Empathy Practices in Social Media Discourse: A Multidimensional and Relational Perspective

Yixin Chen
yixin.chen@oii.ox.ac.uk
Oxford Internet Institute, University of Oxford
Oxford, United Kingdom

Bernie Hogan
bernie.hogan@oii.ox.ac.uk
Oxford Internet Institute, University of Oxford
Oxford, United Kingdom

Dr Scott A. Hale
scott.hale@oii.ox.ac.uk
Oxford Internet Institute, University of Oxford
Oxford, United Kingdom
Meedan
San Francisco, California, United States

Abstract

Empathy is widely regarded as an inherently positive feature of supportive online interactions, but its value is shaped by context. In this study, we argue that empathy should be understood not as a uniform good but as a multidimensional and relational practice. Rather than treating empathy as a binary or unidimensional attribute, we propose a new framework that captures how empathy is solicited in posts and how it is expressed in replies, emphasising that context is critical in determining its appropriateness and effectiveness. Using post–reply data from six communities across Reddit and Stack Exchange, we conduct a three-phase study. First, we develop a fine-grained annotation framework grounded in psychological theory to capture distinct empathy practices in both opening posts and replies. Second, we fine-tune language models to detect these practices, showing that they can effectively capture the distinct empathy practices in our schema. Third, we apply the models at scale and conduct an empirical analysis to examine platform- and community-specific patterns in how empathy is both elicited and expressed. Our findings challenge current assumptions about online empathy and offer a more contextualised understanding of its role in online discourse. We identify future directions for platform design and contextual support for community members.

CCS Concepts

• **Human-centered computing → Collaborative and social computing**.

Keywords

Empathy, Online Communities, Language Models, Fine-Tuning, Thematic Analysis, Reddit, Stack Exchange, Emotional Empathy, Cognitive Empathy, Empathic Concern

ACM Reference Format:
Yixin Chen, Bernie Hogan, and Dr Scott A. Hale. 2026. Empathy Practices in Social Media Discourse: A Multidimensional and Relational Perspective. In *Proceedings of the 2026 CHI Conference on Human Factors in Computing Systems (CHI '26), April 13–17, 2026, Barcelona, Spain.* ACM, New York, NY, USA, 25 pages. https://doi.org/10.1145/3772318.3790967

CHI '26, Barcelona, Spain

ACM ISBN 979-8-4007-2278-3/2026/04
https://doi.org/10.1145/3772318.3790967

1 Introduction

Is it possible to have too much of a good thing? Empathy is often seen as a positive force and something to be maximised in social psychological research and computer-mediated communication [e.g., 10, 38, 86, 106]. Yet, different contexts may require different kinds of empathetic responses, some of which are potentially situationally inappropriate or considered insincere. In competitive or efficiency-driven spaces, empathy may even be perceived as a distraction [72, 99]. It can also carry material [8], emotional [1, 16], and cognitive costs [17, 70]. These concerns raise a fundamental question: when and how should empathy be desired in online settings? Rather than treating empathy as inherently positive, we argue platform designers and community leaders need to consider context and should seek to elicit empathy from their users judiciously.

Current work on identifying online empathy typically treats it as a binary label or a simple scoring dimension [42, 54]. This limits our understanding of its value and effectiveness across different contexts. Therefore, this study starts from the idea that the effects of empathy are not fully understood and may depend on how it is communicated, perceived, and received. In this work, we propose a new framework for studying empathy practices in online communities based on post–reply data from two major online community platforms, Reddit and Stack Exchange. Our study proceeds in three linked phases: (1) defining and labelling empathy practices with a level of granularity that captures their distinct conceptual dimensions, (2) building and evaluating models that can reliably identify these multidimensional and relational empathy practices in posts and replies, consistent with human annotations, and (3) applying these models at scale to measure how empathy practice patterns vary across platforms and communities.

We focus on two platforms (Reddit and Stack Exchange) and three communities apiece (r/Parenting, r/socialskills, r/work, and Parenting Stack Exchange, Interpersonal Skills Stack Exchange, The Workplace Stack Exchange). These sites were chosen to examine empathy-warranting requests and empathy-containing replies in comparable, everyday contexts where empathy is plausible but not structurally mandated. The results show that empathy practices on these sites can be intelligibly coded into a small number of themes (six for seeking, seven for giving), each of which resonates with existing psychological theories while offering considerably more granularity than prevailing labelling approaches. We describe these as "empathy practices" rather than "empathy" on account of empathy's inherent relational dynamics [60, 91, 102]. To be empathetic is to signal to an observer that one is taking the observer's context and circumstances into account in one's communication [12]. This implies both an observer (typically a person who directly seeks a response that will account for the person's context and circumstances) and a respondent (who accounts for these circumstances in their communication).

These relational dynamics are not always accounted for in the analysis of empathy in textual communication. Prior work focuses primarily on the response out of context, for example by labelling a tweet or a comment with an empathy score [107] or extracting lexical or semantic features from text for scoring [42]. Even when scores are multidimensional, they rarely consider different approaches to the solicitation of empathy. Such differences in solicitation can be illustrated as follows: *"I'm wondering which car seat would be the best for a two year old. I'm trying to stay in budget. Any ideas?"* versus '*'My two year old hates his new car seat and makes driving a chore, I can't afford a new one. Any ideas?"*. Offering extensive emotional sympathies in the first case might be seen as untoward (e.g., *"I'm so sorry to hear of your budget restrictions…"*). Yet, in the second case such sympathies might be warranted (e.g., *"I know how hard it is to drive with a screaming toddler…"*).

Merely scoring responses misses the potential contextual triggers that align the responses (replies) with the original requests (opening posts). Thus, this study examines not only how individuals express empathy in their replies, but also how the need for empathy, whether direct or indirect, is conveyed in opening posts and how this influences the replies received. By doing so, we aim to understand not just whether empathy is present, but when it is solicited, how it is interpreted, and whether it meets the needs of the poster.

We also present a robust method for identifying empathy in online discourse. Although several computational approaches have been proposed to detect empathy [e.g., 42, 47, 54, 83, 98, 107], these often lack specificity and granularity, overlook relational dynamics, exhibit limited contextual awareness, and have a narrow scope of analysis. Language models (LMs) have shown promise in capturing nuanced social and psychological signals [26, 59, 104], especially when fine-tuned [41, 58, 80]. Building on this, we fine-tune LMs to detect empathy in online conversations and evaluate their effectiveness in capturing both requests and responses. Our findings on in-context annotations are inconclusive but encouraging, i.e., for only some practices they significantly help with classification.

We then demonstrate the value of our models in an empirical case study. We applied the best-performing models to a substantially larger corpus, and examined platform- and community-specific patterns in empathetic interactions. We focus on the relationships between request and response types, how prevalent these types were across the six communities of interest, and which empathy types were associated with further responses from the original poster (OP). We see clear but heterogeneous differences in empathy practices across online communities despite superficial topic similarity. Our results lead to a discussion about contextual considerations for empathy as well as insights into the applicability and limitations of using language models for this task at scale.

We consolidate these findings in a discussion of implications for design and practice in online collaborative spaces, while also noting plausible future directions for the application of this approach. In particular, we focus on the potential for a more granular application of empathy that accommodates its relational dimensions and contextual differences. We believe a scalable approach to the granular detection of empathy practices will enable researchers and analysts to probe the health of online communities and develop context-specific strategies that resonate with the expectations of community members or chat participants. Additionally, this approach could help with onboarding for newcomers seeking effective contextual support.

The main contributions of this paper are as follows. (1) We develop a comprehensive framework that captures the multidimensional and relational nature of both requests for and expression of online empathy, moving beyond simplistic assumptions that empathy is inherently positive or negative. (2) We annotate and release a dataset of Reddit and Stack Exchange posts and replies annotated for dimensions of empathy as well as our fine-tuned LMs, showing that, when guided by clearly defined themes, LMs can effectively detect empathy dimensions in requests and responses. (3) We demonstrate the value of our models by applying them to a large-scale dataset and showing platform- and community-specific patterns in empathetic interactions. Rather than treating empathy as something to be universally maximised, our findings highlight the importance of context in shaping the presence, form, and potential impacts of empathy.

2 Related Work

2.1 Theoretical Framework of Empathy

Understanding the effects of empathy across different contexts requires a thorough examination of its theoretical underpinnings. Empathy can be viewed as a personal trait [23, 25, 29] or context-specific state [9, 66]. This study adopts the state perspective of empathy; so, we focus on the expression of empathy in specific online conversations. Different from in face-to-face encounters, empathy is communicated asynchronously via text in the online communication we study [67, 73]. Nevertheless, we can still draw on the theoretical framework developed in broader settings.

Prior theoretical and empirical studies primarily in psychology capture the multidimensional nature of empathy by deconstructing the concept and considering multiple components. The first component is emotional empathy, which focuses on the emotional process of reflecting or matching the stimulus person's emotional state [32, 40, 44]. The second component is cognitive empathy, which is defined as the intellectual process of understanding the stimulus person's perspective [46, 95]. Some research also considers a third component, referred to as empathic concern or the motivational component, which describes the desire to care for or improve the stimulus person's welfare [31, 33, 68, 86, 103]. These components are not entirely exclusive but emphasise different dimensions or aspects of empathy, and all contribute to a comprehensive understanding of the concept [30, 37]. This study also focuses on these three components of empathy.

Empathy is relational. Empathy practices involve not only the observer who expresses empathy but also the original speaker who initiates the interaction. A request for empathy is referred to as "a need to be heard and understood" [47]. People may use strategies, either intentionally or unintentionally, to connect with their target audience and elicit the empathetic responses they desire. Disclosure of personal stories or experiences can be considered as an implicit or indirect way to call for listening and empathy [19, 71, 89]. Self-disclosure typically falls into two categories: informational disclosure, which involves sharing personal information, and emotional disclosure, which involves sharing emotions towards someone or something [2]. Furthermore, some people intentionally employ vagueness in their disclosures through coded cues to prevent potential judgement by those who do not understand their experiences and to signal to those who have similar experiences and are more likely to be supportive and empathetic [6]. Therefore, this study focuses on two major aspects from the original speakers' perspective: self-disclosure and requests.

While psychological studies have produced an expanded theoretical framework for empathy, its application to online text-based interactions has been somewhat superficial by comparison, using only one or a small number of dimensions (discussed further below). We nonetheless assert that a deeper investigation of each dimension within this framework is warranted to better understand the effects of empathy.

2.2 Challenges in Measuring Online Empathy Practices

Previous studies have explored various approaches to identifying and assessing empathy in online interactions. Table 1 summarises key studies on this topic and highlights their platforms, interaction types, empathy constructs, methods used, and model performance.

Empathy is often operationalised as a single construct in evaluation systems. Much of the early work exemplifies this view by concentrating on characteristic linguistic patterns associated with empathetic expression. For example, rule-based approaches have been used to match such patterns in online health support communities [42], while other studies incorporate features such as metaphors or idioms into their classifiers (e.g., RoBERTa-twitter-sentiment and T5) to better capture empathetic tone in forum posts [54]. Another line of work draws on the close relationship between empathy and emotion. For example, multi-task learning has been used to transfer knowledge from emotion classification to empathy prediction in news and social media comments [47]. These methods differ in technique but share an underlying assumption that empathy can be treated as a singular, detectable signal.

More recent work challenges this assumption by emphasising empathy's multidimensional nature. Sharma et al. [83] introduced a framework derived from posts on two mental health platforms, which distinguishes emotional reactions, interpretations, and explorations, and trained a multi-task RoBERTa-based bi-encoder model to identify these components. Their framework has since shaped a considerable body of subsequent research [e.g., 27, 52]. Following this direction, Ziems et al. [109] evaluated multiple language models under the same scheme and showed that recognising these dimensions is particularly difficult in zero-shot or few-shot settings. Other studies extend the multidimensional view in different ways: appraisal-based scales (e.g., pleasantness, anticipated effort) have been used to characterise empathy in condolence messages with random forest and RoBERTa models [107], and other work examines empathetic alignment between an observer and a target using pre-trained language models guided only by prompts [98]. Together, this emerging body of research highlights a shift from treating empathy as a monolithic signal to understanding it as a complex, context-sensitive phenomenon.

Despite these developments, identifying empathy in online settings remains challenging due to several limitations:

Lack of specificity and granularity: Although empathy is widely recognised as a multidimensional construct, most empirical studies on identifying online empathy rely on abstract theoretical frameworks without considering its nuanced dimensions [e.g., 51]. Additionally, some studies lack a clear definition or theoretical framework of empathy before training models to identify it. This can hinder further analysis of empathy practices in subsequent research.

Overlooking the relational perspective: Empathy is inherently relational and exists within interactions. However, previous work primarily focuses on responses/replies, while explicit or implicit solicitations for empathy or their connections with empathetic responses remain underexplored. To fully capture the dynamics of empathy in online interactions, it is crucial to consider both perspectives.

Lack of contextual awareness: Even when measuring the expressions of empathy in responses, contextual information in the original posts can provide valuable background that help identify appropriate empathetic reactions. However, much of the existing work relies solely on signals within the responses themselves and lacks awareness of the broader context.

Limited scope of analysis: While empathy is common in online communication, its identification and analysis have primarily been limited to specific areas, such as mental health or online therapy. Additionally, studies using previous classifiers may also overlook variations in empathy practices across different community norms and expectations.

2.3 Language Models for Social Tasks

Language models (LMs) exhibit in-context learning abilities and can generalise well across a wide range of downstream language tasks, such as classifying and explaining various social and psychological constructs. LMs have demonstrated promising performance in many of these tasks using prompting alone or in combination with few-shot examples. For instance, Zhang et al. [104] used prompting strategies, such as Decision-Tree-of-Thought, which improved accuracy and rationale quality by recursively re-prompting the model with more fine-grained context when uncertainty was detected. Their results show that LMs such as GPT-3.5-turbo and T5 can outperform baselines in toxicity detection. Cruickshank and Ng [26] evaluated 10 open-source LLMs, such as LLaMA-2 and Phi-2, and found that zero-shot models with chain-of-thought prompting, as well as few-shot models, achieved performance comparable to in-domain supervised models. Maceda et al. [59] applied one-shot learning using the GPT-4 model for sentiment classification on social media texts and found that it achieved substantial agreement with human annotations. However, while zero-shot and few-shot models with prompt engineering can perform well on many tasks,

Paper	Platform	Interactions	Empathy Construct	Approach	Performance
Han et al. [42]	Cancer support groups	Message posts from 236 participants	Binary empathy: expressions like "I'm so sorry for you" or "My heart goes out to you."	Rule-based matching	Agreement = 90.9% with human coding
Lee et al. [54]	Acne support forum	Posts and replies in the forum	Binary empathy: defined as per Sharma et al. [83]	Feature-based models (e.g., SVM, Naive Bayes) and pre-trained language models (e.g., RoBERTa-twitter-sentiment, T5)	F1=87.5
Zhou and Jurgens [107]	Reddit	Condolence-giving comments	Appraisal-based empathy (five-point Likert scale)	RoBERTa and random forest models	MSE=0.492
Yang and Jurgens [98]	Reddit (mental health-related subreddits)	Post–comment and comment–reply pairs	Alignment-based empathy (binary)	Pre-trained language models (e.g., RoBERTa, SpanBERT) and prompt-based models (e.g., OpenPrompt+RoBERTa, OpenPrompt+T5)	F1=46
Hosseini and Caragea [47]	News platform, Twitter	Messages in reaction to news articles and cancer-related posts	Seeking & providing empathy (binary)	Multi-task learning with knowledge distillation and teacher annealing	F1=68.41 (NewsEmp); 85.71 (TwittEmp)
Sharma et al. [83]	TalkLife, Reddit (mental health-related subreddits)	Seeker post – response post pairs	Emotional reactions, interpretations, & explorations (none, weak, or strong)	RoBERTa-based bi-encoder model	F1=74.29, 67.46, 73.47 (TalkLife); 74.46, 62.6, 72.58 (Reddit)
Ziems et al. [109]	Same as Sharma et al. [83]	Same as Sharma et al. [83]	Same as Sharma et al. [83]	Zero-shot and few-shot language models (Flan-T5, GPT-3, GPT-4)	F1=41.5

Table 1: Precursor studies on identifying online text-based empathy.

they may fall short on more nuanced tasks, particularly those involving psychological constructs. For example, they perform poorly in zero- or few-shot settings when identifying figurative language [57], empathy [109], and implicit hate [109].

