empathy to other behaviors (e.g., counselor strategies) in a way that is based on psychology literature. There may also be descriptions of concepts that are not specifically conveyed as empathy, but suggest an independent thought process that arrived at perspectives consistent with multiple aspects reviewed in §2. **Examples:** Ito et al. (2020); Zhang and Danescu-Niculescu-Mizil (2020); Wu et al. (2021b); Gao et al. (2021).

(2) Descriptions provided are vague or ambiguous with other concepts. The descriptions may go briefly beyond the "understanding the user and responding emotionally" concept by mentioning other aspects (e.g., perspective-taking) in passing or by describing empathetic responses to be those that mimic or mirror the target's response. Other concepts may be mentioned in a way that lacks separable distinction. Some papers reference a psychology definition but leave one or more (usually cognitive) components untouched in their work (Khanpour et al., 2017). This theme typically emerges when the study derives a conception based on their own reasoning (Shen et al., 2021; Zhu et al., 2022), which could benefit by incorporating terminology and distinctions we provide in §2.

(1) No empathy description but abstract conceptualization may be inferred through task or system description. No definition nor description of empathy or empathetic behaviors is explicitly stated, however some abstract conceptualization of empathy can be inferred through the description of the task or requirements of an empathetic system. This includes describing the empathetic response generation task as "to understand the user emotion and respond appropriately" Lin et al. (2019); Rashkin et al. (2019). Elaboration on the process of understanding the user's emotion in an empathetic way or responding appropriately in an empathetic way are not provided.

(0) No empathy description, and the conceptualization is not inferrable through other information. No definition of empathy is provided, despite the apparent relevance of empathy to the study.

Such works may perform a human evaluation of empathy, and an emotional perspective on empathy may be inferred through a description of the evaluation if provided (Phy et al., 2020). Generally, the conceptualization cannot be inferred. This label does not include works that reflect abstract conceptualization in their task description as in Theme 1. This category includes cases where empathy may be ambiguous with concepts such as "politeness" and "courtesy" (Firdaus et al., 2020).

B Evaluation Themes

Here we categorize papers by how they approached evaluating, labeled, or rated empathy in their research results or when constructing datasets.

Multi-item: cognitive & emotional (8). These studies report manual tasks for labeling or rating multiple items (e.g., behaviors, strategies) of empathy and include cognitive and emotional aspects in these items. This category includes studies that developed a new scheme for labeling or rating response intents and behaviors, or labeled counselor behaviors based on schemes designed for a counseling style. In a setup particularly unique among the NLP literature, one recruited participants to complete a multi-item self-report scale for trait empathy (Davis' Interpersonal Reactivity Index (IRI) (Davis, 1980, 1983)) and then analyzed linguistics behaviors with respect to those scales (Litvak et al., 2016). **All papers:** Litvak et al. (2016); Pérez-Rosas et al. (2017); Sharma et al. (2020); Welivita and Pu (2020); Ito et al. (2020); Zhang and Danescu-Niculescu-Mizil (2020); Wambsganss et al. (2021); Svikhnushina et al. (2022).

Single label/rating: cognitive & emotional empathy (3). These studies report manual tasks for labeling or rating empathy as a single item or score based on multiple items or aspects representing both cognitive and emotional empathy. Two are based on theoretical and practical psychology conceptualizations (Appraisal Theory of Empathy (Zhou and Jurgens, 2020) and MISC (Wu et al., 2021b)). The other is based on a description of empathy that includes intents or acts by the observer to help regulate the target's emotions (Sanguinetti et al., 2020). **All papers:** Zhou and Jurgens (2020); Sanguinetti et al. (2020); Wu et al. (2021b).

Single label/rating: emotional empathy (19). These studies report manual tasks for labeling or rating empathy as a single item or score based on

a description of empathy (reported to be provided to the annotators) representing only emotional aspects of empathy. Task setups include comparisons between two items (which item is more empathetic based on the provided description of empathy), binary labels (empathetic or not), and ratings on different Likert scales. The typical descriptions of empathy provided are defined by whether or the degree to which the observer item shows, demonstrates, or expresses an ability to infer or an understanding/awareness of the target's emotions or feelings. Often these descriptions include that the observer should respond in a way that is appropriate, emotionally or otherwise, without guidelines on appropriate vs. inappropriate empathetic responses. Some descriptions refer to emotion sharing, such as saying the observer should manifest, share, or experience the target's emotions (Zhu et al., 2022). Some studies appear to use sympathy and empathy interchangeably (Rashkin et al., 2019; Lin et al., 2019). We determined this category to be the best fit for Buechel et al. (2018)'s study, in which the single empathy ratings are based on a multi-item questionnaire focused on emotions. **All papers:** Alam et al. (2016b); Alam et al. (2016a); Alam et al. (2018); Buechel et al. (2018); Rashkin et al. (2019); Lin et al. (2019); Smith et al. (2020); Majumder et al. (2020); Naous et al. (2020); Phy et al. (2020); Li et al. (2020a); Zeng et al. (2021); Wu et al. (2021a); Shen et al. (2021); Xie and Pu (2021); Sabour et al. (2021); Zheng et al. (2021); Naous et al. (2021); Zhu et al. (2022).

