

# ABLE: Personalized Disability Support with Politeness and Empathy Integration

Kshitij Mishra<sup>1\*</sup> Manisha Burja<sup>1\*</sup> Asif Ekbal<sup>2</sup>

<sup>1</sup>Department of Computer Science and Engineering, Indian Institute of Technology Patna, India

<sup>2</sup>School of AI and Data Science, Indian Institute of Technology Jodhpur, India

mishra.kshitij07@gmail.com, manishaboorja@gmail.com, asif.ekbal@gmail.com

## Abstract

In today’s dynamic world, providing inclusive and personalized support for individuals with physical disabilities is imperative. With diverse needs and preferences, tailored assistance according to user personas is crucial. In this paper, we introduce ABLE (Adaptive, Bespoke, Listen and Empathetic), a Conversational Support System for Physical Disabilities. By tracking user personas, including gender, age, and personality traits based on the OCEAN model, ABLE ensures that support interactions are uniquely tailored to each user’s characteristics and preferences. Moreover, integrating politeness and empathy levels in responses enhances user satisfaction and engagement, fostering a supportive and respectful environment. The development of ABLE involves compiling a comprehensive conversational dataset enriched with user profile annotations. Leveraging reinforcement learning techniques and diverse reward mechanisms, ABLE trains a model to generate responses aligned with individual user profiles while maintaining appropriate levels of politeness and empathy. Based on rigorous empirical analysis encompassing automatic and human evaluation metrics based on persona-consistency, politeness accuracy, empathy accuracy, perplexity, and conversation coherence, the efficacy of ABLE is assessed. Our findings underscore ABLE’s success in delivering tailored support to individuals grappling with physical disabilities. To the best of our knowledge, this is the very first attempt towards building a user’s persona-oriented physical disability support system <sup>1</sup>.

## 1 Introduction

Physical disabilities present significant challenges to individuals, affecting their daily activities and quality of life. According to the World Health Organization (WHO), over a billion people, approximately 15% of the global population, live

with some form of disability (Organization, 2021). Providing effective support for individuals with physical disabilities is crucial in enabling them to navigate their environment, engage in social interactions, and lead fulfilling lives. Conventional support systems (Johnson and Jacob, 2017) tried to address this issue but lacked in fulfilling the diverse needs of this population in facilitating independence, mobility, and access to different resources.

Personalization is the key in providing effective support for individuals with physical disabilities. Each user may have distinct characteristics, preferences, and requirements, necessitating tailored solutions (Cai et al., 2023). Research indicates that gender can influence the experience and perception of disability, with women often facing unique challenges related to societal expectations and access to healthcare (Matin et al., 2021). Individuals may exhibit variations in *gender*, *age*, and *personality traits*, such as those identified in the OCEAN model (Openness, Conscientiousness, Extraversion, Agreeableness, Neuroticism) (McCrae and Costa, 1992). Moreover, *meta-communicative* aspects, including politeness and empathy, are integral in fostering effective communication and rapport with users (Brown and Levinson, 1987). Tailoring support to individual needs and preferences enhances user satisfaction, engagement, and outcomes.

Despite the critical role of support systems, current solutions frequently fail to address the diverse needs of individuals with physical disabilities adequately. Many existing systems rely on standardized, generic responses, lacking the necessary personalization and adaptability to cater to individual user characteristics (Martinez-Cruz et al., 2020). Furthermore, the absence of politeness and empathy in these interactions often leads to impersonal and ineffective communication, ultimately hampering user engagement and satisfaction (Parchomiuk, 2019). The rapid advancements in technology un-

\*Equal contribution.

<sup>1</sup>Dataset and codes can be accessed at [EMNLP2024-ABLE](#)

underscore the urgent demand for more personalized and empathetic support solutions that are finely attuned to the distinct needs and preferences of users.

Focusing on these shortcomings, we propose **ABLE**: an Adaptive, Bespoke, Listen, and Empathetic conversational support system tailored specifically for individuals with physical disabilities, aiming to provide personalized assistance. We begin with the creation of a large-scale persona-tailored physical disability support conversational dataset, PERPDSCD, which encompasses various disabilities and support issues. Leveraging PERPDSCD, **ABLE** is developed in a reinforcement learning framework, where novel rewards are strategically designed to guide its learning process. These rewards guide **ABLE** to generate personalized responses that align with individual user profiles while incorporating politeness and empathy cues. The effectiveness of **ABLE** is assessed through rigorous automatic and human evaluation, focusing on measures, such as persona-consistency, gender-age consistency, politeness correctness, empathy correctness, linguistic fluency, and conversational coherence, to ensure its robustness and efficacy in providing tailored support. Our *key* contributions include:

1. Create a comprehensive conversational dataset, termed as PERPDSCD, encompassing various combinations of user personality traits, agent politeness, and empathy information. This dataset sets the groundwork for future advancements in physical disability support systems.
2. Introduce **ABLE** (Adaptive, Bespoke, Listen, and Empathetic), a physical disability support system prioritizing patient personality traits to tailor its responses with politeness and empathy to create a welcoming environment.
3. Design a novel reward function utilizing four reward models to ensure responses align with appropriate user persona-based politeness and empathy.
4. Through rigorous evaluation, we demonstrate the effectiveness of **ABLE** in providing personalized, polite, and empathetic support.

