1 Overview

In fields like healthcare and education, AI systems are increasingly expected to demonstrate empathy during interactions. Empathy in text-based communication helps build trust, improve user experience, and enhance system effectiveness. However, defining and measuring empathy in AI remains a challenge, making it difficult to evaluate systems claiming to be empathic.

Similar to how emotion recognition progressed—where AI systems learned patterns of emotional cues to eventually generate emotion-related content—detecting empathy patterns is the first step towards generating empathic responses in AI. Understanding empathy in text can lay the foundation for developing systems capable of creating meaningful empathic interactions in areas like healthcare and education.

The goal of this project is to develop an empathy detection system for text. This system will assess how empathic two given texts are, providing a reliable method to evaluate the empathic capabilities of AI systems. Several approaches to empathy detection will be explored, comparing different models to determine the most effective one.

A major challenge in the field is the lack of a universally accepted measure of empathy. This makes it difficult to verify claims of empathic behavior in AI systems. Developing an automatic empathy recognizer will address this gap, providing an objective way to evaluate AI-generated texts.

The empathy detection system will first be applied to score AI-generated text, offering a way to validate systems that claim to be empathic. This project will ultimately lead to a paper detailing the system design, model comparisons, and the overall findings, contributing to the development of empathic AI.

2 Reading List

To understand how empathy is defined, detected, and evaluated in AI, I have compiled a list of key papers. These readings will guide the design of the empathy detection system and help compare the project results with existing work. The list is flexible and may expand based on the evolving needs of the research.

1. **Challenges in Defining and Measuring Empathy:** This section covers theoretical papers on empathy as a concept, which is debated in both psychology and AI. AI cannot "feel" empathy but must simulate it. These papers will help define what we aim to simulate in AI systems.

    - Elliott, R., Bohart, A. C., Watson, J. C., & Greenberg, L. S. (2011). Empathy. Psychotherapy, 48(1), 43. https://doi.org/10.1037/a0022187
    - Cuff, B. M., Brown, S. J., Taylor, L., & Howat, D. J. (2016). Empathy: A review of the concept. Emotion review, 8(2), 144-153. https://doi.org/10.1177/175407391455846

2. **Empathy in AI and Dialogue Systems:** Papers in this section focus on AI's ability to simulate empathy in dialogue systems. Empathy in AI is functional and goal-oriented, improving user interaction. Studying current evaluation methods will help identify gaps in empathy detection metrics.

    - Wang, Y. H., Hsu, J. H., Wu, C. H., & Yang, T. H. (2021, January). Transformer-based empathetic response generation using dialogue situation and advanced-level definition of empathy. In 2021 12th International Symposium on Chinese Spoken Language Processing (ISCSLP) (pp. 1-5). IEEE.https://doi.org/10.1109/ISCSLP49672.2021.9362067
    - Concannon, S., & Tomalin, M. (2023). Measuring perceived empathy in dialogue systems. Ai & Society, 1-15. https://doi.org/10.1007/s00146-023-01715-z
    - Ma, Y., Nguyen, K. L., Xing, F. Z., & Cambria, E. (2020). A survey on empathetic dialogue systems. Information Fusion, 64, 50-70. https://doi.org/10.1016/j.inffus.2020.06.011
    - Maroengsit, W., Piyakulpinyo, T., Phonyiam, K., Pongnumkul, S., Chaovalit, P., & Theeramunkong, T. (2019, March). A survey on evaluation methods for chatbots. In Proceedings of

the 2019 7th International conference on information and education technology (pp. 111-119). https://doi.org/10.1145/3323771.3323824

  - Inkster, B., Sarda, S., & Subramanian, V. (2018). An empathy-driven, conversational artificial intelligence agent (Wysa) for digital mental well-being: real-world data evaluation mixed-methods study. JMIR mHealth and uHealth, 6(11), e12106. https://doi.org/10.2196/12106
  - Rashkin, H. (2018). Towards empathetic open-domain conversation models: A new benchmark and dataset. https://doi.org/10.48550/arXiv.1811.0020
  - Li, B., Wang, A., Strachan, P., Séguin, J. A., Lachgar, S., Schroeder, K. C., ... & Schaekermann, M. (2024, May). Conversational AI in health: Design considerations from a Wizard-of-Oz dermatology case study with users, clinicians and a medical LLM. In Extended Abstracts of the CHI Conference on Human Factors in Computing Systems (pp. 1-10). https://doi.org/10.1145/3613905.3651891

3. **Empathy Detection Models and Metrics:** I will study existing empathy detection systems to understand what models have been used and compare their success with my own. This will help benchmark my results and highlight strengths and weaknesses in previous work.

