

Design Considerations for an NLP-Driven Empathy and Emotion Interface for Clinician Training via Telemedicine

Roxana Girju

Departments of Linguistics & Computer Science,
University of Illinois at Urbana-Champaign
girju@illinois.edu

Marina Girju

Jabs School of Business,
California Baptist University
mgirju@calbaptist.edu

Abstract

As digital social platforms and mobile technologies become more prevalent and robust, the use of Artificial Intelligence (AI) in facilitating human communication will grow. This, in turn, will encourage development of intuitive, adaptive, and effective empathic AI interfaces that better address the needs of socially and culturally diverse communities. In this paper, we present several design considerations of an intelligent digital interface intended to guide the clinicians toward more empathetic communication. This approach allows various communities of practice to investigate how AI, on one side, and human communication and healthcare needs, on the other, can contribute to each other's development.

1 Introduction

Recent years brought both challenges and opportunities to interpersonal communication in all areas of life, especially healthcare. The COVID-19 pandemic, for instance, took an enormous toll on people's mental health. Effective empathic communication is now even more vital.

In healthcare, and Telemedicine (TM) in particular, expression of empathy is essential in building trust with patients. Yet, physicians' empathic communication in TM encounters has remained largely unexplored and not measured. Despite considerable research establishing the clinical efficacy of TM (e.g. in acute stroke care), there is limited research on how TM technology affects physician-patient communication (Cheshire et al., 2021). Research on how to decode human behaviors with respect to empathy expression, perception and action is still nascent (Xiao et al., 2012; Gibson et al., 2015; Alam et al., 2018; Pérez-Rosas et al., 2017; Buechel et al., 2018; Sedoc et al., 2020; Zhou and Jurgens, 2020; Hosseini and Caragea, 2021). Of all the components of professionalism, empathy may be the most challenging to communicate via TM given the physical separation of participants.

AI systems with simple, intuitive, flexible and efficient emotionally-intelligent interfaces to support empathic provider-patient communication during digital visits are urgently needed. With its current developments, AI can help us understand how to implement empathy and compassion in effective patient-provider interactions and guide training for medical personnel. In healthcare, AI initiatives must also be multidisciplinary, using/developing a variety of core sets of requirements and expertise and engaging many participants, e.g. AI designers, developers, health care leadership, frontline clinical teams, ethicists, humanists, patients and caregivers. Health care professional training programs should also incorporate core curricula that trains on using such AI tools appropriately (Matheny et al., 2019).

With this research, we aim to offer a solution to improve empathic patient-physician communication. Specifically, part of a larger inter-disciplinary initiative, we propose to develop a digital interface that integrates with various TM platforms to monitor the emotional state of providers/patients and to guide/train them on how to improve their expression of empathic communication. We use state-of-the-art multimodal Natural Language Processing (NLP) built on cognitive science communication theories (Cuff et al., 2016), operating as a plug-and-play across TM platforms for future scaling.

Our goals are to: (1) Design, build, and test an intelligent digital interface that guides clinicians toward more empathetic communication; (2) Develop a set of objective measures to assess the system's ability to positively impact clinicians' empathetic communication; and (3) Design a scalable plug and play architecture agnostic to TM platforms.

Beyond serving as a tool to improve empathic communication towards increased patient satisfaction, this project lays the groundwork for additional research in helping different professions work together effectively in the TM environment. We believe our research and investigation come at the

right time. Collaborative NLP + HCI developments have been largely unexplored (Blodgett et al., 2021), yet critical for the next-generation AI-driven immersive environments, especially in healthcare.

2 Methodology

Our NLP-driven Empathy system is a multitask multimodal (video, speech, text) machine learning (ML) framework to train a classifier to recognize empathetic language in patient-physician communication. It automatically labels dialogues with sentiment and emotions, recognizes different types of empathy (i.e., cognitive, affective, and prosocial behavior) (Cuff et al., 2016) at the utterance level, and computes an overall empathy score. Throughout the dialogue, when the empathy score falls below a critical level, the system automatically recommends the top three most plausible empathetic response suggestions (predicated on the sentiment and emotion labels, and the dialogue history).¹

With this research, we propose an intelligent interface system design, then present various ways to evaluate it along a number of relevant dimensions. We assume an ideal NLP system that operates at the human-level (i.e., gold standard).