## 2 Related work

In the domain of physical disability support, historical developments lay the groundwork for understanding the current landscape. Initial attempts

focused on rudimentary assistive technologies and human-centered interventions (Johnson and Smith, 1998). In recent years, the importance of providing effective support for individuals with physical disabilities through conversational systems has been emphasized by numerous studies in healthcare (Preum et al., 2021). Several studies underscore the significance of facilitating conversations tailored to the specific needs and preferences of users with disabilities (Montenegro et al., 2019; Cha et al., 2021; Huq et al., 2022; Ha et al., 2023).

The shift towards personalized support systems for individuals with physical disabilities parallels advancements in healthcare and technology. With the advent of Artificial Intelligence (AI)-driven conversational agents, there is a growing recognition of the need for tailored assistance in this population (Wang and Li, 2018). The trajectory outlined by (Smith and Robinson, 1995) and (Alleman, 2002) in mental health counseling sets the stage for the application of personalized conversational agents in healthcare. As discussed by (Kocaballi et al., 2019), the systematic review sheds light on the potential of personalized systems to enhance patient outcomes and engagement. Personalization in physical disability conversations has been highlighted as a crucial factor for enhancing user engagement and satisfaction (Brown and Lee, 2018; Wang and Zhang, 2019).

Moreover, the incorporation of politeness and empathy in support interactions has been shown to improve user experience and foster a supportive environment significantly (Johnson and Adams, 2017; Lee and Tan, 2020). Recent studies have also addressed the importance of incorporating politeness and empathy in conversational systems. Techniques, such as reinforcement learning have been employed to adapt the politeness and empathy levels of system responses (Tan and Zhao, 2020; Huang and Liu, 2021; Mishra et al., 2022a; Samad et al., 2022; Mishra et al., 2022b, 2023b, 2024). However, these approaches have not been extensively applied in the context of physical disability support conversations.

Conversational systems for healthcare vary widely in applications, from behavior change interventions (Dennison et al., 2013), for chronic conditions (Schachner et al., 2020) to aiding cognitive disabilities (Huq et al., 2022). Personalized agents empower diverse population, from adolescents with Autism Spectrum Disorder (Cha et al., 2021) to

older adults promoting physical activity (Wiratunga et al., 2020). Despite advancements, challenges persist: lack of personalization, empathy, and reliance on rule-based models (Smith and Dragone, 2023; Miller and Lee, 2020; Wang and Zhang, 2022). Further, conversational systems explicitly tailored for supporting individuals with physical disabilities remain under-explored (Chen and Wang, 2020; Zhang and Liu, 2021). While recent research has explored the integration of personality traits, such as those defined by the OCEAN model, in conversational systems (Adams and Brown, 2019; Wang and Li, 2020; Mishra et al., 2023a), these works often focus on fixed personas, neglecting the variability and complexity of individual personalities exhibited during conversations (Miller and Wilson, 2021).

Therefore, to build a Support System for Physical Disabilities, we introduce **ABLE** which utilizes a comprehensive novel conversational dataset named PERPDSCD, comprising 18,026 dialogues enriched with user profiles and annotations, to offer tailored support interactions. Using reward different functions, ABLE generates responses aligned with individual user profiles while maintaining appropriate levels of politeness and empathy. Our approach represents a significant advancement in the field, addressing the shortcomings of existing systems and paving the way for further research in this important domain. To the best of our knowledge, PERPDSCD and ABLE constitute the first attempt to create a large-scale dataset and conversational system, specifically tailored for supporting individuals with physical disabilities.

### 3 Resource Creation

We create a large-scale physical disability support conversational dataset PERPDSCD consisting of personalized support conversations with the user’s gender, age, and persona. The PERPDSCD addresses a wide array of challenges related to physical disabilities, *viz.* *Mobility Aids, Home Modifications, Physical Therapy Exercises, Assistive Technology, Pain Management, Activities of Daily Living (ADLs), Emotional Support, Employment and Education, Social Interaction, Fitness and Recreation, Peer Support Groups, Parenting with Disabilities, and Transitions and Life Changes.* It delves into specific issues for these disabilities, such as Mobility Impairments, Visual Impairments, Hearing Impairments, Speech Impairments, Neu-

rological Disorders, Spinal Cord Injuries, Amputations, Orthopedic Disabilities, Cerebral Palsy, Muscular Dystrophy, Balance and Gait Disorders, Chronic Pain, Aging-Related Disabilities, and Mental Health Conditions. The details of each of the disabilities and respective issues can be found in Table 4 of the Appendix.

#### 3.1 Dataset Creation

The dataset consists of conversations between individuals with physical disabilities and a system acting as a doctor, created utilizing the capabilities of the GPT-3.5 (Ouyang et al., 2022) and human intervention. The objective is to create a diverse and realistic collection of dialogues capturing support dynamics across various situations.

- 1. Prompt Crafting:** The PERPDSCD dataset curation followed a structured approach by considering different disability types, gender (*male* and *female*), age (*younger, middle-aged, and older*), and persona. Additionally, various topics with associated physical disabilities were illustrated in the A.1.1 section of the Appendix for comprehensive representation. The prompt template facilitated the generation of multi-turn conversations between individuals with physical disabilities and the support doctor. Iterative feedback by domain experts refined the prompt template, enhancing its effectiveness. Key elements included introducing the patient’s problem, maintaining concise dialogue, and infusing responses with politeness and empathy.
- 2. Persona Variation:** The Five-Factor Model provides a robust framework for comprehending human personality, encompassing openness, conscientiousness, extraversion, agreeableness, and neuroticism (McCrae and Costa, 1992). These traits, ranging from high openness and conscientiousness to low extraversion and neuroticism, underscore the intricate complexity of human behavior and cognition. Through our analysis of common trait configurations observed in real-world population, we constructed 19 valid persona combinations (Costa and McCrae, 1991). This broad spectrum of persona combinations captures the varying trait intensities to portray individual complexity (McCrae and Costa, 1992, 2008). Validation by domain experts ensures the rel-

evance and accuracy of these persona combinations (McCrae et al., 2007). All of these persona combinations are detailed in Section A.1.2 of the Appendix.