To address the limitations of prompt-based approaches in nuanced tasks, recent studies have increasingly focused on fine-tuning language models to improve task-specific performance. Liu et al. [58] proposed a series of EmoLLM by fine-tuning LLMs based on comprehensive affective instruction datasets, and their models outperformed existing sentiment analysis tools, open-source LLMs, and surpassed GPT-4 on most tasks. Shah et al. [80] fine-tuned GPT-3.5-turbo and LLaMA2-7B for depression detection on social media, and these models achieved high accuracy and outperformed state-of-the-art zero-/few-shot models such as GPT-4, GPT-4o, and Gemini, which supports the effectiveness of task-specific fine-tuning. Similarly, Gül et al. [41] showed that fine-tuned GPT-3.5-turbo, LLaMA-2-13B, and Mistral-7B achieved state-of-the-art performance on stance detection across multiple datasets and consistently outperformed traditional baselines and zero-/few-shot models. However, the potential of fine-tuning is still underexplored in many other tasks where zero-/few-shot models struggle, such as empathy detection.

In addition to classification tasks, LMs also provide flexibility in input and output formats, which makes it possible to extract rationales from the text along with predictions. For example, as LMs struggle with hate speech detection, Nirmal et al. [65] used GPT-3.5-turbo to extract features from texts that promote hateful sentiment, and then trained a base hate speech classifier. Huang et al. [48] suggested that LMs provide advantages in interpretability by generating explanations comparable to traditional methods for sentiment analysis and feature attribution. However, they also highlighted the risks of relying solely on these self-generated explanations without further validation.

Given the nuanced and complex nature of empathy practices, and the current limitations of LMs in accurately identifying it [109], this study aims to fine-tune LMs to improve performance on empathy classification tasks. We also take advantage of the flexible input and output formats of LMs to experiment with different settings, such as generating either classification outputs alone or outputs accompanied by rationales. Additionally, we evaluate the performance of few-shot models for comparison.

2.4 Empathy as a Critical Aspect in Online Communities

Online communities provide a valuable source of information and support, enabling people to interact with others who share similar experiences or interests [49, 50]. This shared connection can further foster empathy practices [70, 74]. In online communities, empathy plays a crucial role in establishing trust [38], mitigating negative bystander behaviours [10, 86], and motivating members to provide knowledge and support [106]. Empathy can be triggered by witnessing others' suffering or distress and feeling a desire to ease their pain [11, 45]. It can also arise in response to positive emotions, where people express empathy to celebrate or share another's happiness [45]. In online spaces, empathic expression can be seen as a form of social support or prosocial behaviour [4, 43, 84, 90, 97, 101].

The need for empathy can be a significant motivator that drives individuals to participate in online communities [50, 77, 105]. In online communities such as those focused on health [108] or support for communities for violence victims [22], individuals often expect to receive not only knowledge but also empathy and emotional support. Research shows that people in such communities may prefer support from others who can empathise with their experience [14] and feel discomfort when responses lack empathy [96]. Engaging in empathy practices can also influence individuals' perceptions of online communities, fostering more positive attitudes such as appreciation for the community [64, 69, 78], which can, in turn, enhance user engagement within the community [13]. Considering the positive effects of empathy, previous studies have also explored ways to increase the expressed level of empathy through sentence-level editing [81], collaboration with AI-in-the-loop agents [82], or by automatically generating empathetic responses [56].

However, online communication is not always empathetic, and it varies depending on the focus of communities. For example, in communities focusing on emotional support or patient support, empathy tends to be relatively prevalent [72]. Conversely, in communities where the primary focus is on sharing information and achieving efficiency (e.g., gig worker communities), the competitive nature of these spaces may discourage the display of empathy practices, as it can be seen as detracting from the primary informational function [99]. Moreover, many online communities allow some degree of anonymity (e.g., the use of pseudonyms) [3, 5], which may encourage greater empathy in supportive online communities. However, this is not always the case, especially in communities where norms are centred around harassment or disruption [79].

Given the complexity of empathy across different contexts, the goal of an online community should not simply be to maximise empathetic interactions but to ensure that empathetic responses are provided appropriately to those who genuinely need them. This highlights the need for a comprehensive understanding of empathy practices in online communities.

2.5 Outcome Measures for Empathy Seeking and Provision

Prior work on online interactions has shown that engagement-based behavioural signals, including how users receive responses and how they subsequently respond to others, serve as effective outcome measures for characterising support-seeking and support-providing dynamics.

One common indicator of support-seeking outcomes is whether a post receives replies and how many it receives. Both the presence and volume of replies have been shown to reflect levels of community engagement, responsiveness, and overall supportiveness that a post receives [28, 55] and are essential for OP's future re-engagement [20, 28] and community development [21, 93]. Posts that go unanswered are linked to negative user experiences and unmet emotional or informational needs [62, 76]. Thus, reply activity functions as a direct proxy for how effectively a community responds to a seeker's needs.

On the support-provider side, the perceived quality or appropriateness of a reply can be inferred from the OP's subsequent behaviour. When OPs accept a reply [7, 92] or express acknowledgement [61], it signals that the response was noticed and appreciated. These behaviours represent forms of seeker-side validation, suggesting that the support provided was meaningful or satisfactory. Such reciprocation of knowledge and social interaction are positively associated with sustained contribution [63, 100], indicating that empathetic or helpful exchanges can reinforce ongoing participation from both parties.

In summary, these engagement-based signals, such as reply count and OP re-engagement, serve as meaningful and empirically supported proxies for assessing the effectiveness of empathetic communication. Building on these insights, this study examines how empathy shapes interaction patterns by using these behavioural indicators as outcome measures.

2.6 Research Objectives and Hypotheses

Grounded in the overall theoretical framework, this study deconstructs the granular aspects of empathy practices in online communities and examines how they can be operationalised, identified, and compared across different platforms. Prior work has not fully addressed the multidimensional and relational nature of empathy in online interactions, nor how these practices vary across community contexts. To address these gaps, we structure our investigation into three research objectives (ROs), with hypotheses specified for the third objective.

- ***Research objective 1***: Our conceptual framework captures distinct dimensions of empathy practices in online communities.
- ***Research objective 2***: Our model identifies multidimensional and relational empathy practices in online posts and replies, consistent with human-annotated evidence.
- ***Research objective 3***: We are able to measure how empathy practice patterns vary across different online platforms and communities, such as Reddit and Stack Exchange.
  - ***Hypothesis 3.1***: Empathy practices are distributed differently across platforms and community topics.
  - ***Hypothesis 3.2***: Posts that include empathy-seeking features will receive more replies, and this relationship is platform or topic specific.
  - ***Hypothesis 3.3***: Replies that include empathy-providing features will have a higher probability of receiving a response or being accepted from the original poster, and this relationship is platform or topic specific.

3 Data

In this study, we collected data from two widely used public platforms for peer interactions: Reddit and Stack Exchange. These platforms were selected because they offer similar but complementary environments for examining empathy practices online. Both platforms host diverse communities, where posts and replies are organised according to topics. Their clear interaction structures and post–reply formats provide a shared framework in which empathy can be communicated and interpreted.

These two platforms also differ in their interaction environments. Stack Exchange is primarily designed for efficiently obtaining answers, while Reddit also encourages open-ended discussion in addition to that [36, 88]. This difference allows us to examine the role of empathy in distinct communicative contexts. In each platform, we focus on three communities that share similar structural formats but differ in thematic focus. We first select specific communities on Stack Exchange based on the criteria described below and then identify corresponding communities on Reddit.

Stack Exchange. Stack Exchange is a popular forum which incorporates 182 question-answering communities.[1] In these communities, posts begin with a question, and subsequent replies aim to provide answers. For the purpose of clarification, we consistently refer to the initial question as a "*post*", and the subsequent answers as "*replies*". On Stack Exchange, original posters can accept a certain reply as the best answer, and other users can give their feedback by voting up or down on replies.

For our analysis, we selected communities based on the following criteria:

(a) There was a reasonable analogue between Reddit and Stack Exchange.
(b) The conversations tended to focus on personal challenges for which empathy would figure commonly in the discussion. This contrasts with communities related to technical or professional information, which conceivably would be lower in empathy, or politics, where empathy is likely to be group-specific.

We reviewed the community profiles and conversations of all communities and chose three specific communities: "Parenting Stack Exchange",[2] "Interpersonal Skills Stack Exchange",[3] and "The Workplace Stack Exchange".[4] These communities focus on different aspects of relationships and provide a valuable sample for this study, as they are highly relevant to empathy practices. The Stack Exchange data was obtained through the publicly available data dump [85].

Reddit. Reddit is a popular online platform that consists of numerous communities known as subreddits, each centred around a particular theme. As of 2022, it has over 100k active subreddits.[5] Within these subreddits, a post initiates a conversation, and subsequent replies contribute to the ongoing thread. The structure of these threads follows a nested tree pattern, with each reply directly linked to its parent reply or the original post. For our data collection, we selected the most relevant and popular subreddits that correspond to each of the Stack Exchange communities mentioned earlier, specifically "r/Parenting",[6] "r/socialskills",[7] and "r/work".[8] The Reddit data was collected from the Pushshift dataset [94].

As previously discussed, there are structural differences between the threads in Reddit and Stack Exchange. In this study, we consider top-level (i.e., direct) replies in Reddit and answers in Stack Exchange equally as "replies", as they both represent others' direct responses to the original posts. To ensure consistency across all communities on both platforms, we used data up to the end of 2022. We removed duplicate posts, along with their associated replies from both the Reddit and Stack Exchange datasets. We also filtered out replies made by the original posters themselves. In the Reddit dataset, posts or replies marked as "removed" or "deleted" were excluded. For Stack Exchange data, we extracted the content from its HTML format. Table 2 shows the summary of our datasets.

[1]https://stackexchange.com/, accessed in August 2023
[2]https://parenting.stackexchange.com/
[3]https://interpersonal.stackexchange.com/
[4]https://workplace.stackexchange.com/
[5]https://www.redditinc.com/
[6]https://www.reddit.com/r/Parenting/
[7]https://www.reddit.com/r/socialskills/
[8]https://www.reddit.com/r/work/

		Reddit	Stack Exchange
Parenting		r/Parenting	Parenting Stack Exchange
	Post	218,829	6,382
	Reply	2,272,348	19,758
Social Skills		r/socialskills	Interpersonal Skills Stack Exchange
	Post	153,584	3,572
	Reply	678,978	12,079
Working		r/work	The Workplace Stack Exchange
	Post	34,141	29,244
	Reply	73,868	96,321

Table 2: Number of posts and replies by platform and topic.

4 Methods

To examine empathy practices empirically, we proceeded in four stages building on selected communities from Reddit and Stack Exchange:

(1) Codebook development: Authors derived a codebook from a sample of 600 post–reply pairs.
(2) Annotation: Trained annotators applied the codebook to 1,302 post–reply pairs, annotated both categories and corresponding snippets.
(3) LLM fine-tuning: Trained an ensemble of models and selected the best-performing one based on accuracy scores.
(4) Scaled LLM inference: We applied the selected model to 3,572 posts and all associated replies in each community, yielding a total of 21,432 posts and 97,071 replies for statistical analysis of empathy practices.

4.1 Unpacking the Thematic Dimensions of Online Empathy

4.1.1 Overall Framework. As discussed in the related work (Section 2.1), this study considers the relational nature of empathy by examining the perspectives of both original posters and repliers. Specifically, for original posters, we analyse the types of requests made in the original post. To capture potential implicit empathy-seeking practices, we also incorporate self-disclosure, which includes informational disclosure and emotional disclosure. In summary, we consider three empathy practices from the original posters' perspective:

- Request: Expressing a need for guidance, understanding, or support from others.
- Informational disclosure: Sharing personal information [2].
- Emotional disclosure: Sharing emotions towards someone or something [2].

From the perspective of repliers, this study focuses on three key components of empathy-giving practices: emotional empathy, cognitive empathy, and empathic concern.

- Emotional empathy: The emotional reaction of sharing or experiencing the stimulus person's emotional state [40, 44].
- Cognitive empathy: The intellectual process of understanding the stimulus person's perspective, involving perspective taking, critical thinking, or inference [39, 46, 95].
- Empathic concerns: The desire to care for or improve the stimulus person's welfare [31, 33].

4.1.2 *Themes.*

Following the overall framework of empathy practices in requests and responses, we conducted thematic analysis to examine their specific patterns in online community conversations. We randomly selected 100 post–reply pairs from each community, resulting in a total of 600 pairs. Our sampling is based on the post—reply pair as the unit of analysis, since our focus is on empathy practices between the original poster and the replier. This approach implicitly accounts for post popularity, as more popular posts are associated with more post—reply pairs.

We conducted a hybrid thematic analysis, in which we used the above theoretical framework to guide our initial coding scheme and added codes to capture salient ideas that emerged during coding [34]. The coding process followed the thematic analysis protocol outlined by Braun and Clarke [15]. For posts, we focused on themes relating to requests or personal disclosures; for replies, we identified patterns in empathetic expression.

The coding of the entire sample was performed by the authors. Coders first familiarised themselves with the data and generated initial coding based on the theoretical framework. Working in batches of 100 post–reply pairs, they synthesised these codes into candidate themes and refined them through repeated discussions. Through this iterative process, they jointly reviewed and refined the developing themes to ensure consistency between the coded segments and the thematic structure across the dataset. By the last batch, no new themes emerged and no further merges were needed, indicating that the structure had stabilised. They then finalised the themes and agreed on their names and definitions. This process produced a set of analytically grounded themes that both reflected our theoretical framework and incorporated new, data-driven insights.

To ensure clarity and consistency in our scheme, coders collaboratively developed detailed descriptions and examples for each theme. We consulted five additional researchers to confirm that the descriptions were clear and understandable. This process ensured that each theme was well-defined and recognisable, and helped maintain consistent interpretation across annotators in subsequent steps. The resulting thematic structure then guided the design of our final annotation questions for collecting human annotations and formed the basis for model development in later stages.

4.2 Empathy Annotation Scheme

4.2.1 *Sampling.*

To train models to automatically identify themes in the theoretical framework, we first created a dataset for annotation. In each of the six selected communities on Reddit or Stack Exchange, we sampled 217 posts using a probability proportional to (number of received replies + 1). This sampling method considers the distribution of replies for further sampling and ensures the inclusion of posts without replies. For each selected post, we randomly chose one reply, if available. This process provided us with a subset of 1,302 post–reply pairs for annotation.