Data. In the first phase of the project, we test the interface design on a dataset of six recorded doctor-patient interaction videos (three empathetic and three non-empathetic) collected from a healthcare training initiative² (Haglund et al., 2015). The dialogues are professionally designed simulations of five to seven minute interactions, where a doctor, breaking bad news, is expected to use layperson terms in a highly empathic language to console and guide the patient/family. The dataset was already analyzed and annotated for emotion and empathy content by trained third-party annotators (undergraduate Psychology and Social Work students at the University of Illinois trained in the SPIKES protocol (Baile et al., 2000)) using annotation guidelines consistent with established practices in NLP (Artstein and Poesio, 2008) and socio-behavioral research on empathy (Cuff et al., 2016).

In this step, we transcribed the interactions, converted the audio into .wav format, single-channel recordings, normalized the intensity by -3dB, using Audacity (AudacityTeam, 2017), and annotated

the audio files with Praat (Boersma and Weenink, 2021). The annotators marked utterance boundaries and segmented them by speaker turns, topic changes, and major syntactic boundaries (i.e., sentence and clause breaks) as needed. They, then completed utterance-level empathy annotation on the identified utterances, labeled each dialogue with emotion and sentiment, and suggested a set of three plausible empathetic responses at every time-stamp in the non-empathetic dialogue in need of empathetic intervention (guided by the positive interaction). This dataset/setting was used in developing the user interface and will be used for its evaluation.

3 Intelligent Empathic Interface Design

Given our task and data, we show the proposed intelligent interface with its five major functional regions in Figure 1. This is the doctor’s view.³

R1: Top left shows the basic function icons: account settings; stats; interface modality selection (video, audio, text). It also includes several standard icons to control the sound and video.

R2: Top center shows the account owner’s info (R2a): picture; basic credentials. Top right shows the other participant’s (interlocutor) info (R2b).

R3: In the center of the screen, there is the text dialogue. The default window is limited to two-turn history of the selected time-stamp; with option to see the entire raw/annotated transcript.

R4 and **R5**, in the bottom half of the user interface, give the audio and video streams, respectively. **R5** (Empathy statistics) visually shows measured patient’s distress and doctor’s empathetic score throughout the dialogue interaction. The user can stop, replay, and select various timestamps, etc. For a given time-stamp, a pop-up window suggests more empathetic responses.

Our long-term plan is to use the interface to evaluate the NLP system, for example to identify statistically significant acoustic differences between empathetic/non-empathetic speech; the extent to which emotion/empathy perception is encoded across modalities, etc.

4 Proposed Evaluation

It is important to recognize that effective skills for expressing empathy through TM differ from those used in in-person encounters. Virtual environments

¹The NLP system is currently under development.

²‘How should providers deliver bad news’ initiative: Duke Graduate Medical Center in collaboration with the Institute for Healthcare Improvement, and Open School.

³The patient and nurse pictures used were made available on Wikimedia Commons under the Creative Commons CC0 1.0 Universal Public Domain Dedication.

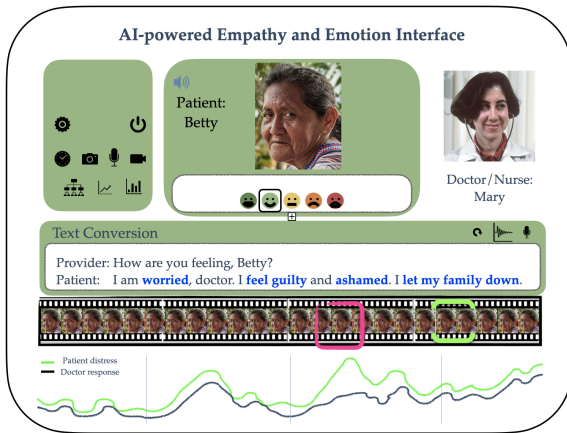


Figure 1: NLP-powered Empathic Interface

force healthcare professionals to adapt communication skills in a way that maintains professionalism and fosters the trust needed in medical care. In this study, we propose to investigate how TM can be used to assist rather than hinder patient-provider interactions, and to identify how the technology can support rather than diminish participants’ perceptions of expression of understanding, compassion and willingness to help. Delivering emotional and empathic suggestions visually as well as presenting them in an approachable way through a minimal interface is not a trivial task. The user must be able to relate to the interface and feel supported. Empathy, however, is a complex construct, its interpretation and significance being task-specific.