### 3.1.1 Dialogue Generation

Initially, we crafted a prompt with specified traits to generate multi-turn conversations between individuals with physical disabilities and the doctor. Additionally, we integrated seed utterances drawn from real dialogues and consulting sources, such as the World Health Organization (WHO) to provide context and set the interaction’s tone. We recruited 10 human experts in post-graduate English Linguistics and relevant experience in physical disability tasks for creating these seed utterances based on specific guidelines:

1. Create seed utterances for all the combinations, encompassing gender, age, persona, issue, and type of physical disability.
2. Tailor the conversation flow as per user’s profile.
3. Formulate responses with varying politeness and empathy, catering to individual needs.
4. Craft responses aimed at providing the necessary support and instilling user confidence.

The prompts and seed utterances were fed into the GPT-3.5 model to generate dialogues. The number of turns in a conversation varied from 8 to 30 turns. After each dialogue generation, automated quality checks are also applied to ensure a high-quality dialogue: (i). Dialogue should start with USER only; (ii). Check blank spaces between conversations; (iii). Check the alternative USER-DOCTOR role switch in the generated dialogue:  $Speaker(t)! = Speaker(t - 1)$ ; (iv). No repetition of USER’s utterance or DOCTOR’s response, i.e. for a dialogue  $d = \{u_i, co_i, \dots, u_T, co_T\}$ , any  $u_i \neq u_k$  or  $co_i \neq co_k$ , where  $0 \leq i, k < T$ ; (v). DOCTOR’s responses should be in a conversational context. For this, we compute a BERT-F1-SCORE (Zhang et al., 2019) between the context  $c_i$  and DOCTOR’s response  $s_i$ , i.e.  $s = BS_{F1}(c_i, s_i)$ . For threshold values of  $s < 0.15$ , the generated  $s_i$  is taken as out-of-context.

Dialogues not satisfying any of the above conditions are generated again. The statistics of generated persona-oriented physical disability support conversational dataset (PERPDSCD) is shown in

Table 1. Prompt and example of seed utterances are detailed in Figure 2 and Table 8 of the appendix, respectively. A sample dialogue and complete conversation generated are shown in Figures 3 and 4 of the Appendix.

| Metrics                      | Train  | Val   | Test  |
|------------------------------|--------|-------|-------|
| # of Dialogues               | 14421  | 1803  | 1800  |
| # of Utterances              | 313495 | 49238 | 40353 |
| Min. Utterances per Dialogue | 10     | 12    | 10    |
| Avg. Utterances per Dialogue | 21.73  | 27.30 | 22.41 |
| Max. Utterances per Dialogue | 33     | 31    | 27    |

Table 1: Dataset statistics of PERPDSCD.

To ensure authenticity, accuracy, and relevance in conversations between the user and doctor we performed data quality control. This involves multiple phases, including manual checks, expert review, and continuous improvement measures. In the initial phase, same 10 human participants conducted manual checks, rating dialogues on a Likert scale of 1-5 to ensure *internal coherence*, *content consistency*, and *naturalness*. Guidelines to participants covered to rectify grammatical correctness, thematic consistency, language appropriateness, user profile consistency, and clinical sensitivity. In the subsequent phase, conversations scoring 1 were discarded, while those scoring 2 or 3 underwent modification, constituting approximately 5%, 12%, and 18% of the dataset, respectively. Corrections included rectifying errors, restructuring sentences, and ensuring grammatical coherence. In the final phase, expert reviews were conducted, with 5% of dialogues evaluated by medical health experts to ensure clinical accuracy and relevance. Their feedbacks contributed to refining and modifying the remaining 95% conversations. Due to space restrictions, guidelines are detailed in Section A.1.3 of the Appendix. The statistics of PERPDSCD concerning quality checks are shown in Table 5 of the Appendix.

### 3.2 Dataset Annotation

Annotations in our dataset PERPDSCD are carried out at the utterance level. During this process, annotations at the utterance level centered on classifying counselors’ responses based on (i) politeness: *polite*, *impolite*, *neutral*, and (ii) empathy levels: *empathetic*, *non-empathetic*, *neutral*. Due to space restrictions, annotation details are given in Section A.1.4 of the Appendix. Further, the statistics of PERPDSCD are shown in Table 6 of the Appendix.