4.2.2 *Annotators Recruiting and Training.*

We recruited three graduate students (ages 20–30; two female, one male) to annotate the sampled dataset. After finishing several training sessions and becoming familiar with the codebook, the annotators practised coding 10 typical post–reply pairs to understand the coding process and the interface. For each post–reply pair, annotators were first presented with the post and asked to select the applicable empathy-seeking themes. If a theme applied, they were asked to highlight the relevant snippets in the original post. Next, the reply was displayed below the post, and annotators were asked to review all empathy-giving themes related to the reply, select the applicable themes, and highlight the corresponding snippets in the reply. The annotation interface was set up using Label Studio [87]. The annotation interface is shown in Appendix A.

After discussing the annotations from the training session and reaching an agreement, the annotators were asked to annotate a set of 100 post–reply pairs. The average Fleiss' Kappa for post-related themes and for reply-related themes were both 0.76, with individual themes scoring 0.67–0.96. This indicates substantial agreement among the annotators [53]. The authors then examined all examples of disagreement, discussed them with the annotators, and reached a consensus. The final agreed-upon annotations for these 100 post–reply pairs were included in the final annotated dataset.

4.2.3 *Main Coding Procedure.*

The remaining 1,202 post–reply pairs in the main coding procedure were divided among the three annotators and split into four batches, with each annotator receiving about 100 pairs per batch. The annotators were asked to flag any uncertain examples in the annotation interface, which the authors reviewed regularly. The authors then discussed these flagged examples with the annotators to finalise the annotations. After each batch was completed, the authors summarised the uncertain examples and the explanations for the annotations, held team meetings to review them collectively, and reached an agreement on all annotations.

4.3 Fine-Tuning Language Models for Empathy Detection

We selected language models (LMs) based on their popularity, accessibility, and performance on relevant tasks. We chose open-weights models due to their customisability and transparency, and used instruct-tuned versions as they are more capable of following task-specific instructions. Specifically, we fine-tuned Llama-3.1-8B-Instruct, Mistral-Nemo-Instruct, and Phi-3-Medium-Instruct models on our human-labelled data. We fine-tuned these models using 4-bit quantization with QLoRA [35] via the Unsloth framework [9] for 3 epochs. Further details of the experimental setup are provided in Appendix B. For comparison, we also included few-shot models in our experiments.

We conducted experiments under two conditions regarding the detection of empathy-giving practices in replies: one using only the reply, and another incorporating the context (i.e., the original post plus the reply). For all empathy practice themes, we tested two approaches: one using only the classification results, and another using both the theme classification and the user's highlights of the segments indicating that theme. When the training dataset includes highlights, the models are also required to provide highlights in their results.

Following this, we report on the accuracy, prevalence, and combinations of the empathy practices using results from the top-performing model as applied to a larger corpus. We analyse how

[9]https://github.com/unslothai/unsloth

frequently these empathy practices occur across the six communities, how different kinds of empathy-eliciting requests shape the nature of empathetic responses, and which practices tend to encourage continued engagement from the original poster.

5 Results

5.1 Empirical Themes of Online Empathy Practices (*RO 1*)

Based on the high-level theoretical framework, our thematic analysis reveals granular empathy practices from the perspectives of both original posters and repliers. Tables 3 and 4 present the identified themes along with corresponding examples for posts and replies, respectively. In summary, we identified six empathy-warranting themes related to requests or self-disclosure in original posts, and seven empathy-giving themes that reflect emotional empathy, cognitive empathy, or empathic concern in responses.

In the next section, we fine-tune language models to detect these distinct dimensions of empathy expression and reception. We then return to our *RO 1* and examine the correlations among these dimensions to evaluate whether our theoretical framework capture distinct aspects of empathy (Section 5.2.4).

5.2 Empathy Identification from a Multidimensional and Relational Perspective (*RO 2*)

The performance of the fine-tuned models across all empathy practice themes is shown in Table 5. To evaluate model performance, we compare our models against two sets of baselines: (1) basic baselines: a distribution-based random predictor (rand.), which samples labels proportionally to the training distribution, and a majority-class predictor (maj.), which assigns the most frequent label in the training set; and (2) few-shot prompted language model baselines, which use pre-trained models without fine-tuning by providing 4–6 demonstration examples per empathy dimension in the prompt (see Appendix C). Due to an insufficient number of data points in the "seeking emotional support" category (see Appendix B.1 for the distribution of labels across posts and replies), we did not test the models on that theme or include it in the subsequent analysis of this study. We maintain that this is a key part of our theoretical focus on empathy as a multidimensional construct. The communities chosen in this study were not oriented toward this type of interaction, while many other communities can be specifically designed for seeking emotional support, such as mental health forums. Accordingly, the empirical analyses in this study focus on empathy practices in comparable, everyday contexts where empathy is plausible but not structurally mandated, rather than on providing a definitive account of all settings where empathy could be applied.

5.2.1 Overall Model Performance across Themes. The results show that LMs can effectively identify empathy practices, both in terms of how they are solicited and how they are expressed in online communication. Compared with few-shot LM baselines, fine-tuning consistently improves performance (see Table 12 in Appendix C). In our settings, Phi-3-Medium-Instruct models outperform other models, achieving macro-F1 scores ranging from 0.72 to 0.90 across all themes. These findings suggest that the best-performing model consistently identifies signals associated with different types of empathy practices across themes. In contrast, in our experiments, models like Llama-3.1-8B-Instruct and Mistral-Nemo-Instruct can identify signals for some themes, but they struggle with those that require more nuanced interpretation, performing similarly to or even worse than basic baseline models.

Regarding the variance in model performance across different themes, some themes are easier to detect, such as "asking for advice or opinions" in posts and "sharing similar experiences" in replies. Different models, with their various training data, processes, and architectures, perform relatively well on these themes. This may be because such practices exhibit clearer linguistic cues and can be inferred from general knowledge without needing a deep understanding of relational dynamics or contextual nuances. However, other practices like "expressing validation," are relatively challenging for all models. In these cases, model performance is noticeably lower compared to other themes. These practices may demand a more context-sensitive understanding of online interactions, and are therefore difficult for language models to detect.

In summary, our theoretical framework breaks down the abstract concept of empathy into distinct, interpretable themes of empathy practices. This enables LMs to effectively recognise and differentiate these themes, though some remain challenging depending on model capability and theme complexity.

5.2.2 Use of Contexts and Highlights. For all models and themes, we experimented with variations of input in two dimensions: 1) using only classification (referred to as "class" in Table 5), or using classification along with quotes from the original text as rationale (referred to as "+quotes") in fine-tuning; 2) for themes related to replies, using only the reply as input, or using the reply plus the original post as context (referred to as "post+reply"). Overall, the best-performing model (Ph-3-Medium) in this study demonstrates relatively stable performance across various input settings. However, the performance of Llama-3.1-8B-Instruct and Mistral-Nemo-Instruct depends a lot on the input settings. Under some conditions, they fail and perform similarly to majority or random guessing baselines.

Although the original post offers context for interpreting empathy practices in replies, and the corresponding post is presented when collecting ground-truth labels, our results show that including the post as input only improves model performance in some cases. In several cases, models with less input perform better. For example, Llama-3.1-8B-Instruct is not good at handling inputs that include both the post and the target reply, while excluding the post often yields better results. This suggests that while adding the post can provide the model with additional background information, it can also introduce noise or distraction compared to models that rely on more focused input. Most of the information to classify the empathy practice in a reply may already be contained within the reply itself.

In addition to binary classifications for each theme, we collected highlighted quotes from posts and replies that indicate the presence of each theme. We experimented with incorporating this information during model fine-tuning and observed performance gains in some cases. This suggests that using highlights can help the model more accurately capture how empathy practices are expressed,

Category	Theme	Example
Request	Asking for advice or opinions	*"How do I deal with the problem of . . . "; "How to support a friend with. . . "; "Is there any way I can find someone that'll..."; "Why is my toddler. . . "; "What do you think of..."*
	Seeking similar experiences or feelings	*"Has anyone else here gone through such a phase?"; "Does someone feel the same?"; "Any other parents dealing with something similar?"; "Has anyone encountered a similar situation and handled it..."*
	Seeking circumstance calibration	*"Is that how it should be? Or is it just my problem?"; "Should I have handled differently?"; "Or if I have to mend the way I feel about my friends interacting?"; "Am I overthinking this?"*
	Seeking emotional support	*"Give me your strength, parents around the world"; "Needing all the positive vibes"; "I Need Some Support/Hugs"; "Would like some kind words of reassurance"; "New mommy-to-be looking for some encouragement, empathy or whatever"*
Self-disclosure	Sharing personal information	*"I (29f) never..."; "I work as a web developer"; "I'm a trans man"; "I live about 10 miles outside Cambridge"; "I suffer from Cerebral Palsy"; "I am a 15-year-old Egyptian Muslim"*
	Sharing emotions	*"So annoyed right now"; "I have an overwhelming sense of dread"; "Been really depressed for the past 1 month"; "I instantly felt disliked and hurt because of it"; "I'm glad that it's done but I'm sad, too"; "I am extremely positive/happy"*

Table 3: Themes identified in *posts*, and examples for each theme.

thereby improving its performance. However, these improvements are not consistent, and in some situations, models trained with classification alone outperform those given additional highlight inputs. Despite this, adding highlights in fine-tuning always offers an advantage by guiding the model to identify specific expressions of empathy practices rather than just generating a binary classification.

5.2.3 Error Analysis. We qualitatively examine the discrepancies between the model output and the annotated labels. This analysis focuses on our best-performing model structure, Phi-3-Medium-Instruct. For false negative examples, we reviewed the annotations of highlighted texts that indicate specific themes. For false positives, we analysed the model-generated highlights to understand the rationale behind the model's output.

Over-Inference. The models sometimes over-interpreted and inferred implicit expressions that did not necessarily align with the intended themes. For example, a post saying, *"My son was unjustly cut from the basketball team"* was identified as "sharing emotions," even though it did not explicitly mention any emotions. Similarly, a post including *"Hey all, I feel discomfort. This thing kills me."* was detected as "asking for advice or opinions," despite the lack of a direct request. The models may sometimes struggle to differentiate between nuanced themes, leading to the misclassification of one empathy practice as another. For instance, a replier shared similar emotions to the original poster with the words, *"It has been incredibly difficult, frustrating, and disheartening..."*, but this was inappropriately identified as an expression of validation.

Misinterpretation of the Subject or Target. Another error is the mismatch of the subject or target of the emotions or experiences. For example, the model incorrectly interpreted a reference to someone else's information as the poster sharing their own personal details, such as in a post, *"my husband is 40 years old"*. It might also fail to accurately identify whose emotions are being discussed, e.g., a reply that shares experiences similar to those of the poster's friend, rather than the poster themselves, was inaccurately classified as "sharing similar experiences". Moreover, the model could incorrectly view the quoted sentences in the reply as direct communication with the original poster. For instance, the snippet *"If that's true, say something like 'What are your goals here?',"* which was part of a piece of advice, was misidentified as showing interest in the poster's further elaboration.

When given both the post and its reply, the models can sometimes, though rarely, attribute themes in the post to the reply. For example, the model incorrectly identified the sentence in a post saying, *"I love being a parent because of funny situations like this,"* as the replier recalling similar feelings or emotions in their response. Such confusion between source and target text may help explain why, in some cases, including the post even reduced the model's performance in identifying empathy practices in replies. It required the model to accurately distinguish between empathetic expressions in the reply and emotional or contextual cues in the original post.

Lack of Generalisability. The models' performance may not extend well beyond the specific examples present in the prompts or training set. For instance, the model failed to identify the disclosure in a post stating, *"My real name is on my profile: xxx"*, as this direct

Category	Theme	Example
Emotional empathy	Expressing emotional resonance	*"Oof, sorry that is happening."; "That sounds incredibly hard and frustrating"; "Wow that sucks."; "Awesome!"; "Congrats!"; "Great job there"; "I love love love that you care and are so aware how important your relationship is!"*
	Recalling similar feelings or emotions	*"Just know that you are not alone. I felt the same that you felt"; "I also get very uncomfortable when she rubs my back"; "Yup, it's awful. You're not alone"; "Oh, man, this drives me completely crazy. My daughter has been doing it for 6 months"; "I think I can relate with your feeling of ... from my high school days"*
Cognitive empathy	Expressing validation	*"I understand wanting to get to the bottom of what happened"; "I know that confronting your brother is a scary thing"; "Yep. The owner is quite rich, and all the money is for..."; "It's totally reasonable to feel overwhelmed when trying to make new friends"; "I work in the same field so I think I have an idea of what you mean haha"; "It's normal to feel anxious in your situation."*
	Sharing similar experiences	*"I am in a similar situation right now too"; "I've been in this spot before"; "Wow I think I have the same problem"; "I had the same problem myself in high school too"; "I've always thought the same about FORD"; "I do this, too!"*
Empathic concern	Offering reassurance, encouragement, or good wishes	*"Take it with good humor and don't overthink it too much, it'll get better!"; "There is nothing to worry about. It is just a matter of time"; "hang in there"; "Good luck, and you are a good dad and she does love you."; "Keep it up. You can do it."*
	Showing interest in further elaboration	*"Is this your first job like this?"; "I could recommend tons of audiobooks if you tell me what you might like?"; "What are the rules of your job?"; "Please define what you mean by 'SPD'"; "I could go on for ages, but I'd need more information on you and what you are going for"; "Is there evidence that you requested time off?"*
	Offering personalised advice	*"You sound stressed. Take a complete break for 3-4 days or a week"; "The best you can do is ease out over the course of several months"; "If you can, I would be very direct in explaining to him that…"; "I recommend teaching her how to relax and meditate"; "You probably should only explain that to them"*

Table 4: Themes identified in *replies*, and examples for each theme.

form of disclosure was not provided in the instructions or the training set. Similarly, posts that described similar experiences with less obvious cues, such as *"I caused my company to lose 200k because of a bug in my code"* (where the original poster also shared a story of a job-related mistake), could be overlooked by the model, likely due to their relative infrequency in the training data. Additionally, some expressions not encountered during training or fine-tuning, such as the abbreviation used in self-disclosure: *"FTM over here"* (referring to female-to-male), could be challenging for the model to interpret accurately.

Annotation Error. A few differences between the model's predictions and the ground-truth labels can be attributed to the inaccuracies in the labels provided by the annotators. For instance, the annotators missed *"I'm a musician"* under the category of "sharing personal information" in a long post. It should be noted that since annotators were able to identify similar expressions in other examples, these errors are likely due to occasional lapses in attention rather than a misunderstanding of the codebook or systematic errors. This suggests an advantage of LMs in detecting patterns in online texts. Unlike humans, LMs can process long texts and identify empathy practices quickly and efficiently, without experiencing fatigue that can lead humans to overlook expressions of specific themes.