To address the challenges of the empathic construct, we start by focusing on specific tasks of empathic behaviors: breaking bad news to a patient. Our primary focus then becomes identifying elements of both affective and cognitive empathy, or perspective taking, in which one person attempts to view the scenario from another person’s perspective. We analyze empathy as it reflects in the verbal and nonverbal aspects of the conversation.

To investigate the interface efficacy as an assistive, communication mediation tool, we will conduct specific Human-Computer Interaction user studies of the intelligent interface. We propose to evaluate the interface along a number of dimensions with two third-party evaluators: (1) Healthcare App Developers; (2) Doctors and Patients. They are first briefed on the task and shown a demo of the interface. They will then watch the video interactions as they use the interface. Their findings will be recorded and transcribed, and their answers to all our dedicated questions and open ended ques-

tions (i.e., their concerns and suggestions) will be captured. We will then analyze their work and use it to better design and implement NLP-powered tools that can give both the doctor and the patient a frictionless and more accessible healthcare experience. Our focus is on making mainstream TM healthcare interfaces accessible and easy to use which, in turn, can lower development costs, increase availability, and lead to better tech acceptability (Agha et al., 2002; Annaswamy et al., 2020).

Paying attention to providers’ interactions with patients can encourage not only empathy but also the formation of professional identities that embody desirable values such as integrity and respect. Here, we want to build an AI communication mediation system that takes an experiential approach, putting experience and functionality on the same level. Besides ease of use, efficiency, and computational aspects, we also want to explore the *felt experience* and what really matters to human users and what it takes to make technology more meaningful. We intend to design a tool that does not only mediate communication, but also shapes experience. Most theoretical and practical HCI (Rubin and Chisnell, 2008) and NLP (Bird et al., 2008) systems and models focus primarily on quantitative metrics of evaluation. However, experience is subjective and dynamic, and thus, it emerges, shapes and reshapes through interactions with objects, people, environment and how these respond back to the experiencer (Hassenzahl, 2010). We believe that, besides required specific medical training, there is a need to create a space for clinicians to increase emotional awareness and discuss distressing aspects of their work.

A. Evaluation with Healthcare App Developers.

To examine our interface’s usability, we will seek feedback from healthcare app developers on ease of use (overcrowded interface; too many icons; functionality vs. aesthetics, etc.), user control/freedom, consistency, easiness in navigating/finding info, etc. The interface will give a wide range of feedback statistics during and after the dialogue interaction, including sentiment, emotion classes and intensity levels, and empathy scores. Data visualizations will then make it easy to analyze trend insights.

B. Evaluation with Doctors and Patients.

We plan to use a convenience sample of medical students/nurses (male/female) from a major US university (School of Nursing) and 25 trained and calibrated Standardized Participants (SPs) prepared for

the patient role.

We test five ways for physicians to foster empathy during interaction (i.e., ask participants to consider the doctor’s/patient’s point of view in the simulation, respectively): (1) recognize one’s own as well as other’s emotions, (2) address negative emotions over time, (3) attune to patients’ verbal/nonverbal emotional messages, and (3) be receptive to negative feedback. The participants also identify the use of relevant empathic language features in their evaluation, e.g. offer reassurance/support, express concern, repeat information, listen well, give enough time to the patient to process the news, and elicit open ended questions.

A final participant evaluation (a five-point Likert scale) captures the overall score of the patient’s perception of physician’s empathy during the visit (evaluator ‘as if’ the patient) and the overall score reflecting the observer’s evaluation of the intelligent interface. Once any of the three output dimensions of empathy drops beyond a threshold level, the system recommends an immediate action: (1) make the physician aware of their behavior and urge them to adjust (i.e., ‘be respectful’, ‘slow down’, ‘be more inclusive’, ‘be more friendly’, etc.); (2) make the physician aware of the patient’s behavior and urge them to respond compassionately (i.e., “calm them down, if angry”; “offer compassion, if anxious, sad”; “offer encouragement, if there is desire for positive change”, etc.).

5 Limitations and Potential Risks

A. Privacy, data concerns, accessibility, and personalization need to be addressed because AI models often rely on sensitive patient data to make decisions and predictions. Emotion AI models are increasingly better at understanding patient emotions, but expert human supervision is necessary, hence part of the interface design. Without earning users’ trust and confidence, AI for emotional support will not achieve its potential to help people. In our system, we will make it clear that we use aggregate, de-identified user data collected solely for research purposes (subjects decide to participate).