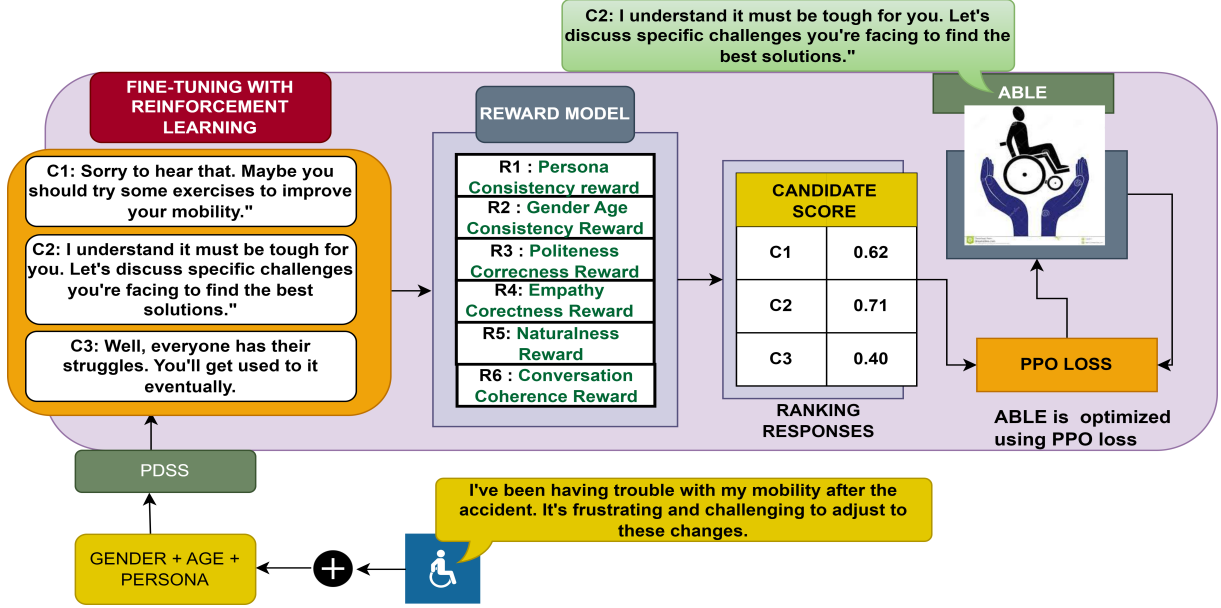


Figure 1: Overall architecture of the proposed system ABLE. First, we train a cross entropy loss-based PDSS model on PERPDSCD dataset. Then, it is fine-tuned by employing proximal policy optimization loss with six rewards to generate a user’s profile-oriented polite and empathetic response.

## 4 Methodology

We first warm-start by fine-tuning the Phi-2 (Li et al., 2023) model using the LORA (Hu et al., 2021) parameter efficient technique on PERPDSCD dataset, where PERPDSCD contains  $N$  conversations between a user (physically disabled) and a system (doctor). Each conversation includes information about the user’s gender, age, and persona. The model takes as input  $x_i$  the context, user’s persona, age, and gender, given as  $x_i = [c_i + p_i + g_i + a_i]$ , where  $c_i = [c_{i-1} + u_i]$ , and the output is  $y_i = s_i$ , where  $u_i$  and  $s_i$  are the user’s and system’s responses at the  $0 \leq i^{th} < T$  turn in the  $0 \leq d^{th} < N$  conversation.

$$PDS_{\theta} = \prod_{d=0}^N \prod_{i=0}^{i=T} \rho(y_i | x_i, x_{i-1}, \dots, x_0) \quad (1)$$

We aim to predict  $\hat{y}_i \approx y_i$ . The fine-tuning process involves minimizing the cross-entropy loss between the predicted and actual system responses:

$$\mathcal{L}_{CE} = -\frac{1}{N} \sum_{i=1}^N \sum_{j=1}^M y_{ij} \log(\hat{y}_{ij}) \quad (2)$$

where,  $M$  represents the vocabulary size, and  $\hat{y}_{ij}$  is the predicted probability of the  $j$ -th token in the vocabulary for the  $i$ -th conversation.

### 4.1 ABLE

In the second step, we further fine-tune the  $PDS_{\theta}$  in a reinforcement learning framework with the

Proximal Policy Optimization (PPO) loss (Schulman et al., 2017). We initialize the policy  $\pi_{\theta}(a_t | s_t) = PDS_{\theta}$  as the probability distribution over actions  $a_t$  given the state  $s_t$  under the current policy parameters  $\theta$ . In our context, an action  $a_t$  corresponds to selecting a response token from the vocabulary  $V$ . The state  $s_t$  at time step  $t$  is represented by the current conversation context and the model’s internal memory. Formally,  $s_t = [c_t, m_t]$ , where  $c_t$  is the conversation context and  $m_t$  is the model’s memory.

#### 4.1.1 Rewards

To guide the learning process, we design six novel rewards. These rewards ensure that the  $PDS_{\theta}$ ’s generated response ( $\hat{y}$ ) is natural and consistent with user persona, gender, and age with the incorporation of correct polite and empathy levels.

1. **Persona-Consistency Reward:** Encourages the model to generate responses consistent with the user’s persona information.

$$R_1 = \text{CLS}_{\text{per}^k}(y) - \alpha \text{CLS}_{\text{per}^k}(\hat{y}) \quad (3)$$

where  $\text{CLS}_{\text{per}^k}()$  computes the probability of  $0 \leq k^{th} < P$  persona class out of  $P$  classes.

2. **Gender-Age-Consistency Reward:** Promotes responses that are consistent with the user’s gender and age.

$$R_2 = \text{CLS}_{\text{ga}^k}(y) - \alpha \text{CLS}_{\text{ga}^k}(\hat{y}) \quad (4)$$

where  $\text{CLS}_{\text{ga}(\cdot)}$  computes the probability of  $0 \leq k^{\text{th}} < G$  gender-age class out of  $G$  classes.

- 3. Politeness Correctness Reward:** Rewards polite responses that adhere to predefined politeness criteria.

$$R_3 = \text{CLS}_{\text{pol}^k}(y) - \alpha \text{CLS}_{\text{pol}^k}(\hat{y}) \quad (5)$$

where  $\text{CLS}_{\text{pol}(\cdot)}$  computes the probability of  $0 \leq k^{\text{th}} < Q$  politeness class out of  $Q$  classes.