5.2.4 Co-occurrence Patterns Between Empathy Practices in Posts/Replies. The heatmaps below show the one-mode correlation between different empathy practices in posts (Figure 1) and replies (Figure 2) in each community. These correlations indicate how different dimensions of empathy practices co-occur. Overall, the empathy dimensions show mostly small correlations (<0.3) and sometimes moderate correlations (0.3–0.5), with only a few exceptions (e.g., *emotional resonance* with *recall feelings* (0.51) and with *validation* (0.59) in r/Parenting, and *recall feelings* with *shared experiences* (0.51) in r/socialskills) [24]. Additionally, these few higher-correlation cases may reflect frequent co-occurrence of certain empathy practices in specific samples rather than conceptual overlap between the dimensions themselves. Taken together, the correlation patterns provide additional empirical evidence that the theoretical

	Input / Category	Phi-3-Medium-Instruct reply/post class	Phi-3-Medium-Instruct reply/post +quotes	Phi-3-Medium-Instruct reply+post class	Phi-3-Medium-Instruct reply+post +quotes	Llama-3.1-8B-Instruct reply/post class	Llama-3.1-8B-Instruct reply/post +quotes	Llama-3.1-8B-Instruct reply+post class	Llama-3.1-8B-Instruct reply+post +quotes	Mistral-Nemo-Instruct reply/post class	Mistral-Nemo-Instruct reply/post +quotes	Mistral-Nemo-Instruct reply+post class	Mistral-Nemo-Instruct reply+post +quotes	Baseline rand.	Baseline maj.
	Request														
	Asking for advice or opinions	86.71	**87.32**			**80.09**	59.80	-	-	84.66	**85.24**	-	-	45.14	45.51
	Seeking similar experiences or feelings	88.68	**90.01**			**67.88**	48.52	-	-	72.71	**86.33**	-	-	49.86	48.52
Post	Seeking circumstance calibration	**80.00**	79.24			**66.18**	50.71	-	-	47.17	**73.75**	-	-	48.75	47.17
	Self-disclosure														
	Sharing personal information	79.26	**80.92**	-	-	**66.60**	64.02	-	-	**72.42**	68.72	-	-	48.04	42.64
	Sharing emotions	**83.15**	81.16	-	-	77.30	**79.44**	-	-	80.45	**82.29**	-	-	49.96	38.00
	Emotional empathy														
	Expressing emotional resonance	72.64	**73.75**	68.43	65.00	54.38	48.68	**55.80**	48.68	**69.38**	64.69	48.68	63.54	55.39	48.68
	Recalling similar feelings or emotions	**71.61**	64.45	70.78	70.65	**53.73**	48.47	48.47	48.47	58.98	**62.71**	48.47	48.47	47.06	48.47
	Cognitive empathy														
	Expressing validation	69.50	**72.44**	56.94	56.92	53.73	**56.25**	48.36	48.36	**58.12**	54.89	55.84	52.19	50.84	48.36
Reply	Sharing similar experiences	81.43	80.05	**83.18**	83.00	**69.31**	56.25	52.56	55.42	57.81	72.29	**74.94**	57.44	47.46	45.57
	Empathic concern														
	Offering reassurance, encouragement, or good wishes	74.96	76.74	78.05	**78.97**	71.97	**71.97**	58.88	69.01	69.34	67.32	68.25	**76.12**	53.12	45.57
	Showing interest in further elaboration	**73.31**	72.44	63.64	64.12	**54.10**	47.83	51.94	50.38	58.91	**62.71**	51.88	60.54	52.96	47.83
	Offering personalised advice	78.65	**79.78**	76.54	78.34	73.83	**77.87**	70.13	76.54	73.25	**79.69**	74.71	78.19	45.48	44.49

Table 5: Classification results (macro-F1) of fine-tuned models across model structures and input combinations. The best performance of input combination within each model structure is in bold, and the best model for each theme is highlighted in green .

framework developed earlier successfully disentangles the key dimensions of empathy practices.

5.3 Empathy Practices Across Online Communities and Platforms (*RO 3*)

In this section, we randomly selected 3,572 posts and all their associated replies from each community, based on the minimum number of posts in our target communities. The statistics of this subset are provided in Appendix D. We applied the best-performing model to this sampled dataset and examined the distribution of empathy practices in both posts and replies across the communities. We also compared the posts with and without empathy-warranting themes, as well as replies with and without empathy-giving themes. Additionally, we explored the relationship between empathy practices in posts and those in replies.

5.3.1 Community Difference in Empathy Practices. Our results show differences in the distribution of empathy practices across various dimensions, including platforms and community topics. As Figure 3 shows, Reddit and Stack Exchange display different patterns in both empathy-warranting requests and empathy-containing replies. Table 6 further shows the platform different in seeking and providing empathy. Compared to Stack Exchange communities, subreddits are less likely to include *"requests for advice or opinions"* (71.95% vs. 97.97%, $p < .001$), but are more likely to *"seek similar experiences or feelings"* (7.89% vs. 1.29%, $p < .001$) and *"share emotions"* (45.26% vs. 26.82%, $p < .001$) in posts. These subreddits also tend to provide more emotional empathy in their replies, including *"expressing emotional resonance"* (7.75% vs. 1.14%, $p < .001$) and *"recalling similar feelings or emotions"* (8.41% vs. 1.02%, $p < .001$), but offer less *"personalised advice"* (59.32% vs. 93.33%, $p < .001$) as concrete support.

The distribution of empathy practices is also influenced by community topics, regardless of the platform. Table 7 illustrates the distribution across three community topics (parenting, social skills, and work), each including both platforms. For example, on both Reddit and Stack Exchange, communities focused on social skills show more *"requests for advice or opinions"* (91.91% vs. 85.79% and 77.18%, $p < .001$), as well as more *"informational and emotional self-disclosure"* (informational: 22.37% vs. 9.06% and 20.58%, $p < .001$; emotional: (47.96% vs. 30.12% and 30.04%, $p < .001$) in original posts, compared to communities on other topics. In terms of empathy expressed in replies, parenting-focused communities show higher levels of *emotional empathy* (emotional resonance: 8.03% vs. 3.51% and 1.34%, $p < .001$; (7.46% vs. 5.32% and 2.06%, $p < .001$)), *"sharing similar experiences"* (27.12% vs. 20.58% and 11.69%, $p < .001$), and *"offering reassurance, encouragement, or good wishes"* (27.35% vs. 17.49% and 10.31%, $p < .001$) across both platforms. On the other hand, work-related communities demonstrate the lowest percentage of *emotional and cognitive empathy* in their replies.

5.3.2 Effects of Empathy Practices. We examine the effects of each empathy-seeking practice on the number of replies that posts receive. For each community, we compare posts with requests or self-disclosures to those without them. As shown in Table 8, including a request or self-disclosure in a post is associated with significantly more replies in several communities (e.g., r/work and r/parenting). In other communities, however, the significant effects are limited to specific dimensions. For example, in Parenting Stack Exchange, *"asking for advice or opinions"* yields more replies (diff. = +1.06, $p < .001$); in Interpersonal Skills Stack Exchange, only *"sharing personal information"* is associated with significantly more replies (diff. = +0.46, $p < .001$). Requests or self-disclosures therefore do not uniformly lead to greater engagement. In Interpersonal

(a) r/Parenting (b) r/socialskills (c) r/work

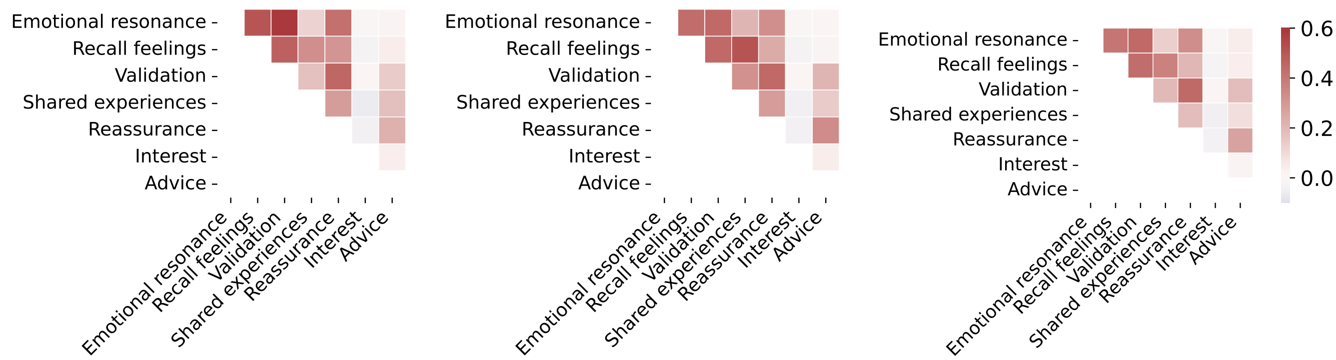
(d) Parenting Stack Exchange (e) Interpersonal Skills Stack Exchange (f) The Workplace Stack Exchange

Figure 1: One-mode correlation heatmaps showing co-occurrence patterns between request/self-disclosure dimensions in *posts*. Dimensions: asking advice (asking for advice or opinions), seeking similarity (seeking similar experiences or feelings), circumstance calibration (seeking circumstance calibration), personal info (sharing personal information), emotions (sharing emotions).

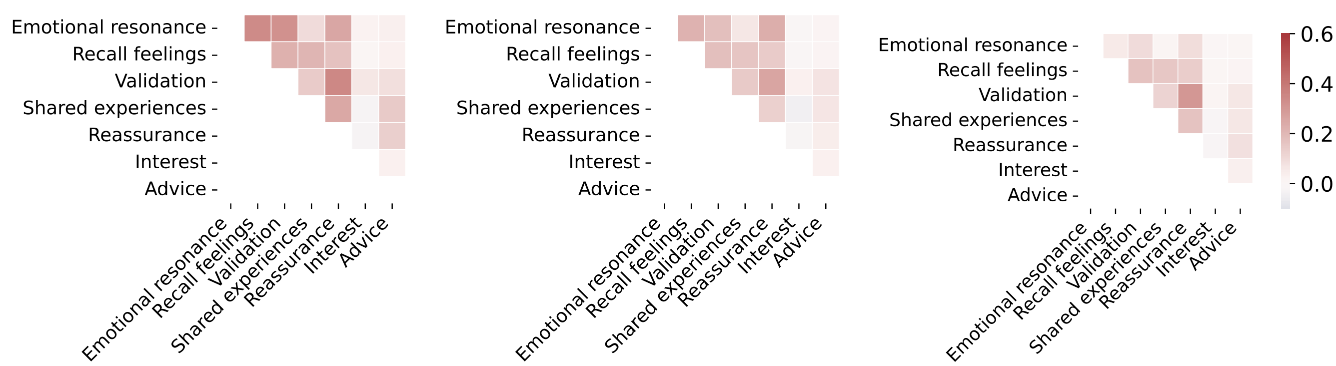
Skills Stack Exchange, for instance, posts that include *"sharing emotions"* receive significantly fewer replies than those that do not (diff. = −0.31, $p < .01$). Likewise, in communities such as r/socialskills, none of the examined empathy-seeking practices were found to be significantly associated with the number of replies.

For replies, we analyse the probability of a reply being accepted by the original poster (OP) in Stack Exchange communities. Since Reddit does not have this feature, we instead examine the probability of a reply being responded to by the OP there. As shown in Table 9, empathic concern practices such as *"showing interest in further elaboration"* (diff. = +25–27%, all $p < .001$) and *"offering personalised advice* are strongly associated with a higher OP response rate on Reddit (diff. = +4–7%, all $p < .001$). In contrast, Stack Exchange communities sometimes show significantly negative associations for these same practices (e.g., *"offering personalised advice"* in Parenting Stack Exchange (diff. = -3.68%, $p < .01$)). Stack Exchange communities also differ from Reddit in how they respond to emotional versus cognitive empathy. Cognitive empathy is sometimes linked to higher acceptance rates, most notably in Interpersonal Skills Stack Exchange, where *"expressing validation"* (diff. = +6.61%, $p < .001$) and *"sharing similar experiences"* show substantial positive effects (diff. = 4.35%, $p < .001$). Emotional empathy, however, is occasionally associated with lower acceptance rates, such as *"recalling similar feelings or emotions* in Parenting Stack Exchange (diff. = −6.45%, $p < .001$) and The Workplace Stack Exchange (diff. = −13.58%, $p < .001$), possibly reflecting the platform's norms favouring efficiency and task-focused responses. On Reddit, the patterns regarding cognitive empathy are more mixed, and we find no significant positive effects for emotional empathy.

5.3.3 Connection between Requests and Expression of Empathy. We consider all post–reply pairs and examine the relationship between requests/self-disclosures in the posts and empathetic responses in the replies. The correlations in all six communities are shown in Figure 4. The results reveal a positive correlation between the OP's *self-disclosure* and the replier's expression of *emotional and cognitive empathy*, as well as *personalised advice.* Additionally, certain types of requests or self-disclosures are more closely linked to specific dimensions of empathy. For instance, *"asking for advice or opinions"* often leads to *"offering personalised advice"*, *"sharing emotions"* typically evokes *"offering reassurance, encouragement, or good wishes"*,

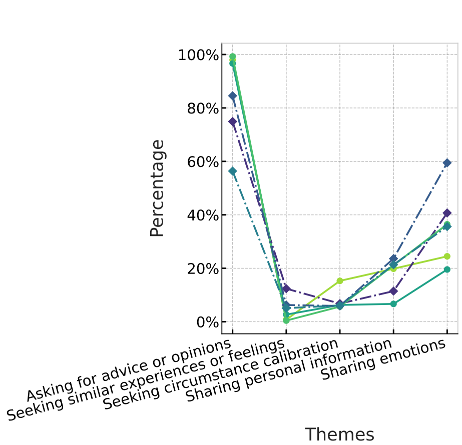


(a) r/Parenting (b) r/socialskills (c) r/work

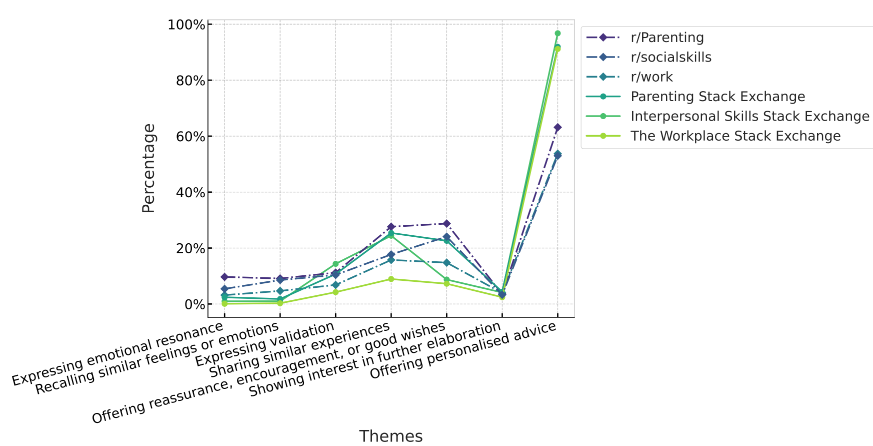


(d) Parenting Stack Exchange (e) Interpersonal Skills Stack Exchange (f) The Workplace Stack Exchange

Figure 2: One-mode correlation heatmaps showing co-occurrence patterns between empathetic expression dimensions in *replies*. Dimensions: emotional resonance (expressing emotional resonance), recall feelings (recalling similar feelings or emotions), validation (expressing validation), shared experiences (sharing similar experiences), reassurance (offering reassurance, encouragement, or good wishes), interest (showing interest in further elaboration), advice (offering personalised advice).



(a) Post



(b) Reply

Figure 3: Distribution of themes across communities.

and *"sharing personal information"* usually elicits *"expressing validation"*. This suggests that the effectiveness of strategies used in the original post to elicit empathy depends on the specific dimension of empathy the OP is seeking.