The pandemic drove people of all abilities to use digital products they never used, products where accessibility was often overlooked. Most TM platforms do not have custom features to ease healthcare communications (Annaswamy et al., 2020). Moreover, TM providers may not be able to understand/address the accessibility issues with their

patients even if the system was designed properly. Web accessibility standards also need to be adjusted to TM platforms (W3C, 2021). We plan to make our digital experience accessible, and also consider aspects that were less explored in TM.

Our system allows interface developers to customize the default visualizations/feedback to match the system’s aesthetics and goals. Customization should further be available to the end-user and meet her individual healthcare preferences and needs (i.e., privacy controls around revealing one’s abilities, security controls towards third-party devices combined with personal assistive technologies).

B. Limitations of TM Setting. The TM technology brings benefits to medical care but also adds limitations, as it changes the verbal/nonverbal doctor-patient communication, and mandates focused attention of doctor and patient. Unlike in traditional medical visits, where doctors/patients have physical proximity and communicate fully, with TM, non-verbal communication is limited and visual communication might be obstructed/distorted.

To counter this loss of patient-doctor information, both the doctor and the patient need to be intentionally focused. Doctors must address patients/family by name, nod, smile and provide auditory feedback to show they understand and empathize. Both doctors and patients must avoid disruptions outside the medical TM visual field.

However, even with the TM limitations, research so far found no reduction in patients’ perceived level of physician empathy (Nelson and Spaulding, 2005). In fact, in TM visits, with the doctor driving about two-thirds of the medical dialog (Ong et al., 1995), TM patients reported higher satisfaction. We argue that, in order to make up for the lack of non-verbal communication, in TM visits, doctors increase verbal communication, voicing agreement more and overall, providing more varied verbal feedback that improves the socio-emotional connection with patients. Even though more research is needed over a longer period of time, we believe there is a TM technology paradox: the limitations introduced by the TM technology (reduce communication - non-verbal) in fact force developing the very behaviors they were expected to hinder (increase communication - verbal).

C. Potential Risks in Emotional AI. AI is a necessary tool in TM solutions to assist with emotion detection. At the same time, it increases the risk for emotion mis-identification and, worse,

has the potential to generalize this across large groups of patients. For example, emotional AI can fail to capture how neurodivergent and neurotypical patients (Jurgens, 2020), or patients of various ethnicities/cultures, ages, genders express emotions, and thus easily mis-identify negative for positive emotions ((Rhue, 2018). The smile of a Japanese patient might be used to show respect or hide her true emotion, while for an American/Australian/Canadian patient it might be a sign of happiness. Some research found that, as compared to men, women not only seem to smile more but they might also do so on purpose, to diffuse a negative situation (LaFrance, 2002). Without intentional, situated research and implementation, the TM solution can easily stereotype some patients and miss-qualify the experience of others.

For successful TM, the emotional AI algorithms must account at least for cultural, age and gender differences in patient behaviors. They must also be able to identify extreme views, (e.g. racism, xenophobia, homophobia or ageism) that can lead to miss-interpreting doctor-patient communications in TM visits. This is possible only when intentionally hiring diverse teams to develop the TM solution, e.g. psychologists, ethicists, healthcare professionals and software engineers. Allowing for multi-modal inputs, e.g. not only facial recognition (of smiles) but also voice inflections, tone, or choice of words, is crucial to correctly identify emotions and avoid bias and stereotyping. Previous research has shown that multi-modal information, grace to complementarity benefits, is much more valuable than individual information – e.g. when used individually, accuracy in facial coding, biometrics, and electroencephalography (EEG) was 9% - 62%, but increased to 77%-84% when combined (Nielsen, 2016). In AI, multimodal emotion and empathy detection architectures are still in their infancy with their own challenges (for a survey, see (Zhao et al., 2021)). In our study, we intend to contribute to multi-model TM solution development.

6 Future Considerations

As healthcare technologies advance, NLP solutions also need to evolve to address the changing needs of TM providers, in particular, to improve the patient/family/caregiver - clinician communication with empathy and compassion. Our proposal to design and build an AI-powered interface to better guide/train medical professionals is timely.

With our reliably-evaluated interface, we can develop objective data-driven measures of empathy and foresee that they can leverage the promise of data analytics, thus shedding new light, from a novel quantitative perspective, on the construct of empathy (as a psychological and socio-behavioral phenomenon) and its indicators in linguistic behavior. These resources have the potential to present an entirely new framework to investigate, analyze, understand, and automatically detect empathy using advanced language processing technologies.