- 4. Empathy Correctness Reward:** Rewards empathetic responses that demonstrate understanding and empathy towards the user.

$$R_4 = \text{CLS}_{\text{emp}^k}(y) - \alpha \text{CLS}_{\text{emp}^k}(\hat{y}) \quad (6)$$

where  $\text{CLS}_{\text{emp}(\cdot)}$  computes the probability of  $0 \leq k^{\text{th}} < E$  empathy class out of  $E$  classes.

- 5. Naturalness Reward:** Encourages responses that are linguistically natural and fluent.

$$R_5 = \tanh(\text{Loss}(y, \hat{y})) \quad (7)$$

$\text{Loss}(y, \hat{y})$  gives the  $PDS_\theta$  loss in predicting  $\hat{y}$  for given  $y$ .

- 6. Conversation-Coherence Reward:** Promotes responses that maintain coherence and flow within the conversation.

$$R_6 = \beta \text{BS}_{F1}(y, \hat{y}_i) + \gamma \text{BS}_{F1}(c_i, \hat{y}_i) \quad (8)$$

$\text{BS}_{F1}(\text{zhang2019bertscore})$  gives the BERT-F1 score (Zhang et al., 2019).  $\beta, \gamma$  acts as weight we want to give to both the quantities where  $\beta + \gamma = 1$

In each of the rewards,  $\alpha = [1, 2]$  acts as a penalization factor. We define the overall reward  $R$  as the sum of all individual rewards weighted by their respective coefficients:

$$R = \sum_{i=1}^6 w_i \cdot R_i \quad (9)$$

where  $w_i$  are the weights corresponding to each reward  $R_i$ , where  $\sum w_i = 1$ . Then, the advantage function  $\hat{A}_t$  is computed using the rewards obtained from the environment.

$$\hat{A}_t = R_t - V(s_t) \quad (10)$$

where  $R_t$  is the total reward obtained at time step  $t$ , and  $V(s_t)$  is the state-value function representing the expected cumulative reward from state  $s_t$  onwards.

## 4.2 Policy Update with PPO Loss

The policy  $\pi_\theta$  is updated using the proximal policy optimization (PPO) loss function:

$$L^{PPO}(\theta) = -\mathbb{E}[\min(r(\theta)\hat{A}_t, \text{clip}(r(\theta), 1 - \epsilon, 1 + \epsilon)\hat{A}_t)] \quad (11)$$

where  $r(\theta)$  is the probability ratio,  $\hat{A}_t$  is the advantage function, and  $\epsilon$  is the clipping parameter. The parameters  $\theta$  of the policy  $\pi_\theta$  are updated using gradient descent with the modified PPO loss incorporating the reward:

$$\theta_{t+1} = \theta_t - \alpha \nabla_\theta L^{PPO}(\theta) \quad (12)$$

where  $\alpha$  is the learning rate.

## 5 Experiments

Due to space restrictions, implementation details of all the models are given in Section A.2 of the Appendix.

### 5.1 Evaluation Metrics

Both automatic and human evaluations are conducted to assess the performance of the proposed system **ABLE**.

**Automatic Evaluation Metrics:** We employ four metrics to evaluate persona accuracy (PCA), gender-age accuracy (GAA), politeness accuracy (PA), and empathy accuracy (EA). These metrics are defined as follows:

$$PCA = \mathbb{E}_{x_i, y_i} \mathbb{1}\{\text{CLS}_{\text{per}}(y_i) = \text{CLS}_{\text{per}}(\hat{y})\}, \quad (13)$$

$$GAA = \mathbb{E}_{x_i, y_i} \mathbb{1}\{\text{CLS}_{\text{ga}}(y_i) = \text{CLS}_{\text{ga}}(\hat{y})\}, \quad (14)$$

$$PA = \mathbb{E}_{x_i, y_i} \mathbb{1}\{\text{CLS}_{\text{pol}}(y_i) = \text{CLS}_{\text{pol}}(\hat{y})\}, \quad (15)$$

$$EA = \mathbb{E}_{x_i, y_i} \mathbb{1}\{\text{CLS}_{\text{emp}}(y_i) = \text{CLS}_{\text{emp}}(\hat{y})\}. \quad (16)$$

Additionally, we evaluate **ABLE** in terms of language and dialogue quality using three metrics: Perplexity ( $PPL$ ) (Brown et al., 1992), Response Length Ratio ( $R_{\text{len}}$ ), Non-repetitiveness ( $N_{\text{rep}}$ ).

$$PPL = \frac{\sum_r \exp(-\frac{1}{n} \sum_{i=1}^n \log P(y_i|x_i))}{r} \quad (17)$$

where  $n$  is the total number of tokens in the generated responses,  $r$  is the total number of the generated responses, and  $P(y_i|x_i)$  is the probability assigned by the language model to the  $i^{\text{th}}$  token given the input  $x_i$ .

$$R_{\text{len}} = \frac{\sum_r(n)}{r}. \quad (18)$$

$$N_{\text{rep}} = \frac{1}{2}(\text{BS}_{F1}(y_i, y_{i-1}) + \text{BS}_{F1}(y_i, y_{i-2})), \quad (19)$$

**Human Evaluation Metrics:** Human evaluation involves 10 evaluators, who were compensated according to the university norms. The evaluation consists of two phases: In the first phase, each evaluator interacts with **ABLE** five times, using different sets of utterances. They rate the conversations based on a Likert scale of 1-5 for seven metrics: persona accuracy, gender-age accuracy, politeness accuracy, empathy accuracy, fluency (FY), consistency (CY), and non-repetitiveness (NR). The scale denotes low-to-high intensity, e.g., a rating of 1 for persona accuracy indicates low consistency, while 5 denotes high consistency. These 50 evaluations are reviewed by medical experts. Based on the experts’ feedback, evaluators re-evaluate the initial 50 interactions. In the second phase, following obtained feedback, evaluators assess an additional 15 interactions each. This gave us a total of 200 evaluated interactions. Lastly, scores of each of the seven metrics are computed by taking the average of all 200 interactions.