Type	Category	Mean ($\overline{\%}$)	Reddit − $\overline{\%}$	Stack Exchange − $\overline{\%}$	Cramér's V	χ^2
Post	*Request*					
	Asking for advice or opinions	84.96%	-13.01%	13.01%	0.364	2838.58***
	Seeking similar experiences or feelings	4.59%	3.30%	-3.30%	0.158	532.42***
	Seeking circumstance calibration	7.67%	-1.40%	1.40%	0.053	59.33***
	Self-disclosure					
	Sharing personal information	17.33%	1.46%	-1.46%	0.038	31.70***
	Sharing emotions	36.04%	9.22%	-9.22%	0.192	789.55***
Reply	*Emotional empathy*					
	Expressing emotional resonance	4.44%	3.31%	-3.31%	0.141	1923.28***
	Recalling similar feelings or emotions	4.71%	3.70%	-3.70%	0.152	2253.70***
	Cognitive empathy					
	Expressing validation	10.13%	0.32%	-0.32%	0.010	9.74**
	Sharing similar experiences	21.56%	1.99%	-1.99%	0.046	206.30***
	Empathic concern					
	Offering reassurance, encouragement, or good wishes	19.23%	6.54%	-6.54%	0.154	2299.58***
	Showing interest in further elaboration	3.56%	-0.08%	0.08%	0.004	1.46
	Offering personalised advice	76.33%	-17.00%	17.00%	0.362	12701.58***

Table 6: Proportions of empathy practices in posts and replies across *platforms* (Reddit and Stack Exchange). Mean ($\overline{\%}$) represents the average proportion across platforms; deviation columns show percentage differences from the mean. χ^2 tests assess platform differences, and Cramér's V indicates effect sizes. $N_{\text{post}} = 21{,}432$, $N_{\text{reply}} = 97{,}071$. $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Type	Category	Mean ($\overline{\%}$)	Parenting − $\overline{\%}$	Social Skills − $\overline{\%}$	Work − $\overline{\%}$	Cramér's V	χ^2
Post	*Request*						
	Asking for advice or opinions	84.96%	0.83%	6.95%	-7.78%	0.169	612.02***
	Seeking similar experiences or feelings	4.59%	2.93%	-1.26%	-1.67%	0.099	210.72***
	Seeking circumstance calibration	7.67%	-1.14%	-1.89%	3.03%	0.081	141.60***
	Self-disclosure						
	Sharing personal information	17.33%	-8.28%	5.03%	3.24%	0.156	520.38***
	Sharing emotions	36.04%	-5.92%	11.92%	-6.00%	0.176	660.18***
Reply	*Emotional empathy*						
	Expressing emotional resonance	4.29%	3.74%	-0.78%	-2.95%	0.124	1504.02***
	Recalling similar feelings or emotions	4.95%	2.51%	0.37%	-2.89%	0.089	767.41***
	Cognitive empathy						
	Expressing validation	9.49%	1.60%	2.60%	-4.21%	0.083	668.31***
	Sharing similar experiences	19.80%	7.32%	0.78%	-8.11%	0.143	1989.75***
	Empathic concern						
	Offering reassurance, encouragement, or good wishes	18.38%	8.97%	-0.89%	-8.08%	0.168	2752.57***
	Showing interest in further elaboration	3.46%	0.22%	0.20%	-0.42%	0.014	18.39***
	Offering personalised advice	72.48%	-2.81%	-0.71%	3.52%	0.053	276.92***

Table 7: Proportions of empathy practices in posts and replies across *community topics* (Parenting, Social Skills, and Work). Mean ($\overline{\%}$) represents the average proportion across community topics; deviation columns show percentage differences from the mean. χ^2 tests assess differences across topics, and Cramér's V indicates effect sizes. $N_{\text{post}} = 21{,}432$, $N_{\text{reply}} = 97{,}071$. $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

The impacts of the OP's requests and self-disclosure on triggering empathetic replies also vary by community. For example, in communities such as Parenting Stack Exchange and The Workplace Stack Exchange, most types of requests and self-disclosures are positively related to expressions of empathy in almost all dimensions. However, in other communities such as r/Parenting, r/socialskills, and r/work, some of the requests, like *"asking for advice or opinions"* and *"seeking circumstance calibration"*, are not always linked to increased empathy in the replies received.

6 Discussion

6.1 Deconstructing Online Empathy from A Multidimensional and Relational Perspective (*RO 1*)

Extensive research has examined the multidimensional aspects of empathy in in-person interactions, including emotional empathy [32, 40, 44], cognitive empathy [46, 95], and empathic concern [31, 33, 68, 86, 103]. Moving online, empathy practices manifest differently in text-based and asynchronous communication [67, 73]. However, previous studies on online empathy often reduce it to a single [42, 47, 54] or a few [83] general and abstract variables,

	Difference in Number of Replies											
	r/Parenting		r/socialskills		r/work		Parenting Stack Exchange		Interpersonal Skills Stack Exchange		The Workplace Stack Exchange	
	Diff.	(t-stat)	Diff.	(t-stat)	Diff.	(t-stat)	Diff.	(t-stat)	Diff.	(t-stat)	Diff.	(t-stat)
Request												
Asking for advice or opinions	2.42^{**}	(2.75)	−2.66	(−1.94)	2.17^{***}	(19.61)	1.06^{***}	(6.08)	−0.51	(−0.57)	0.86^{***}	(3.31)
Seeking similar experiences or feelings	1.77	(1.31)	0.75	(0.80)	2.52^{***}	(6.31)	0.49	(1.38)	2.43	(1.45)	−0.37	(−1.05)
Seeking circumstance calibration	8.46^{***}	(3.55)	0.91	(0.65)	2.05^{***}	(6.68)	−0.02	(−0.13)	0.12	(0.43)	0.51^{***}	(3.76)
Self-disclosure												
Sharing personal information	3.27^{**}	(2.39)	0.62	(0.96)	1.60^{***}	(9.86)	0.47	(1.89)	0.46^{***}	(3.38)	0.23	(1.96)
Sharing emotions	6.44^{***}	(7.08)	0.11	(0.18)	1.96^{***}	(15.34)	0.11	(0.82)	-0.31^{**}	(−2.94)	0.15	(1.39)

Table 8: Difference in the number of replies between posts containing each request or self-disclosure practice and posts not containing that practice. $N_{\textbf{post}} = 21{,}432$. * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

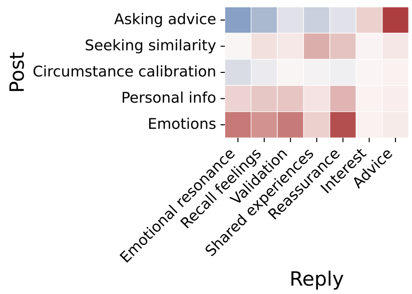
	Diff. in % Responded by OP						**Diff. in % Accepted by OP**					
	r/Parenting		r/socialskills		r/work		Parenting Stack Exchange		Interpersonal Skills Stack Exchange		The Workplace Stack Exchange	
	Diff.	(t-stat)	Diff.	(t-stat)	Diff.	(t-stat)	Diff.	(t-stat)	Diff.	(t-stat)	Diff.	(t-stat)
Emotional empathy												
Expressing emotional resonance	0.34%	(0.51)	0.74%	(0.52)	5.96%	(1.98)	−0.60%	(−0.30)	0.46%	(0.14)	−3.55%	(−0.35)
Recalling similar feelings or emotions	0.15%	(0.22)	−2.22%	(−2.04)	−2.33%	(−1.01)	$-6.45\%^{***}$	(−3.65)	0.65%	(0.20)	$-13.58\%^{***}$	(−42.72)
Cognitive empathy												
Expressing validation	$1.80\%^{**}$	(2.79)	$3.72\%^{***}$	(3.42)	4.41%	(2.13)	2.41%	(2.22)	$6.61\%^{***}$	(6.42)	4.25%	(2.43)
Sharing similar experiences	$-1.26\%^{**}$	(−2.87)	0.00%	(0.00)	$-3.49\%^{**}$	(−2.61)	0.70%	(0.96)	$4.35\%^{***}$	(5.45)	1.07%	(0.94)
Empathic concern												
Offering reassurance, encouragement, or good wishes	0.21%	(0.48)	1.62%	(2.16)	1.63%	(1.14)	1.07%	(1.39)	$4.10\%^{***}$	(3.29)	1.86%	(1.45)
Showing interest in further elaboration	$27.08\%^{***}$	(19.64)	$24.70\%^{***}$	(11.32)	$25.03\%^{***}$	(8.67)	−2.70%	(−1.91)	0.77%	(0.46)	−2.33%	(−1.22)
Offering personalised advice	$4.11\%^{***}$	(10.24)	$5.18\%^{***}$	(8.25)	$7.31\%^{***}$	(7.32)	$-3.68\%^{**}$	(−2.91)	$5.41\%^{***}$	(3.53)	2.39%	(2.28)

Table 9: Difference in the probability of being responded to (Reddit) or accepted (Stack Exchange) by the original poster between replies containing each empathetic expression and replies not containing that practice. $N_{\textbf{reply}} = 97{,}071$. $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

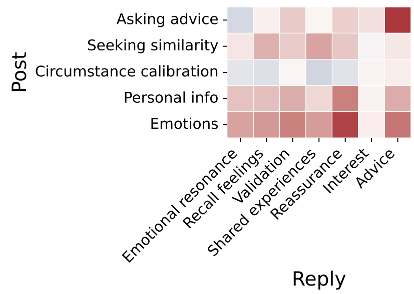
leaving theoretical frameworks for online empathy underexplored. Additionally, although empathy is a relational, interactional process [60, 91, 102], much existing research focuses on the expression of empathy while limited attention has been given to the original posts or solicitations that warrant empathetic responses. This study therefore bridges the gap between the nuanced insights from theoretical work and the often oversimplified operationalisations of online empathy.

To address *Research Objective 1*, this study aims to develop a framework that systematically captures the distinct dimensions of empathy practices in online communities. Inspired by existing theoretical research [2, 31, 33, 39, 40, 44, 46, 95] and grounded in a thematic analysis of post–reply interactions, the framework offers a comprehensive understanding of how empathy is sought and expressed online. It identifies six distinct empathy practices related to requests or self-disclosures in original posts, and seven distinct practices related to expressions of empathy in replies. This study achieves *Research Objective 1* by demonstrating a multidimensional and relational perspective that includes conceptually distinct themes that are often obscured in widely used binary or responder-oriented frameworks in previous research. Specifically, the framework enriches the conceptualisation of empathy by tracing the linguistic, functional, and relational signals through which different dimensions are expressed. Additional, in line with our relational perspective, the framework explicitly considers the interactive dynamics between the empathy-seeker and the responder, emphasising that empathy practices emerge through interaction and are shaped by contextual cues. Our empirical findings of co-occurrence patterns across these dimensions further support *Research Objective 1* by showing that the proposed framework can reasonably distinguish dimensions that are not interchangeable, allowing online empathy to be examined with greater conceptual precision.

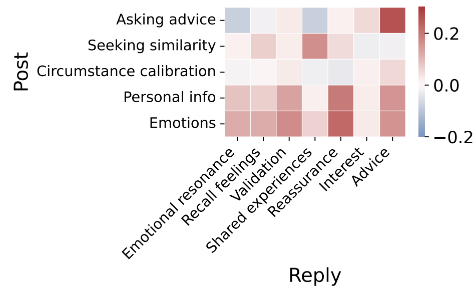
Based on our framework, this study also introduces a high-quality, human-annotated dataset of online empathy. The task of annotating empathy practices is usually challenging due to its subjective nature [47]. However, by breaking down online empathy practices into specific and well-defined themes, this study enables annotators to consistently identify distinct practices. Our comprehensive framework and corresponding dataset facilitate future research to conduct granular and context-sensitive analyses of empathy in downstream fields. This approach allows for further exploration of differences (e.g., effects or individual preferences)



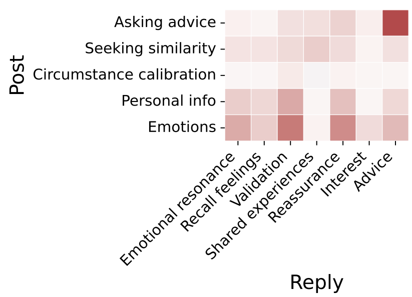
(a) r/Parenting



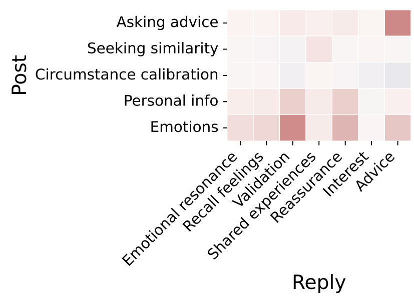
(b) r/socialskills



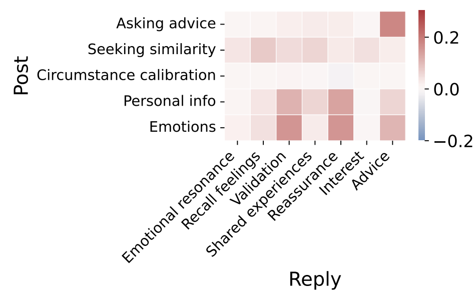
(c) r/work



(d) Parenting Stack Exchange



(e) Interpersonal Skills Stack Exchange



(f) The Workplace Stack Exchange

Figure 4: Correlation between requests/self-disclosure in posts and empathetic expressions in replies. Post: asking advice (asking for advice or opinions), seeking similarity (seeking similar experiences or feelings), circumstance calibration (seeking circumstance calibration), personal info (sharing personal information), emotions (sharing emotions). Reply: emotional resonance (expressing emotional resonance), recall feelings (recalling similar feelings or emotions), validation (expressing validation), shared experiences (sharing similar experiences), reassurance (offering reassurance, encouragement, or good wishes), interest (showing interest in further elaboration), advice (offering personalised advice).

across the various dimensions of empathy practices, rather than treating it as a single construct from the observer's perspective. Additionally, the pipeline developed in this study can be applied to other fields, providing insights for identifying other concepts in online spaces.

6.2 Effectiveness and Challenges in Using LMs for Empathy Detection (*RO 2*)

Detecting empathy in online texts has been shown to be a challenging task [47, 83, 98], and prior efforts on this task have typically relied on the generalised and oversimplified frameworks. To achieve *Research Objective 2*, which focuses on capturing multidimensional and relational empathy practices consistent with human-annotated evidence, we evaluated whether language models (LMs) could effectively detect a broad range of empathy practices grounded in our granular theoretical framework.

The results show that our best-performing model, Phi-3-Medium-Instruct, can effectively identify empathy practices in online communities after fine-tuning, achieving macro-F1 scores of 0.72–0.90 for all themes. These results indicate that the model can recognise empathic cues and capture the multidimensional structure of empathy practices encoded in our framework. The use of LMs allows for more customised configurations and extended context windows, and yields competitive results in this challenging field. In contrast to few-shot models based on general frameworks [109], our fine-tuned LMs, which are built on a more detailed and concrete theoretical foundation, demonstrate consistently higher performance in detecting both original posters' requests and repliers' expressions of empathy.