Our proposed emotionally-intelligent interface contributes to research on how to decode human behaviors with respect to empathy expression, perception and action. We combine computer science, engineering, language, medicine, human-centered design and education to extend our understanding of one another during the two-way audiovisual communication that has become ubiquitous in the lives of many patients seeking health care. Such a system is a novel knowledge-rich resource that could unlock new breakthroughs in our understanding of linguistic discourse-analytic and behavioral indicators of empathy to help shape communication training for physicians and others.

For future successful TM solutions, in our opinion, the following system, doctor and patient needs will drive continued development. First, we see a multi-country trend to develop healthcare systems that provide both traditional and modern medicine intentionally integrated (healthcare system need). For this, TM would greatly benefit from a multimodal and multisensory patient evaluation, the basis of traditional medical practices (Girju, 2021). Second, with increasing invested interest and medical knowledge, patients and their families want to be active co-contributors in the healthcare process (patient need). The future TM interface must welcome patients to share private pictures, videos, notes about their health journey. Third, doctors need future TM solutions to be best training tools not only to meet their varied individual learning styles (visual, auditory, kinesthetic) but also tools that they can use for self-training on demand (healthcare professional need). To meet all these needs, we believe only TM solutions with smart, immersive and empathic interfaces designed as interactive, adaptive environments that facilitate versatile multimodal and multisensory engagement for more efficient, aesthetic, memorable, and healing medical experiences will be successful.

References

- Zia Agha, Ralph M. Schapira, and Azmaira H. Maker. 2002. Cost effectiveness of telemedicine for the delivery of outpatient pulmonary care to a rural population. *Telemedicine Journal of eHealth*, 8(3):281–291.
- Firoj Alam, Morena Danieli, and Giuseppe Riccardi. 2018. Annotating and modeling empathy in spoken conversations. *Computer Speech & Language*, 50:40–61.
- Thiru M. Annaswamy, Monica Verduzco-Gutierrez, and Lex Frieden. 2020. Telemedicine barriers and challenges for persons with disabilities: COVID-19 and beyond. *Disability and health journal*, 13(4):100973.
- Ron Artstein and Massimo Poesio. 2008. Inter-Coder Agreement for Computational Linguistics. *Computational Linguistics*, 34(4):555–596.
- AudacityTeam. 2017. *Audacity(R): Free audio editor and recorder [computer application]*, volume 1. online, <https://audacityteam.org/>.
- Walter F. Baile, Robert Buckman, Renato Lenzi, Gary Gloger, Estela A. Beale, and Andrzej P. Kudelka. 2000. SPIKES - A Six-Step Protocol for Delivering Bad News: Application to the Patient with Cancer. *Oncologist*, 5(4).
- Steven Bird, Robert Dale, Bonnie J Dorr, Bryan Gibson, Mark Thomas Joseph, Min-Yen Kan, Dongwon Lee, Brett Powley, Dragomir R Radev, and Yee Fan Tan. 2008. *The ACL anthology reference corpus: A reference dataset for bibliographic research in computational linguistics*. European Language Resources Association - ELRA.
- Su Lin Blodgett, Michael Madaio, Brendan O’Connor, Hanna Wallach, and Qian Yang, editors. 2021. *Proceedings of the First Workshop on Bridging Human-Computer Interaction and Natural Language Processing*. Association for Computational Linguistics, Online.
- Paul Boersma and David Weenink. 2021. *Praat: doing phonetics by computer [computer program] version 6.1.50*, volume 1. online, <http://www.praat.org/>.
- Sven Buechel, Anneke Buffone, Barry Slaff, Lyle Ungar, and João Sedoc. 2018. Modeling empathy and distress in reaction to news stories. In *Conference on Empirical Methods in Natural Language Processing (EMNLP)*.
- William P Cheshire, Kevin M Barrett, Benjamin H Eidelman, Elizabeth A Mauricio, Josephine F Huang, William D Freeman, Maisha T Robinson, Gary R Salomon, Colleen T Ball, Dale M Gamble, Vickie S Melton, and James F Meschia. 2021. Patient perception of physician empathy in stroke telemedicine. *Journal of Telemedicine and Telecare*, 27(9):572–581.
- Benjamin MP Cuff, Sarah J Brown, Laura Taylor, and Douglas J Howat. 2016. Empathy: A review of the concept. *Emotion review*, 8(2):144–153.
- James Gibson, Nikolaos Malandrakis, Francisco Romero, David C Atkins, and Shrikanth S Narayanan. 2015. Predicting therapist empathy in motivational interviews using language features inspired by psycholinguistic norms. In *16th Conference of the International Speech Communication Association*.
- Roxana Girju. 2021. Adaptive multimodal and multi-sensory empathic technologies for enhanced human communication. In *Rethinking the Senses: A Workshop on Multisensory Embodied Experiences and Disability Interactions, the ACM CHI Conference on Human Factors in Computing Systems*. arXiv preprint arXiv:2110.15054.
- Michael M. Haglund, Mariah Rudd, Alisa Nagler, and Neil S. Prose. 2015. Difficult conversations: a national course for neurosurgery residents in physician-patient communication. *Journal of Surgical Education*, 72(3):394–401.
- Marc Hassenzahl. 2010. Experience design: Technology for all the right reasons. *Synthesis lectures on human-centered informatics*, 3(1):1–95.
- Mahshid Hosseini and Cornelia Caragea. 2021. Distilling knowledge for empathy detection. In *Findings of the Association for Computational Linguistics: EMNLP 2021*, pages 3713–3724, Punta Cana, Dominican Republic. Association for Computational Linguistics.
- Alan Jurgens. 2020. Neurodiversity in a neurotypical world: an enactive framework for investigating autism and social institutions. *Neurodiversity studies: A new critical paradigm*, pages 73–88.
- Marianne LaFrance. 2002. II. smile boycotts and other body politics. *Feminism & Psychology*, 12(3):319–323.
- Michael Matheny, Sonoo Thadaney Israni, Mahnoor Ahmed, and Danielle Whicher (editors). 2019. *Artificial Intelligence in Health Care: The Hope, the Hype, the Promise, the Peril*. National Academy of Medicine, Washington, DC.
- Eve-Lynn Nelson and Ryan Spaulding. 2005. Adapting the Roter interaction analysis system for telemedicine: lessons from four specialty clinics. *Journal of Telemedicine and Telecare*, 11(1_suppl):105–107.
- Nielsen. 2016. Nielsen consumer neuroscience unveils trailblazing ad testing solution. <https://www.prnewswire.com/news-releases/nielsen-consumer-neuroscience-unveils-trailblazing-ad-testing-solution-300283682.html>.
- Lucille ML Ong, Johanna CJM De Haes, Alaysia M Hoos, and Frits B Lammes. 1995. Doctor-patient communication: a review of the literature. *Social science & medicine*, 40(7):903–918.