## 5.2 Baselines

We compare our proposed **ABLE** with eight strong baselines *viz.* GPT2-large (Radford et al., 2019), ARDM (Wu et al., 2021), Llama2-7B (Touvron et al., 2023), Mistral-7B (Jiang et al., 2023), Zephyr-7B (Tunstall et al., 2023), Phi-1.5 (Li et al., 2023), PDSS:  $PDS_\theta$ , ABLE-R: ABLE with  $R = 0$ , ABLE-TR: ABLE with  $R = R_5 + R_6$ , and ABLE-GR: ABLE with  $R = R_1 + R_2 + R_3 + R_4$ .

## 6 Results and Analysis

**Automatic Evaluations:** Table 2 presents the results of automatic evaluation metrics for various physical disability support systems: GPT2-large, ARDM, Phi-1.5, Zephyr-7B, PDSS, ABLE-R, ABLE-TR, and ABLE-GR, to compare with our proposed model, ABLE. Significant differences were observed between **ABLE** and all other models ( $p < 0.05$ ). Among the compared models, ABLE consistently outperforms others across all the metrics.

In examining task-specific metrics:  $PCA$ ,  $GAA$ ,  $PA$ , and  $EA$ , a discernible pattern is seen i.e.  $GPT2\text{-large} < ARDM < Llama2\text{-7B} < Mistral\text{-7B} < Zephyr\text{-7B} < Phi\text{-1.5} < PDSS \approx ABLE\text{-R} < ABLE\text{-TR} < ABLE\text{-GR} < ABLE$ . Notably, PDSS and ABLE-R exhibit similar performance, attributed to ABLE’s initialization from  $PDS_\theta$ . It can be observed that LLAMA2-7B, Mistral-

7B, Zephyr-7B, and Phi-3 are outperformed by both ABLE-TR, and ABLE-GR which suggests that we do need RL to steer the model towards persona-consistent supportive dialogues. The better performance of ABLE-GR can be traced back to the influence of  $R_1$ ,  $R_2$ ,  $R_3$ , and  $R_4$ , underscoring the pivotal role of persona, gender, age, politeness, and empathy in guiding ABLE to formulate persona-consistent, polite, and compassionate responses. Moreover, Table 2 demonstrates that ABLE outperforms all eight baselines in terms of  $PPL$ ,  $R_{len}$ , and  $N_{rep}$ , following the same order as above:  $GPT2\text{-large} < ARDM < Zephyr\text{-7B} < Phi\text{-1.5} < PDSS \approx ABLE\text{-R} < ABLE\text{-TR} < ABLE\text{-GR} < ABLE$ . The better performance of ABLE-TR is attributed to  $R_5$  and  $R_6$ , which steer it towards more natural and contextually consistent responses. Language understanding and ability to generate coherent and contextually relevant responses.

ABLE’s success across all metrics can be attributed to its assimilation of patient profile information and adept adaptation of politeness and empathy levels. The integration of task-specific rewards aids ABLE in approximating a more precise distribution, further enhancing its competitive edge over the eight baselines. The inclusion of response-quality rewards fosters a dynamic rapport between the system and the user, enabling ABLE to focus on pertinent details and craft refined responses. This results in better language understanding ability to generate contextually relevant, diverse, and engaging responses. This underscores the dual necessity of all six rewards in yielding responses of elevated quality, validating our initial hypothesis. Generated responses of different models are illustrated in Figure 5.

**Human Evaluation:** Table 3 showcases human evaluation results for GPT2-large, ARDM, Zephyr-7B, Phi-1.5, PDSS, ABLE-R, ABLE-TR, and ABLE-GR, compared against ABLE. Similar to the automatic evaluation, ABLE outperforms all other models with respect to all the metrics:  $PCA$ ,  $GAA$ ,  $PA$ ,  $EA$ ,  $FY$ ,  $CY$ , and  $N_{rep}$ . A nuanced contrast emerges between PDSS and ABLE-TR, emphasizing the significance of task-specific rewards— $R_1$ ,  $R_2$ ,  $R_3$ , and  $R_4$ —in crafting persona-sensitive, polite, and empathetic responses. Notably, ABLE surpasses ABLE-TR and ABLE-GR, indicating the pivotal role of all six rewards in achieving fluent, consistent, non-repetitive, courte-