   - Shen, J., Kim, Y., Hulse, M., Zulfikar, W., Alghowinem, S., Breazeal, C., & Park, H. W. (2024). EmpathicStories++: A Multimodal Dataset for Empathy towards Personal Experiences. arXiv preprint arXiv:2405.15708.
   - Shen, J. (2023). Modeling empathic similarity in personal narratives (Doctoral dissertation, Massachusetts Institute of Technology). arXiv:2305.1424

4. **Empathic Datasets:** Since empathy-labeled datasets are scarce, I plan to investigate all available datasets, analyze their structure and tasks, and select one for training and evaluation. The dataset selection will be critical to the project's success. Please refer to Table1, where the names of the datasets are hyperlinked to the corresponding papers.

3 Datasets to Explore

The lack of empathy-labeled datasets presents a challenge in this project. I will explore various empathic datasets, focusing on those suited for text-based empathy detection tasks. Each dataset will be assessed for its structure, annotations, and the tasks it supports. Please check Table1. The dataset selection will involve:

- Reviewing research papers to understand dataset use.
- Analyzing the dataset's structure, including empathy annotations.
- Evaluating the tasks the dataset supports, such as empathy classification or rating.

After selecting a dataset, I will prepare it for model training and ensure it meets the requirements of the models I plan to develop. The quality and relevance of the dataset will directly affect the accuracy of the empathy detection models.

4 System Design and Methods

This project will design and compare different models for detecting empathy in text. The methods described are potential approaches that will be tested as the project progresses. The system design will focus on finding the most effective method for empathy detection.

- **Fine-Tuning Pre-Trained Models:** This approach involves fine-tuning models like BERT or RoBERTa to rate empathy between two texts. These models will serve as a baseline for comparison.

Dataset Name	Description	Tasks Supported	Access Status
Empathic Dialogues	Open-domain conversations focused on producing empathetic responses	Empathy classification, response generation	Open (downloaded)
Empathic Stories+++	Multimodal dataset with personal narratives, focusing on text modality	Empathy classification	Requested (accessed)
SocialIQA	Dataset for commonsense reasoning about social interactions	Social interaction QA, empathy detection	Open (downloaded)
EPITOME	Dataset on empathy in mental health support conversations	Empathy in mental health, classification	Partially Open (downloaded)
SEND	Emotional narratives dataset focusing on complex stories	Emotion recognition, empathy classification	Accessed (downloaded)
CANDOR	Multimodal dataset with naturalistic conversations (requested text only)	Empathy in conversation	Requested
OMG-Empathy	Multimodal empathy dataset used in affective behavior evaluation	Empathy prediction, storytelling	Requested (pending)
GoEmotions	Dataset of fine-grained emotions in text	Emotion classification, empathy tasks	Open (downloaded)
SEMAINE	Multimodal dataset of emotionally colored conversations	Emotion recognition, empathy detection	Requested (pending)
COMET	Commonsense transformers for knowledge graph construction	Knowledge graph generation, empathy-related tasks	Open (downloaded)
Empathic Reactions	Dataset focusing on empathy responses to text-based stimuli	Empathy reaction prediction	Open (downloaded)
MEDIC	Multimodal empathy dataset in counseling	Counseling empathy detection	Requested (access pending)
Empathic Intents	Dataset for identifying empathetic intentions in text	Empathy detection, intent classification	Open (downloaded)
EDOS	Large-scale dataset for empathetic response generation	Response generation	Open (downloaded)
Empathic Stories	Dataset modeling empathic similarity in personal narratives	Empathy similarity detection	Open (downloaded)
Empathic Conversations	Contextualized conversations with multiple empathy levels	Contextualized conversation analysis	Requested (email sent)

Table 1: Empathy-related datasets explored in this project

- **Zero-Shot/Few-Shot Learning with LLMs:** Large language models like GPT, Claude, and LLaMA will be explored for zero-shot or few-shot learning, potentially offering strong performance without extensive fine-tuning.
- **Neural Network on Extracted Features:** A shallow neural network trained on features like text embeddings or emotion-related word lists will allow experimentation with handcrafted features.
- **Joint Fine-Tuning Architecture:** An innovative approach will involve using a text encoder (e.g., BERT) in a joint architecture, where the model generates encodings that capture empathy-related traits for comparison.

The models will be trained and tested on the selected dataset. Evaluation metrics will include:

- **Performance:** Metrics such as accuracy, F1 score, and precision/recall will be used, depending on the task (e.g., classification or regression).
- **Empathy Detection Accuracy:** Each model's ability to detect empathy in text will be benchmarked.
- **Generalization:** The models' ability to generalize across different types of texts will be assessed to ensure robustness.

This project addresses the research gap in understanding empathic behavior in AI. By exploring existing research on empathy challenges, datasets, and evaluation methods, this project ensures that the developed methods are informed by the most relevant studies, contributing to the development of AI capable of detecting and eventually generating empathic responses.

5 Deliverables

1. **Final Paper:** By the end of the semester, I will deliver a paper documenting the project. The introduction and related work sections will be based on readings and research, while the methods, benchmarking, and results sections will detail the system design, model training, and evaluation. The paper will conclude with a discussion of the findings and lessons learned.
2. **Project's Data and Files:** The project's data and code will be made available in a GitHub repository. This will include the implemented models, preprocessing scripts, and instructions for reproducing the experiments.