- Verónica Pérez-Rosas, Rada Mihalcea, Kenneth Resnicow, Satinder Singh, and Lawrence An. 2017. Understanding and predicting empathic behavior in counseling therapy. In *The 55th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 1426–1435.
- Lauren Rhue. 2018. Racial influence on automated perceptions of emotions. Available at SSRN 3281765.
- Jeffrey Rubin and Dana Chisnell. 2008. *Handbook of usability testing: how to plan, design and conduct effective tests*. John Wiley & Sons.
- João Sedoc, Sven Buechel, Yehonathan Nachmany, Anke Buffone, and Lyle Ungar. 2020. Learning word ratings for empathy and distress from document-level user responses. In *Proceedings of the 12th Language Resources and Evaluation Conference*, pages 1664–1673, Marseille, France. European Language Resources Association.
- W3C. 2021. *W3C - The website of the world wide web consortium's web accessibility initiative*. online, <https://www.w3.org/WAI/>. www.w3.org.
- Bo Xiao, Dogan Can, Panayiotis G Georgiou, David Atkins, and Shrikanth S Narayanan. 2012. Analyzing the language of therapist empathy in motivational interview based psychotherapy. In *Proceedings of The 2012 Asia Pacific Signal and Information Processing Association Annual Summit and Conference*, pages 1–4. IEEE.
- Sicheng Zhao, Guoli Jia, Jufeng Yang, Guiguang Ding, and Kurt Keutzer. 2021. Emotion recognition from multiple modalities: Fundamentals and methodologies. *IEEE Signal Processing Magazine*, 38(6):59–73.
- Naitian Zhou and David Jurgens. 2020. Condolence and empathy in online communities. In *Proceedings of the 2020 Conference on Empirical Methods in Natural Language Processing (EMNLP)*, pages 609–626, Online. Association for Computational Linguistics.