| Model                             | PCA          | GAA          | PA           | EA           | PPL         | $R_{len}$    | $N_{rep}$   |
|-----------------------------------|--------------|--------------|--------------|--------------|-------------|--------------|-------------|
| GPT2-large (Radford et al., 2019) | 50.3%        | 60.1%        | 72.8%        | 70.2%        | 14.93       | 11.19        | 0.39        |
| ARDM (Wu et al., 2021)            | 55.2%        | 67.9%        | 77.6%        | 75.6%        | 11.14       | 13.49        | 0.31        |
| Llama2-7B (Touvron et al., 2023)  | 54.7%        | 67.2%        | 78.6%        | 77.1%        | 7.01        | 16.94        | 0.22        |
| Mistral-7B (Jiang et al., 2023)   | 55.4%        | 68.3%        | 79.2%        | 78.4%        | 6.85        | 17.10        | 0.21        |
| Zephyr-7B (Tunstall et al., 2023) | 56.3%        | 69.6%        | 80.7%        | 78.9%        | 6.59        | 17.23        | 0.21        |
| Phi-1.5 (Li et al., 2023)         | 56.8%        | 70.1%        | 80.5%        | 78.7%        | 6.67        | 17.15        | 0.20        |
| PDSS                              | 58.0%        | 71.0%        | 83.7%        | 81.2%        | 5.01        | 18.31        | 0.15        |
| ABLE-R                            | 57.9%        | 71.3%        | 83.5%        | 81.6%        | 5.08        | 18.12        | 0.14        |
| ABLE-TR                           | 58.4%        | 71.9%        | 85.4%        | 83.0%        | 4.94        | 18.28        | 0.11        |
| ABLE-GR                           | 60.7%        | 73.1%        | 86.7%        | 84.2%        | 4.86        | 18.35        | 0.10        |
| <b>ABLE</b>                       | <b>61.5%</b> | <b>74.0%</b> | <b>87.6%</b> | <b>85.8%</b> | <b>4.30</b> | <b>19.95</b> | <b>0.07</b> |

Table 2: Results of automatic evaluation. Significant differences were observed between **ABLE** and all other models ( $p < 0.05$ ).

| Model       | PCA         | GAA         | PA          | EA          | FY          | CY          | $N_{rep}$   |
|-------------|-------------|-------------|-------------|-------------|-------------|-------------|-------------|
| GPT2-large  | 1.89        | 2.61        | 1.70        | 1.70        | 2.67        | 2.00        | 2.20        |
| ARDM        | 2.38        | 2.95        | 2.64        | 2.55        | 3.85        | 2.36        | 2.40        |
| Llama2-7B   | 2.66        | 2.98        | 3.26        | 3.44        | 4.01        | 3.25        | 2.48        |
| Mistral-7B  | 2.75        | 3.05        | 3.37        | 3.53        | 4.08        | 3.38        | 2.56        |
| Zephyr-7B   | 2.81        | 3.11        | 3.43        | 3.61        | 4.17        | 3.49        | 2.60        |
| Phi-1.5     | 2.79        | 3.15        | 4.43        | 3.88        | 4.27        | 3.70        | 2.80        |
| PDSS        | 3.06        | 3.74        | 4.53        | 4.06        | 4.09        | 3.80        | 3.00        |
| ABLE-R      | 3.02        | 3.70        | 4.63        | 4.16        | 4.00        | 4.13        | 3.40        |
| ABLE-TR     | 3.16        | 3.75        | 4.69        | 4.24        | 4.18        | 4.20        | 3.60        |
| ABLE-GR     | 3.29        | 3.82        | 4.81        | 4.36        | 4.27        | 4.32        | 3.80        |
| <b>ABLE</b> | <b>3.42</b> | <b>3.97</b> | <b>4.92</b> | <b>4.49</b> | <b>4.36</b> | <b>4.46</b> | <b>4.00</b> |

Table 3: Results of human evaluation

ous, and compassionate responses. These enhancements reflect ABLE’s ability to generate human-like and engaging conversations, thus boosting user satisfaction. The superior performance of ABLE is attributed to its reward-based architecture, optimizing response quality.

Both automatic and human evaluations validate ABLE’s efficacy in delivering high-quality conversational support to individuals with physical disabilities, suggesting its potential to significantly enhance user experience and overall well-being.

## 7 Error Analysis

While the results of our empirical analysis demonstrate the overall effectiveness of **ABLE** in delivering tailored support to individuals with physical disabilities, areas for improvement can be identified. One notable aspect of error stems from the misalignment between user personas and the generated responses. Despite our efforts to track user characteristics, there are instances where the generated responses do not fully align with the identified personas. This discrepancy may be attributed to the complexity of human personality traits and the inherent challenges in accurately capturing and representing them in the conversational dataset.

Additionally, we observed instances of sub-optimal politeness and empathy levels in certain

responses, which can lead to decreased user satisfaction and engagement. While ABLE integrates politeness and empathy levels into its response generation process, further refinement is needed to ensure consistency and appropriateness across all interactions. Furthermore, variations in conversation coherence were noted in some interactions, resulting in disjointed or fragmented dialogue flow. This may be attributed to limitations in the training data or deficiencies in the model’s ability to capture and maintain context over extended conversations.

## 8 Conclusion

In this paper, we introduce ABLE (Adaptive, Bespoke, Listen and Empathetic), a Conversational Support System tailored for individuals with physical disabilities. ABLE leverages user personas based on the OCEAN model to provide personalized assistance, integrating politeness and empathy to enhance user satisfaction. First, a physical disability support conversational dataset PERPDSCD with user profile annotations is curated. Then leveraging diverse rewards, ABLE effectively generates responses aligned with individual user profiles while maintaining appropriate levels of politeness and empathy. Through empirical analysis of the evaluation results, we demonstrate that ABLE’s efficacy in delivering tailored support for individuals



with physical disabilities. This study represents a significant step towards building user persona-oriented physical disability support systems and sets a foundation for further research in this domain. Future work could explore enhancements to ABLE's architecture, incorporate additional user profile factors, and extend its applicability to other domains beyond physical disabilities.