Although prior work has reported F1 scores for simpler binary or low-dimensional empathy categories, we do not conduct a formal hypothesis test against those results, as such comparisons would not be methodologically valid given the substantial mismatch in classification granularity. Earlier models classify only one or two categories, whereas our framework includes eight categories for posts and nine for replies. Nonetheless, our macro-F1 results demonstrate performance that is comparable to or better than prior systems [47, 54, 83, 98, 109] despite substantially higher label granularity, thereby providing evidence that our models successfully operationalise *Research Objective 2* by detecting fine-grained, relationally oriented empathy practices that align with human annotations. To further contextualise our performance, we include two types of baselines: basic baselines (random guessing and majority-class classifiers) and few-shot language models, which are widely used in computational social science. The strong performance gap between these baselines and our fine-tuned models further validates the robustness of our approach. Collectively, these findings demonstrate that LMs, when carefully aligned with a theoretically grounded schema, can serve as scalable tools for identifying relational empathy processes in online interactions. More broadly, this study highlights the potential of LMs to detect signals of complex constructs in online texts, particularly when guided by a well-specified theoretical framework. Our methodology is also adaptable to other domains where contextually grounded relational constructs are of interest.

However, not all LMs perform well in understanding nuanced empathy practices under all configurations. Effective application of LMs in this domain requires careful decisions about both model selection and input settings. In this study, models such as Mistral-Nemo-Instruct and Llama-3.1-8B-Instruct produced less accurate outputs, particularly for nuanced themes such as *expressing validation* and *showing interest in further elaboration*. These results highlight that not all LMs are equally capable of capturing the subtle cues necessary for *Research Objective 2*. Model performance can also vary depending on the input configuration. For example, adding more background information or rationale behind the classification can sometimes improve model performance, but excessive contextual information may distract the model from the target text and lead to lower accuracy. Thus, effective deployment of LMs in this domain requires a careful balance between providing helpful context and maintaining model focus to avoid confusion.

6.3 Contextual Variability of Empathy Practices Across Platforms and Communities (*RO 3*)

In addition to deconstructing and identifying online empathy practices, this study sought to understand how these practices manifest differently across platforms and whether such variation supports our hypotheses about contextual and community-level influences. Guided by *Research Objective 3*, our analyses demonstrate how platform and topical contexts condition the presence, form, and potential effects of empathy practices.

Consistent with *Hypothesis 3.1*, our cross-platform comparison reveals systematic divergence in empathy-seeking and empathy-giving patterns, supporting that platform architectures and interactional norms can shape users' communicative styles [36, 88]. Whereas Reddit's discussion-oriented affordances facilitate more emotionally expressive forms of empathy seeking and provision, Stack Exchange's design encourages informationally framed requests and concrete forms of support. These contrasts support the idea that empathy is not distributed uniformly but is sensitive to the communicative goals prioritised by each platform. While previous qualitative efforts or manual analyses of small datasets have suggested differences in empathy across various contexts [72, 99], our results extend this work by applying a more granular analytical framework and automatically identifying nuanced, platform- and community-specific patterns at scale. Such patterns reinforce the hypothesis that empathy practices are relational and responsive to community expectations. We also observe topic-level variability that further substantiates *Hypothesis 3.1*. Across the three community topics examined, communities oriented toward parenting or social skills show a higher prevalence of emotional and cognitive empathy practices in both seeking and provision, whereas work-focused communities display these practices less frequently and exhibit a greater emphasis on concrete advice seeking and problem-solving responses. These patterns indicate that topic-specific expectations, rather than platform identity alone, are associated with differences in how empathy-related behaviours occur.

Regarding outcome of empathy-seeking practices, our analyses provide partial support for Hypothesis 3.2, which posits that empathy-seeking features would be associated with more replies in a platform- and topic-specific manner. Instead of uniformly motivating engagement, the effects of empathy-seeking practices on reply behaviour are highly context dependent. Although some communities, most notably r/parenting and r/work, show broad increases in reply counts when posts include any form of request or self-disclosure, aligned with previous work [6, 14, 71, 89], other communities exhibit effects that are limited to specific dimensions of empathy seeking, where some empathy-seeking practices can even suppress engagement. These heterogeneous patterns emphasise that empathy-seeking cues elicit engagement only when they align with local norms of appropriateness and relevance, highlighting the conditional nature of their behavioural effects. The effects of empathy-giving practices also depend on platform norms, providing partial support for *Hypothesis 3.3*. On Reddit, empathic concern practices are reliably linked to increases in OP response rates, while on Stack Exchange these same practices exhibit neutral or even negative relationships with answer acceptance rates. A similar divergence emerges in the effects of cognitive and emotional empathy across communities.

We also examine the relationship between empathy practices in posts and replies. The findings suggest a systematic alignment between certain types of empathy-seeking and empathy-giving practices, such as *sharing emotions* in posts and *offering reassurance, encouragement, or good wishes* in replies. However, not all requests or self-disclosures are associated with increased expressions of empathy. This suggests that while some post features can act as strong social cues that invite empathy [11, 45, 45], their influence is contextually bounded and shaped by community norms and the perceived appropriateness of empathic engagement.

6.4 Design Implications

In addition to research implications, this study offers broader insights into platform design, community support practices, and the development of other systems. Rather than aiming to maximise empathy indiscriminately, our findings highlight the importance of context-sensitive engagement. This suggests that more tailored strategies may be necessary to enhance specific aspects of empathy.

First, our insights have broader implications for the design of online communities and suggest potential approaches for intervention. Engaging with empathy can be cognitively and emotionally demanding for peer-supporters [1, 8, 16, 17, 70], especially when they are not typically trained professionals [82]. By identifying nuanced themes in both empathy requests and responses, our framework offers targeted guidance on when and how specific empathy practices may be most appropriate in online interactions. For example, platform affordances or assistant systems could help reduce users' burdens by suggesting bullet points or rephrasings to support empathetic responses.

This also sheds light on the effectiveness of different approaches for soliciting empathy, as various components of empathy may be triggered by different mechanisms. Users who either desire or wish to avoid specific aspects of empathy can adjust their requests accordingly to suit their particular needs. For example, in goal-oriented environments where users prioritise efficiency and problem-solving, empathetic responses may be more limited or only welcomed when they align with the task at hand. In contrast, discussion-based platforms may afford greater flexibility for emotional expression and mutual support. These differences point to the need for platform-specific approaches that reflect the community's communication norms and support affordances.

For platform designers and community managers, our framework enables more adaptive and context-aware support strategies. It helps detect high-need posts, such as those involving emotional disclosure or requests, and identify responses that lack appropriately aligned empathy practices. In practice, platforms could provide integrated design affordances or plug-ins that users can adopt to offer more targeted support. Moreover, this framework can reduce the burden on community support teams and help with the development of scalable, semi-automated tools that assist in identifying sensitive posts and recommending context-appropriate actions. This can contribute to building more responsive and supportive communities while avoiding the cost of indiscriminate intervention.

Beyond post–reply interactions, this study is also relevant for broader systems such as conversational agents, recommender systems, and other interactive support tools. While prior work has focused on generating empathetic responses automatically [e.g., 18, 75, 81], our findings highlight the importance of grounding these responses in the specific context and needs of the user. Rather than aiming for generic expressions of empathy, systems should be designed to recognise the type of support being requested and respond accordingly. This approach can lead to more meaningful and appropriate interactions in support-oriented settings. Future work could build on this framework through experimental studies that assess how different types of empathy practices influence user outcomes across platforms. Such research could inform the design of context-aware and user-sensitive interaction design.

6.5 Ethical Considerations

The application of language models to detect and analyse empathy in online conversations raises several ethical concerns that require careful consideration. Empathy is highly subjective and shaped by context, culture, and individual interpretation. Though this study aims to reduce bias by breaking down online empathy into specific and well-defined themes and ensuring inter-annotator agreement, it still presents challenges that call for caution in future applications or adaptations. Moreover, people's preferences and understanding of empathy may evolve over time, which introduces additional risks. To help address this, we recommend testing models on a wide range of datasets and regularly evaluating their outputs for fairness and consistency.

It's also recommended to keep humans in the loop in any practical application, especially in areas like mental health support or response refinement where decisions can carry significant consequences. Rather than relying solely on models to make final decisions, these systems could assist humans by highlighting relevant parts of a conversation or offering possible rationales. This helps maintain accountability while reducing the risk of over-reliance on potentially biased or oversimplified outputs.

If language models are integrated into online platforms or other systems, it is essential to be transparent about how they are involved, e.g., whether they are generating responses, flagging segments, or recommending posts. Users should always be informed when an AI system contributes to content creation or affects how content is surfaced or managed. This transparency helps set appropriate expectations and allows users to make informed judgments about the content they see. Misrepresenting AI-generated content as human-authored undermines trust and poses ethical risks.

7 Limitations

This study developed and applied a granular framework to deconstruct empathy practices in online discourse, which enables a consistent understanding of their structure among annotators. The overall evaluation is still from third-person perspectives, however, and our identification of empathy practices is therefore limited to more explicit and direct expressions to avoid confusion. Further work could collect first-person reports from both seekers and repliers to better understand the mental states behind their textual interactions and the underlying mechanisms of empathy exchange. This approach may also help identify more implicit signals of empathy, which are common and contribute significantly to online communication.

Our results demonstrate the potential of fine-tuned language models in detecting online empathy. However, as our dataset is drawn from communities that do not equally emphasise all dimensions of empathy, some themes are underrepresented. This imbalance may limit the direct applicability of our models to those specific themes. Specifically, we focused our analysis on themes with sufficient data to robustly evaluate model performance, and therefore excluded the category of *seeking emotional support* from our analysis due to its low frequency in the dataset. Importantly, this exclusion does not reflect a dismissal of the theme's relevance, but rather aligns with our goal of examining empathy practices in comparable, everyday contexts where empathy is plausible but not structurally mandated. Future work in contexts that specifically target this theme (e.g., online therapy platforms and mental health communities) may require separate modelling approaches and may not be able to directly use our fine-tuned models. Despite this, such research can still apply the framework and modelling pipeline proposed in this study and develop models tailored to identifying this specific empathy practice.

Moreover, this study collected high-quality annotated data including both classification results for each theme and highlights of supporting segments. While we tested the usefulness of these highlights for the classification task, identifying the highlights themselves could be a valuable task. Further work can apply the datasets to predict specific expressions of each theme and analyse the patterns of online empathy in various contexts. This could provide deeper insights into the quality and specificity of empathy practices in online interactions.

Finally, this study focuses on the conversation-level effects of empathetic interactions. An important direction for future work is to explore how such exchanges influence long-term user engagement, retention, and overall community health. Further analysis using longitudinal data, such as survival analysis or role identification methods, could help assess whether receiving or expressing empathy contributes to sustained participation and supports more inclusive and supportive community environments. This may also provide valuable insights into the downstream effects of empathy in digital spaces.

8 Conclusion

This study advances the understanding of empathy in digital interactions by demonstrating that it is not a uniform or inherently positive phenomenon, but rather a multidimensional and relational practice. We proposed a nuanced framework through thematic analysis that captures how empathy is solicited in posts and expressed in replies across distinct dimensions. We then fine-tuned language models (LMs) to recognise these distinct empathy practices. This study demonstrates the effectiveness of LMs in detecting such complex and nuanced constructs compared to baselines, while also highlighting their limitations and risks. We subsequently applied the best-performing LMs to a large-scale dataset, and the results of the statistical analyses challenge the conventional view of empathy as inherently positive and supportive. Instead, this study reveals that platform- and topic-specific contexts play a crucial role in shaping the appropriateness and impact of empathy practices across varied online communities. These findings enhance our understanding of empathy in online interactions across various stages (requests, replies, and follow-up responses) and perspectives (seekers and providers). They also offer practical insights for designing online platforms and support systems that foster more context-sensitive, responsive interactions, improve onboarding for new users, and enable scalable monitoring of community health.

Acknowledgments

We are thankful for funding from The Alan Turing Institute within the project entitled "Effective discovery, tracking, and response to mis- and disinformation" and from the ESRC Digital Good Network (grant reference ES/X502352/1). This first author was supported by the Economic and Social Research Council [grant number ES/P000649/1]. We also thank Label Studio (https://labelstud.io) for providing the annotation interface used in this research.

References

[1] 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*, Vol. 11. 448–451.

[2] Chandan Akiti, Anna Squicciarini, and Sarah Rajtmajer. 2020. A semantics-based approach to disclosure classification in user-generated online content. In *Findings of the Association for Computational Linguistics: EMNLP 2020*. 3490–3499.

[3] Nazanin Andalibi and Andrea Forte. 2018. Responding to sensitive disclosures on social media: A decision-making framework. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 6 (2018), 1–29.

[4] Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. 2018. Social support, reciprocity, and anonymity in responses to sexual abuse disclosures on social media. *ACM Transactions on Computer-Human Interaction (TOCHI)* 25, 5 (2018), 1–35.

[5] Nazanin Andalibi, Oliver L Haimson, Munmun De Choudhury, and Andrea Forte. 2016. Understanding social media disclosures of sexual abuse through the lenses of support seeking and anonymity. In *Proceedings of the 2016 CHI conference on human factors in computing systems*. 3906–3918.

[6] Nazanin Andalibi, Margaret E Morris, and Andrea Forte. 2018. Testing waters, sending clues: Indirect disclosures of socially stigmatized experiences on social media. *Proceedings of the ACM on Human-Computer Interaction* 2, CSCW (2018), 1–23.

[7] Ashton Anderson, Daniel Huttenlocher, Jon Kleinberg, and Jure Leskovec. 2012. Discovering value from community activity on focused question answering sites: a case study of stack overflow. In *Proceedings of the 18th ACM SIGKDD international conference on Knowledge discovery and data mining*. 850–858.

[8] James Andreoni, Justin M Rao, and Hannah Trachtman. 2017. Avoiding the ask: A field experiment on altruism, empathy, and charitable giving. *Journal of political Economy* 125, 3 (2017), 625–653.

[9] Robert L Barker. 2003. *The social work dictionary*. NASW press.

[10] Julia Barlińska, Anna Szuster, and Mikołaj Winiewski. 2013. Cyberbullying among adolescent bystanders: Role of the communication medium, form of violence, and empathy. *Journal of Community & Applied Social Psychology* 23, 1 (2013), 37–51.

[11] C Daniel Batson. 2009. These things called empathy: eight related but distinct phenomena. (2009).

[12] C Daniel Batson. 2014. *The altruism question: Toward a social-psychological answer*. Psychology Press.

[13] Judith Baxter. 2018. 'Keep strong, remember everything you have learnt': Constructing support and solidarity through online interaction within a UK cancer support group. *Discourse & society* 29, 4 (2018), 363–379.

[14] Lindsay Blackwell, Jill Dimond, Sarita Schoenebeck, and Cliff Lampe. 2017. Classification and its consequences for online harassment: Design insights from heartmob. *Proceedings of the ACM on Human-Computer Interaction* 1, CSCW (2017), 1–19.

[15] Virginia Braun and Victoria Clarke. 2006. Using thematic analysis in psychology. *Qualitative research in psychology* 3, 2 (2006), 77–101.

[16] C Daryl Cameron, Lasana T Harris, and B Keith Payne. 2016. The emotional cost of humanity: Anticipated exhaustion motivates dehumanization of stigmatized targets. *Social Psychological and Personality Science* 7, 2 (2016), 105–112.

[17] C Daryl Cameron, Cendri A Hutcherson, Amanda M Ferguson, Julian A Scheffer, Eliana Hadjiandreou, and Michael Inzlicht. 2019. Empathy is hard work: People choose to avoid empathy because of its cognitive costs. *Journal of Experimental Psychology: General* 148, 6 (2019), 962.

[18] Jacky Casas, Timo Spring, Karl Daher, Elena Mugellini, Omar Abou Khaled, and Philippe Cudré-Mauroux. 2021. Enhancing conversational agents with empathic abilities. In *Proceedings of the 21st ACM international conference on intelligent virtual agents*. 41–47.