Single label/rating: no specification (8). These studies refer to or report a manual task performed to label or rate empathy without a clear description of empathy upon which the task was based. Task setups include asking annotators to label items as "empathic responses" and to compare two items of observer text (empathetic response generation output) for which is "more empathetic." Four of these studies ask annotators to rate empathy on Likert scales (3-point, 4-point, 5-point, and 7-point). A few studies indicated they recruited annotators with linguistics or psychology backgrounds (Tu et al., 2022). One study describes a procedure for familiarizing the annotators with the concept based on selected papers from psychology and unifying their understanding with the help of psychologists (Khanpour et al., 2017). **All papers:** Abdul-Mageed et al. (2017); Khanpour et al. (2017); Kim et al. (2021); Gao et al. (2021); Chen et al. (2021);

Ishii et al. (2021); Jang et al. (2021); Tu et al. (2022).

Heuristic empathy labels or ratings (4). These studies include heuristic methods to gather or label empathetic data automatically. One study created an "empathy lexicon" based on associations with single-item empathy ratings on a prior dataset (Sedoc et al., 2020). Another built a dataset by training a model on data labeled with scheme of empathetic response intents (Welivita et al., 2021). One study selected conversations from two subreddits (r/happy and r/offmychest) and manually labeled samples from them and a control group (r/CausalConversations) as empathetic or non-empathetic (1 or 0), and compared the averages of the selected subreddits to the control to demonstrate that they were more empathetic on average (Zhong et al., 2020). Another considered an "empathetic language style" to be the language style of "more empathetic" of two Myers-Briggs Type Indicator (MBTI) personality types (Vanderlyn et al., 2021). We labeled Rashkin et al. (2019)'s study based on their human evaluation of the empathetic response generation model. However, we also consider their data curation, in which crowdworkers have conversations grounded on specific emotions, to be a heuristic-based approach. **All papers:** Sedoc et al. (2020); Zhong et al. (2020); Welivita et al. (2021); Vanderlyn et al. (2021).

Role labeling (2). Instead of labeling, rating, or associating behaviors with empathy, these studies label target and observer roles (seeker vs. provider (Hosseini and Caragea, 2021a) and seeking empathy vs. providing empathy (Hosseini and Caragea, 2021b)). **All papers:** Hosseini and Caragea (2021b); Hosseini and Caragea (2021a).

Only automatic or no manual evaluation (4). This category includes studies that report empathetic systems or results without reporting manual evaluation (Siddique et al., 2017; Zhou et al., 2020; Firdaus et al., 2020) or that evaluate empathy automatically only (Tsai et al., 2021). **All papers:** Siddique et al. (2017); Zhou et al. (2020); Firdaus et al. (2020); Tsai et al. (2021).

Buechel/WASSA (15). All papers: Fornaciari et al. (2021); Tafreshi et al. (2021); Butala et al. (2021); Vettigli and Sorgente (2021); Mundra et al. (2021); Kulkarni et al. (2021); Guda et al. (2021); Qian et al. (2022); Chen et al. (2022); Del Arco et al. (2022); Vasava et al. (2022); Ghosh et al. (2022);

Lahnala et al. (2022); Barriere et al. (2022).

Not categorized (38). Not empathy task or does not evaluate. All papers: Suarez et al. (2012); Castellano et al. (2013); Bhargava et al. (2013); Denis et al. (2014); Fung et al. (2016a); Hastie et al. (2016); Fung et al. (2016b); Addawood et al. (2017); Iserman and Ireland (2017); Hazarika et al. (2018a); Hazarika et al. (2018b); Guerini et al. (2018); Zhou and Wang (2018); Mahajan and Shaikh (2019); Demasi et al. (2019); Demszky et al. (2020); Shen and Feng (2020); Shuster et al. (2020); Inoue et al. (2020); Wang et al. (2020); Roller et al. (2021); Hu et al. (2021a); Varshney et al. (2021); Yoo et al. (2021); Inoue et al. (2021); Guo and Choi (2021); Lu et al. (2021); Hu et al. (2021b); Li et al. (2022); Maheshwari and Varma (2022); Falk and Lapesa (2022); Ide and Kawahara (2022); Bhandari and Goyal (2022); Stephan (2015); Langedijk and Ham (2021); Pruksachatkun et al. (2019); Sabour et al. (2021).

C Emotion recognition, contagion, and mimicry

Two underlying processes of emotional empathy, emotion recognition, and contagion, are closely related (Shamay-Tsoory, 2011). *Emotional contagion* ("catching feelings") refers to when an observer's brain activates similarly to a target's (e.g., perception of pain), where the observer lacks awareness of the origin of their emotion (Decety and Lamm, 2006; Ickes, 2011). Research on the human mechanism of imitating an affective state (Goldman, 1993; Carr et al., 2003) relates to some response generation approaches in NLP (Majumder et al., 2020). However, mimicking emotions without experiencing contagion is a separate concept referred to as "mimpathy" (Ickes, 2011).

D Non-English Datasets

Arabic Naous et al., 2021, Italian Alam et al., 2018; Sanguinetti et al., 2020, Chinese Sun et al., 2021, Japanese Ito et al., 2020; Sanguinetti et al., 2020, German Wambsganss et al., 2021.