[19] Yixin Chen, Scott Hale, and Bernie Hogan. 2024. "I Am 30F and Need Advice!": A Mixed-Method Analysis of the Effects of Advice-Seekers' Self-Disclosure on Received Replies. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 18. 276–288.

[20] Yixin Chen and Yang Xu. 2021. Social support is contagious: Exploring the effect of social support in online mental health communities. In *Extended abstracts of the 2021 CHI conference on human factors in computing systems*. 1–6.

[21] Chao-Min Chiu, Hsin-Yi Huang, Hsiang-Lan Cheng, and Pei-Chen Sun. 2015. Understanding online community citizenship behaviors through social support and social identity. *International journal of information management* 35, 4 (2015), 504–519.

[22] Tsz Hang Chu, Youzhen Su, Hanxiao Kong, Jingyuan Shi, and Xiaohui Wang. 2021. Online social support for intimate partner violence victims in China: quantitative and automatic content analysis. *Violence against women* 27, 3-4 (2021), 339–358.

[23] Douglas Cohen and Janet Strayer. 1996. Empathy in conduct-disordered and comparison youth. *Developmental psychology* 32, 6 (1996), 988.

[24] Jacob Cohen. 1988. *Statistical power analysis for the behavioral sciences*. Lawrence Erlbaum Associates. 79–81 pages.

[25] Andrew M Colman. 2015. *A dictionary of psychology*. Oxford University Press.

[26] Iain J Cruickshank and Lynnette Hui Xian Ng. 2023. Prompting and fine-tuning open-sourced large language models for stance classification. *arXiv preprint arXiv:2309.13734* (2023).

[27] Andrea Cuadra, Maria Wang, Lynn Andrea Stein, Malte F Jung, Nicola Dell, Deborah Estrin, and James A Landay. 2024. The illusion of empathy? notes on displays of emotion in human-computer interaction. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–18.

[28] Tiago Oliveira Cunha, Ingmar Weber, Hamed Haddadi, and Gisele L Pappa. 2016. The effect of social feedback in a reddit weight loss community. In *Proceedings of the 6th international conference on digital health conference*. 99–103.

[29] Mark H Davis. 1980. A multidimensional approach to individual differences in empathy. (1980).
[30] Mark H Davis. 2018. *Empathy: A social psychological approach.* Routledge.
[31] Jean Decety and Jason M Cowell. 2014. The complex relation between morality and empathy. *Trends in cognitive sciences* 18, 7 (2014), 337–339.
[32] Jean Decety and Philip L Jackson. 2004. The functional architecture of human empathy. *Behavioral and cognitive neuroscience reviews* 3, 2 (2004), 71–100.
[33] Jean Decety and Keith J Yoder. 2016. Empathy and motivation for justice: Cognitive empathy and concern, but not emotional empathy, predict sensitivity to injustice for others. *Social neuroscience* 11, 1 (2016), 1–14.
[34] Melissa DeJonckheere, Lisa M Vaughn, Tyler G James, and Amanda C Schondelmeyer. 2024. Qualitative thematic analysis in a mixed methods study: Guidelines and considerations for integration. *Journal of Mixed Methods Research* 18, 3 (2024), 258–269.
[35] Tim Dettmers, Artidoro Pagnoni, Ari Holtzman, and Luke Zettlemoyer. 2024. Qlora: Efficient finetuning of quantized llms. *Advances in Neural Information Processing Systems* 36 (2024).
[36] Himel Dev, Karrie Karahalios, and Hari Sundaram. 2019. Quantifying voter biases in online platforms: An instrumental variable approach. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–27.
[37] Nancy Eisenberg and Janet Strayer. 1990. *Empathy and its development.* CUP Archive.
[38] Hanmei Fan, Reeva Lederman, Stephen P Smith, and Shanton Chang. 2014. How trust is formed in online health communities: a process perspective. *Communications of the Association for Information Systems* 34, 1 (2014), 28.
[39] Anthony Vincent Fernandez and Dan Zahavi. 2020. Basic empathy: Developing the concept of empathy from the ground up. *International Journal of Nursing Studies* 110 (2020), 103695.
[40] Alvin I Goldman. 1993. Ethics and cognitive science. *Ethics* 103, 2 (1993), 337–360.
[41] Ilker Gül, Rémi Lebret, and Karl Aberer. 2024. Stance detection on social media with fine-tuned large language models. *arXiv preprint arXiv:2404.12171* (2024).
[42] Jeong Yeob Han, Eunkyung Kim, Yen-I Lee, Dhavan V Shah, and David H Gustafson. 2019. A longitudinal investigation of empathic exchanges in online cancer support groups: Message reception and expression effects on patients' psychosocial health outcomes. *Journal of health communication* 24, 6 (2019), 615–623.
[43] Jeong Yeob Han, Dhavan V Shah, Eunkyung Kim, Kang Namkoong, Sun-Young Lee, Tae Joon Moon, Rich Cleland, Q Lisa Bu, Fiona M McTavish, and David H Gustafson. 2011. Empathic exchanges in online cancer support groups: distinguishing message expression and reception effects. *Health communication* 26, 2 (2011), 185–197.
[44] Grit Hein and Tania Singer. 2008. I feel how you feel but not always: the empathic brain and its modulation. *Current opinion in neurobiology* 18, 2 (2008), 153–158.
[45] CT Rodríguez Hidalgo, Eduard S-H Tan, and Peeter WJ Verlegh. 2015. The social sharing of emotion (SSE) in online social networks: A case study in Live Journal. *Computers in Human Behavior* 52 (2015), 364–372.
[46] Robert Hogan. 1969. Development of an empathy scale. *Journal of consulting and clinical psychology* 33, 3 (1969), 307.
[47] Mahshid Hosseini and Cornelia Caragea. 2021. Distilling knowledge for empathy detection. In *Findings of the Association for Computational Linguistics: EMNLP 2021*. 3713–3724.
[48] Shiyuan Huang, Siddarth Mamidanna, Shreedhar Jangam, Yilun Zhou, and Leilani H Gilpin. 2023. Can large language models explain themselves? a study of llm-generated self-explanations. *arXiv preprint arXiv:2310.11207* (2023).
[49] Quentin Jones. 1997. Virtual-communities, virtual settlements & cyber-archaeology: A theoretical outline. *Journal of Computer-Mediated Communication* 3, 3 (1997), JCMC331.
[50] Megan Knittel, Faye Kollig, Abrielle Mason, and Rick Wash. 2021. Anyone else have this experience: Sharing the emotional labor of tracking data about me. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW1 (2021), 1–30.
[51] Allison Lahnala, Charles Welch, David Jurgens, and Lucie Flek. 2022. A Critical Reflection and Forward Perspective on Empathy and Natural Language Processing. In *Findings of the Association for Computational Linguistics: EMNLP 2022*. 2139–2158.
[52] Allison Lahnala, Charles Welch, Béla Neuendorf, and Lucie Flek. 2022. Mitigating Toxic Degeneration with Empathetic Data: Exploring the Relationship Between Toxicity and Empathy. In *Proceedings of the 2022 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*. 4926–4938.
[53] JR Landis. 1977. The Measurement of Observer Agreement for Categorical Data. *Biometrics* (1977).
[54] Gyeongeun Lee, Christina Wong, Meghan Guo, and Natalie Parde. 2024. Pouring Your Heart Out: Investigating the Role of Figurative Language in Online Expressions of Empathy. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*. 519–529.
[55] Andrea C Lewallen, Jason E Owen, Erin O'Carroll Bantum, and Annette L Stanton. 2014. How language affects peer responsiveness in an online cancer support group: implications for treatment design and facilitation. *Psycho-Oncology* 23, 7 (2014), 766–772.
[56] Zhaojiang Lin, Peng Xu, Genta Indra Winata, Farhad Bin Siddique, Zihan Liu, Jamin Shin, and Pascale Fung. 2020. Caire: An end-to-end empathetic chatbot. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 34. 13622–13623.
[57] Emmy Liu, Chen Cui, Kenneth Zheng, and Graham Neubig. 2022. Testing the ability of language models to interpret figurative language. *arXiv preprint arXiv:2204.12632* (2022).
[58] Zhiwei Liu, Kailai Yang, Qianqian Xie, Tianlin Zhang, and Sophia Ananiadou. 2024. Emollms: A series of emotional large language models and annotation tools for comprehensive affective analysis. In *Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining*. 5487–5496.
[59] Lany Laguna Maceda, Jennifer Laraya Llovido, Miles Biago Artiaga, and Mideth Balawiswis Abisado. 2023. Classifying sentiments on social media texts: A gpt-4 preliminary study. In *Proceedings of the 2023 7th International Conference on Natural Language Processing and Information Retrieval*. 19–24.
[60] Alexandra Main and Carmen Kho. 2020. A relational framework for integrating the study of empathy in children and adults. *Emotion Review* 12, 4 (2020), 280–290.
[61] Stephann Makri and Sophie Turner. 2020. "I can't express my thanks enough": The "gratitude cycle" in online communities. *Journal of the Association for Information Science and Technology* 71, 5 (2020), 503–515.
[62] Paul Marshall, Millissa Booth, Matthew Coole, Lauren Fothergill, Zoe Glossop, Jade Haines, Andrew Harding, Rose Johnston, Steven Jones, Christopher Lodge, et al. 2024. Understanding the impacts of online mental health peer support forums: realist synthesis. *JMIR Mental Health* 11 (2024), e55750.
[63] Sohaib Mustafa, Wen Zhang, and Muhammad Mateen Naveed. 2023. What motivates online community contributors to contribute consistently? A case study on Stackoverflow netizens. *Current Psychology* 42, 13 (2023), 10468–10481.
[64] Priya Nambisan. 2011. Evaluating patient experience in online health communities: implications for health care organizations. *Health care management review* 36, 2 (2011), 124–133.
[65] Ayushi Nirmal, Amrita Bhattacharjee, Paras Sheth, and Huan Liu. 2024. Towards interpretable hate speech detection using large language model-extracted rationales. *arXiv preprint arXiv:2403.12403* (2024).
[66] Morris A Okun, Stephanie A Shepard, and Nancy Eisenberg. 2000. The relations of emotionality and regulation to dispositional empathy-related responding among volunteers-in-training. *Personality and Individual Differences* 28, 2 (2000), 367–382.
[67] Lucy Osler. 2021. Taking empathy online. *Inquiry* (2021), 1–28.
[68] Anat Perry. 2023. AI will never convey the essence of human empathy. *Nature Human Behaviour* (2023), 1–2.
[69] Dominik Petko, Nives Egger, Felix Schmitz, Alexandra Totter, Thomas Hermann, and Sissel Guttormsen. 2015. Coping through blogging: A review of studies on the potential benefits of weblogs for stress reduction. *Cyberpsychology-journal of psychosocial research on cyberspace* 9, 2 (2015).
[70] Ulrike Pfeil and Panayiotis Zaphiris. 2007. Patterns of empathy in online communication. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 919–928.
[71] Jenny Preece. 1998. Empathic communities: Reaching out across the web. *interactions* 5, 2 (1998), 32–43.
[72] Jenny Preece and Kambiz Ghozati. 1998. In search of empathy online: A review of 100 online communities. *Proceedings of the Americas Conference on Information Systems* (1998).
[73] Jennifer Preece and Kambiz Ghozati. 2001. Experiencing empathy online. *The Internet and health communication: Experiences and expectations* 1 (2001), 147–166.
[74] Jenny Preece and Diane Maloney-Krichmar. 2003. Online communities: focusing on sociability and usability. *Handbook of human-computer interaction* (2003), 596–620.
[75] Yushan Qian, Wei-Nan Zhang, and Ting Liu. 2023. Harnessing the power of large language models for empathetic response generation: Empirical investigations and improvements. *arXiv preprint arXiv:2310.05140* (2023).
[76] Baojun Qiu, Kang Zhao, Prasenjit Mitra, Dinghao Wu, Cornelia Caragea, John Yen, Greta E Greer, and Kenneth Portier. 2011. Get online support, feel better–sentiment analysis and dynamics in an online cancer survivor community. In *2011 IEEE third international conference on privacy, security, risk and trust and 2011 IEEE third international conference on social computing*. IEEE, 274–281.
[77] Stephen A Rains and Valerie Young. 2009. A meta-analysis of research on formal computer-mediated support groups: Examining group characteristics and health outcomes. *Human communication research* 35, 3 (2009), 309–336.
[78] E Soo Rhee and Hyang-Sook Kim. 2023. Understanding the Dynamics of Online Social Support Among Postpartum Mothers in Online Communities. *Maternal and Child Health Journal* 27, 4 (2023), 690–697.
[79] Joseph Seering, Felicia Ng, Zheng Yao, and Geoff Kaufman. 2018. Applications of social identity theory to research and design in computer-supported cooperative

work. *Proceedings of the ACM on human-computer interaction* 2, CSCW (2018), 1–34.
[80] Shahid Munir Shah, Syeda Anshrah Gillani, Mirza Samad Ahmed Baig, Muhammad Aamer Saleem, and Muhammad Hamzah Siddiqui. 2025. Advancing depression detection on social media platforms through fine-tuned large language models. *Online Social Networks and Media* 46 (2025), 100311.
[81] 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*. 194–205.
[82] Ashish Sharma, Inna W Lin, Adam S Miner, David C Atkins, and Tim Althoff. 2023. Human–AI collaboration enables more empathic conversations in text-based peer-to-peer mental health support. *Nature Machine Intelligence* 5, 1 (2023), 46–57.
[83] Ashish Sharma, Adam Miner, David Atkins, and Tim Althoff. 2020. A Computational Approach to Understanding Empathy Expressed in Text-Based Mental Health Support. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 5263–5276.
[84] Eva Sharma and Munmun De Choudhury. 2018. Mental health support and its relationship to linguistic accommodation in online communities. In *Proceedings of the 2018 CHI conference on human factors in computing systems*. 1–13.
[85] Inc. Stack Exchange. 2023. Stack exchange data dump. https://archive.org/details/stackexchange
[86] Samuel Hardman Taylor, Dominic DiFranzo, Yoon Hyung Choi, Shruti Sannon, and Natalya N Bazarova. 2019. Accountability and empathy by design: Encouraging bystander intervention to cyberbullying on social media. *Proceedings of the ACM on Human-Computer Interaction* 3, CSCW (2019), 1–26.
[87] Maxim Tkachenko, Mikhail Malyuk, Andrey Holmanyuk, and Nikolai Liubimov. 2020-2024. Label Studio: Data labeling software. https://labelstud.io Open source software available from https://github.com/HumanSignal/label-studio.
[88] Trang Tran and Mari Ostendorf. 2016. Characterizing the Language of Online Communities and its Relation to Community Reception. In *Proceedings of the 2016 Conference on Empirical Methods in Natural Language Processing*. 1030–1035.
[89] Olivier Turbide, Maria Cherba, and Vincent Denault. 2019. Responding to self-disclosure in an online discussion forum for people living with cancer: A conversational approach. In *The 10th International Conference on Social Media & Society*.
[90] Jirassaya Uttarapong, Nina LaMastra, Reesha Gandhi, Yu-hao Lee, Chien Wen Yuan, and Donghee Yvette Wohn. 2022. Twitch Users' Motivations and Practices during Community Mental Health Discussions. *Proceedings of the ACM on Human-Computer Interaction* 6, GROUP (2022), 1–23.
[91] Jolanda Van Dijke, Inge van Nistelrooij, Pien Bos, and Joachim Duyndam. 2020. Towards a relational conceptualization of empathy. *Nursing Philosophy* 21, 3 (2020), e12297.
[92] Shaowei Wang, Tse-Hsun Chen, and Ahmed E Hassan. 2018. Understanding the factors for fast answers in technical Q&A websites: An empirical study of four stack exchange websites. *Empirical Software Engineering* 23, 3 (2018), 1552–1593.
[93] Xi Wang, Kang Zhao, and Nick Street. 2017. Analyzing and predicting user participations in online health communities: a social support perspective. *Journal of medical Internet research* 19, 4 (2017), e6834.
[94] Watchful1. 2023. *Subreddit comments/submissions 2005-06 to 2022-12*. https://www.reddit.com/r/pushshift/comments/11ef9if/separate_dump_files_for_the_top_20k_subreddits/
[95] Lauren Wispé. 1986. The distinction between sympathy and empathy: To call forth a concept, a word is needed. *Journal of personality and social psychology* 50, 2 (1986), 314.
[96] Adam Worrall, Alicia Cappello, and Rachel Osolen. 2021. The importance of socio-emotional considerations in online communities, social informatics, and information science. *Journal of the Association for Information Science and Technology* 72, 10 (2021), 1247–1260.
[97] Diyi Yang and Scott Counts. 2018. Understanding self-narration of personally experienced racism on reddit. In *Proceedings of the International AAAI Conference on Web and Social Media*, Vol. 12.
[98] Jiamin Yang and David Jurgens. 2024. Modeling Empathetic Alignment in Conversation. In *Proceedings of the 2024 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies (Volume 1: Long Papers)*. 3127–3148.
[99] Zheng Yao, Silas Weden, Lea Emerlyn, Haiyi Zhu, and Robert E Kraut. 2021. Together but alone: Atomization and peer support among gig workers. *Proceedings of the ACM on Human-Computer Interaction* 5, CSCW2 (2021), 1–29.
[100] Hua Jonathan Ye, Yuanyue Feng, and Ben CF Choi. 2015. Understanding knowledge contribution in online knowledge communities: A model of community support and forum leader support. *Electron. Commer. Res. Appl.* 14, 1 (2015), 34–45.
[101] Woohyun Yoo, Ming-Yuan Chih, Min-Woo Kwon, JungHwan Yang, Eunji Cho, Bryan McLaughlin, Kang Namkoong, Dhavan V Shah, and David H Gustafson. 2013. Predictors of the change in the expression of emotional support within an online breast cancer support group: a longitudinal study. *Patient education and counseling* 90, 1 (2013), 88–95.
[102] Jamil Zaki. 2014. Empathy: a motivated account. *Psychological bulletin* 140, 6 (2014), 1608.
[103] Jamil Zaki and Kevin N Ochsner. 2012. The neuroscience of empathy: progress, pitfalls and promise. *Nature neuroscience* 15, 5 (2012), 675–680.
[104] Jiang Zhang, Qiong Wu, Yiming Xu, Cheng Cao, Zheng Du, and Konstantinos Psounis. 2024. Efficient toxic content detection by bootstrapping and distilling large language models. In *Proceedings of the AAAI conference on artificial intelligence*, Vol. 38. 21779–21787.
[105] Renwen Zhang, Jordan Eschler, and Madhu Reddy. 2018. Online support groups for depression in China: Culturally shaped interactions and motivations. *Computer Supported Cooperative Work (CSCW)* 27 (2018), 327–354.
[106] Jing Zhao, Kathleen Abrahamson, James G Anderson, Sejin Ha, and Richard Widdows. 2013. Trust, empathy, social identity, and contribution of knowledge within patient online communities. *Behaviour & Information Technology* 32, 10 (2013), 1041–1048.
[107] Naitian Zhou and David Jurgens. 2020. Condolence and empathy in online communities. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*. 609–626.
[108] Tao Zhou. 2020. Understanding users' participation in online health communities: a social capital perspective. *Information Development* 36, 3 (2020), 403–413.
[109] Caleb Ziems, William Held, Omar Shaikh, Jiaao Chen, Zhehao Zhang, and Diyi Yang. 2024. Can large language models transform computational social science? *Computational Linguistics* 50, 1 (2024), 237–291.

A Annotation Questions and Interface

Figure 5 shows an example of the annotation interface and annotation questions.

B Experimental Setup

B.1 Dataset Split and Theme Frequency

We randomly split the dataset into train, dev, and test sets with a ratio of 6:2:2. Table 10 shows the frequency of each theme in these splits.

B.2 Fine-Tuning Hyperparameters

- Learning rate: [1e-5, 2e-5, 3e-5, 4e-5, 5e-5]
- Weight decay (λ): 0.01
- Effective batch size: $8 \times 4 = 32$
- Maximum sequence length: 2048
- LoRA r: 16
- LoRA alpha: 16

C Few-Shot Learning Results

We conducted few-shot learning experiments across all themes. The results of the few-shot models are presented in Table 11, and the performance gain from fine-tuning (i.e., the improvement over the few-shot learning model) is shown in Table 12.

D Dataset for Analysis of Empathy in Online Communities

Table 13 summarises the dataset used for the analysis of empathy in online communities, as discussed in Section 5.3.

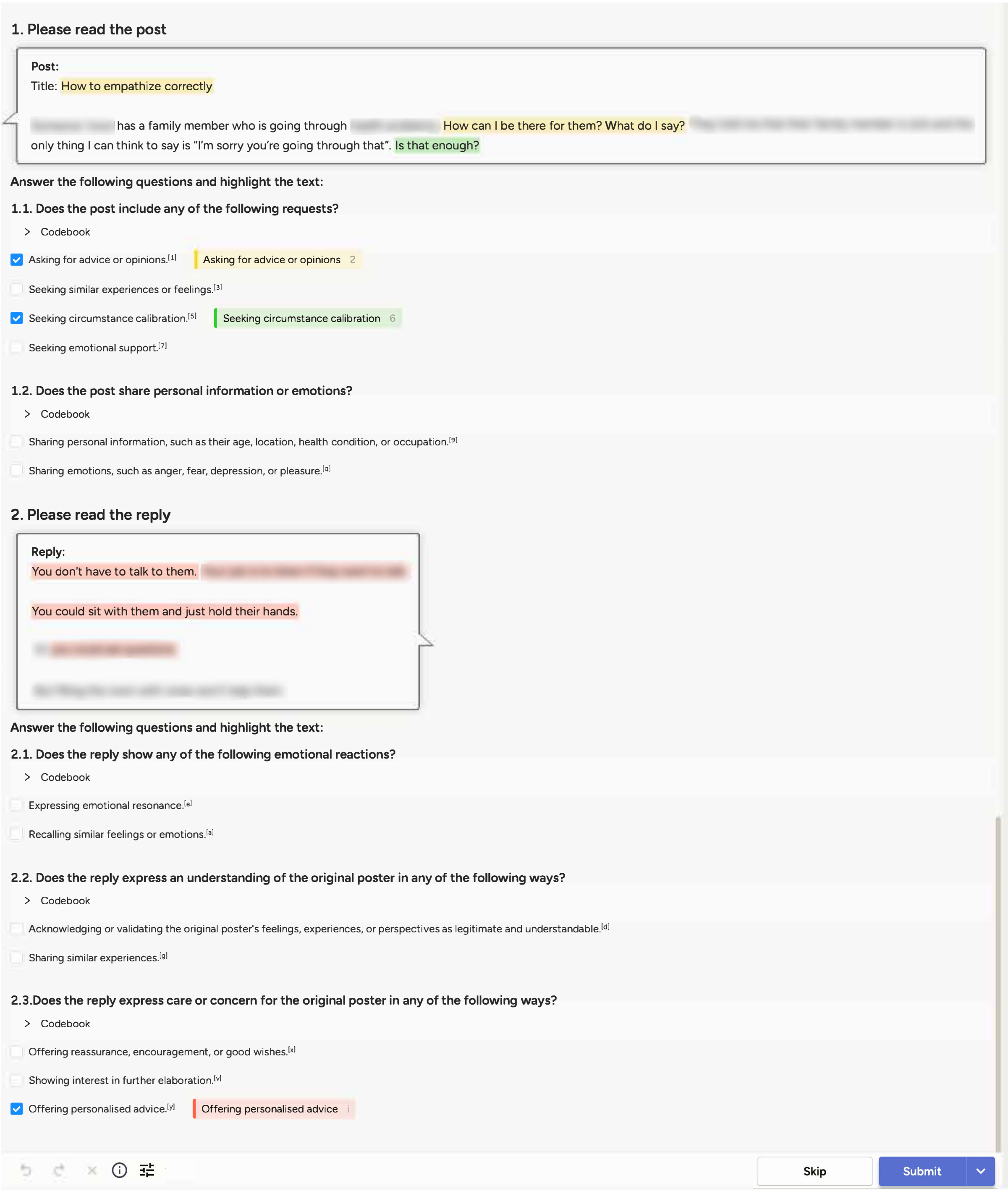


Figure 5: Screenshot of the annotation interface on Label Studio.

Category	Theme	Train	Dev	Test
Request	Asking for advice or opinions	650	223	218
	Seeking similar experiences or feelings	54	19	15
	Seeking circumstance calibration	88	33	28
	Seeking emotional support	9	3	0
Self-disclosure	Sharing personal information	204	60	67
	Sharing emotions	334	113	101
	Total	781	260	261

(a) Post ($N = 1,302$)

Category	Theme	Train	Dev	Test
Emotional empathy	Expressing emotional resonance	41	6	13
	Recalling similar feelings or emotions	39	8	15
Cognitive empathy	Expressing validation	38	15	16
	Sharing similar experiences	144	44	41
Empathic concern	Offering reassurance, encouragement, or good wishes	116	43	41
	Showing interest in further elaboration	68	22	21
	Offering personalised advice	598	200	202
	Total	756	253	252

(b) Reply ($N = 1,261$)

Table 10: Themes distribution across the train, dev, and test sets. The total number of replies is less than the total number of posts in the annotated dataset because the dataset includes both post–reply pairs and posts with no replies. The theme "seeking emotional support" is not included in further analysis due to a lack of data points.

			Phi-3-Medium-Instruct				Llama-3.1-8B-Instruct				Mistral-Nemo-Instruct			
Category		Input	reply/post class	reply/post +quotes	reply+post class	reply+post +quotes	reply/post class	reply/post +quotes	reply+post class	reply+post +quotes	reply/post class	reply/post +quotes	reply+post class	reply+post +quotes
Post	*Request*													
		Asking for advice or opinions	**88.65**	86.43	-	-	**69.87**	52.32	-	-	**52.32**	45.51	-	-
		Seeking similar experiences or feelings	**76.51**	71.23	-	-	**6.70**	5.86	-	-	**27.14**	5.43	-	-
		Seeking circumstance calibration	**79.29**	76.33	-	-	**21.46**	12.85	-	-	**37.56**	9.69	-	-
		Seeking emotional support	-	-	-	-	-	-	-	-	-	-	-	-
	Self-disclosure													
		Sharing personal information	54.07	**56.58**	-	-	**56.35**	42.43	-	-	**45.05**	25.44	-	-
		Sharing emotions	77.84	**81.14**	-	-	**53.93**	40.38	-	-	**65.14**	35.21	-	-
Reply	*Emotional empathy*													
		Expressing emotional resonance	**65.85**	60.44	56.17	53.79	28.17	16.33	**29.88**	24.69	**39.49**	13.78	35.17	9.92
		Recalling similar feelings or emotions	68.76	**74.94**	69.75	71.65	**18.28**	8.64	11.11	11.11	**28.12**	7.36	18.28	6.93
	Cognitive empathy													
		Expressing validation	**53.19**	47.85	39.96	39.96	**29.24**	29.00	13.88	14.65	35.57	17.92	**36.86**	11.51
		Sharing similar experiences	77.77	76.60	**81.43**	80.62	**40.73**	21.39	19.01	15.54	**63.37**	18.04	51.09	14.51
	Empathic concern													
		Offering reassurance, encouragement, or good wishes	72.10	71.17	**74.87**	74.73	40.73	32.94	31.34	27.61	60.88	43.41	**63.41**	34.88
		Showing interest in further elaboration	**70.78**	69.20	64.96	63.10	26.27	11.34	**29.48**	13.03	**55.80**	18.62	45.33	9.51
		Offering personalised advice	55.39	**61.82**	57.08	59.71	58.60	**67.27**	54.92	63.06	34.77	**78.03**	29.52	76.54

Table 11: Classification results (macro-F1) of few-shot models across model structures and input combinations. The best performance of input combination within each model structure is in bold, and the best model for each theme is highlighted in orange .

			Phi-3-Medium-Instruct				Llama-3.1-8B-Instruct				Mistral-Nemo-Instruct			
Category		Input	reply/post class	reply/post +quotes	reply+post class	reply+post +quotes	reply/post class	reply/post +quotes	reply+post class	reply+post +quotes	reply/post class	reply/post +quotes	reply+post class	reply+post +quotes
Post	*Request*													
		Asking for advice or opinions	-1.94	0.89	-	-	10.22	7.48	-	-	32.34	39.73	-	-
		Seeking similar experiences or feelings	12.17	18.78	-	-	61.18	42.66	-	-	45.57	80.90	-	-
		Seeking circumstance calibration	0.71	2.91	-	-	44.73	37.86	-	-	9.61	64.06	-	-
		Seeking emotional support	-	-	-	-	-	-	-	-	-	-	-	-
	Self-disclosure													
		Sharing personal information	25.19	24.34	-	-	10.26	21.59	-	-	27.37	43.29	-	-
		Sharing emotions	5.31	0.02	-	-	23.37	39.07	-	-	15.31	47.08	-	-
Reply	*Emotional empathy*													
		Expressing emotional resonance	6.80	13.31	12.26	11.21	26.21	32.35	25.92	23.99	29.89	50.91	13.51	53.62
		Recalling similar feelings or emotions	2.85	-10.49	1.04	-1.00	35.45	39.83	37.36	37.36	30.86	55.35	30.19	41.53
	Cognitive empathy													
		Expressing validation	16.31	24.58	16.98	16.96	24.49	27.25	34.48	33.71	22.55	36.97	18.98	40.69
		Sharing similar experiences	3.66	3.46	1.75	2.38	28.58	34.86	33.55	39.88	-5.56	54.25	23.85	42.93
	Empathic concern													
		Offering reassurance, encouragement, or good wishes	2.85	5.58	3.18	4.24	31.23	39.03	27.54	41.40	8.46	23.90	4.85	41.24
		Showing interest in further elaboration	2.52	3.24	-1.32	1.02	27.83	36.49	22.45	37.35	3.11	44.09	6.54	51.03
		Offering personalised advice	23.26	17.96	19.46	18.63	15.23	10.60	15.21	13.49	38.48	1.66	45.19	1.65

Table 12: Performance gain from fine-tuning compared to few-shot learning.

		Reddit	Stack Exchange
Parenting		r/Parenting	Parenting Stack Exchange
	Post	3,572	3,572
	Reply	38,173	11,198
Social Skills		r/socialskills	Interpersonal Skills Stack Exchange
	Post	3,572	3,572
	Reply	16,061	12,079
Working		r/work	The Workplace Stack Exchange
	Post	3,572	3,572
	Reply	7,917	11,643

Table 13: Summary of the sampled dataset for analysis.