## Limitations

While ABLE demonstrates promising performance in providing tailored support to individuals with physical disabilities, it comes with some limitations. As it is trained using a large language model, Phi-2 (Li et al., 2023), it comes with its challenges, such as there could be cases where it may hallucinate. Hence, knowledge grounding is required for the responses with critical information. This constitutes our future direction for this work. Further, it is seen that continuous one-word or two-words user queries like 'yes', 'no', and 'is it?' may lead to out-of-context response generation. Despite efforts to integrate politeness and empathy levels into response generation, ABLE may occasionally produce responses that do not adequately reflect the desired level of politeness or empathy.

Variations in conversation coherence were observed in certain interactions, indicating room for improvement in maintaining context and coherence over extended dialogues. This could be addressed through more sophisticated dialogue management techniques and the incorporation of contextual information from previous turns. While ABLE has been evaluated on specific metrics such as persona-consistency, politeness accuracy, empathy accuracy, perplexity, and conversation coherence, there may be other important aspects of conversational quality that have not been fully explored. Future research could delve deeper into these aspects to provide a more comprehensive assessment of ABLE's performance.

Refinement of the model's language generation capabilities, particularly in understanding nuanced social cues, is necessary to enhance the quality of interactions. This could involve fine-tuning the model parameters or incorporating additional contextual cues to enhance the system's understanding of social dynamics and conversational norms. Addressing these issues could involve augmenting the training dataset with more diverse and context-

ually rich conversations or exploring advanced techniques for context-aware response generation. Lastly, while our study focuses on individuals with physical disabilities, it is essential to acknowledge the inherent biases and limitations in the dataset and model architecture. The system may not fully address the diverse needs and preferences of all users.

## Ethics Statement

Ethical considerations are critical in the development of conversational support systems like ABLE, especially when catering to vulnerable populations such as individuals with physical disabilities. Throughout the development process, ethical guidelines and principles were rigorously adhered to, with a focus on user privacy, autonomy, and well-being. Data privacy and security were prioritized to safeguard user information and ensure compliance with data protection regulations. Measures were implemented to anonymize and protect sensitive information. Additionally, efforts were made to mitigate potential biases in the model and dataset, ensuring fair and equitable treatment of persona combinations. Overall, ethical considerations were central to the development and deployment of ABLE, with a commitment to upholding the dignity, rights, and well-being of individuals with physical disabilities. The data had been approved by the Institute review board. The dataset would be made available only for research purposes with proper request.

## Acknowledgements

Kshitij Mishra acknowledges the prestigious Prime Minister's Research Fellowship (PMRF) by the Government of India for providing financial support during this research. We also thank our colleagues at the Department of Computer Science and Engineering, Indian Institute of Technology Patna, for their continuous support and valuable feedback.

## References

- S. Adams and D. Brown. 2019. Personality-aware conversational agents: A survey. *IEEE Transactions on Human-Machine Systems*.
- James R Alleman. 2002. Online counseling: The internet and mental health treatment. *Psychotherapy: Theory, Research, Practice, Training*, 39(2):199.

- G. Brown and H. Lee. 2018. Enhancing user engagement through personalized conversations. *Journal of Human-Computer Interaction*.
- Penelope Brown and Stephen C. Levinson. 1987. Politeness: Some universals in language usage. *Cambridge University Press*.
- Peter F Brown, Stephen A Della Pietra, Vincent J Della Pietra, Jennifer C Lai, and Robert L Mercer. 1992. An estimate of an upper bound for the entropy of english. *Computational Linguistics*, 18(1):31–40.
- Lin Cai, Jiaxin He, Yibo Wu, and Xuji Jia. 2023. The relationship between big five personality and quality of life of people with disabilities: The mediating effect of social support. *Frontiers in psychology*, 13:1061455.
- Inha Cha, Sung-In Kim, Hwajung Hong, Heejeong Yoo, and Youn-kyung Lim. 2021. Exploring the use of a voice-based conversational agent to empower adolescents with autism spectrum disorder. In *Proceedings of the 2021 CHI conference on human factors in computing systems*, pages 1–15.
- X. Chen and Y. Wang. 2020. Conversational ai for healthcare: A review. *IEEE Journal of Biomedical and Health Informatics*.
- Paul T. Jr. Costa and Robert R. McCrae. 1991. The neo personality inventory: Using the five-factor model in counseling. *Journal of Counseling & Development*, 69(4):367–372.
- Laura Dennison, Leanne Morrison, Gemma Conway, Lucy Yardley, et al. 2013. Opportunities and challenges for smartphone applications in supporting health behavior change: qualitative study. *Journal of medical Internet research*, 15(4):e2583.
- Sandeul Ha, Seung Hee Ho, Young-Hyeon Bae, Minyoung Lee, Ju Hee Kim, Ju Han Kim, and Jisan Lee. 2023. Digital health equity and tailored health care service for people with disability: User-centered design and usability study. *Journal of Medical Internet Research*, 25:e50029.
- Edward J Hu, Yelong Shen, Phillip Wallis, Zeyuan Allen-Zhu, Yuanzhi Li, Shean Wang, Lu Wang, and Weizhu Chen. 2021. Lora: Low-rank adaptation of large language models. *arXiv preprint arXiv:2106.09685*.
- L. Huang and Q. Liu. 2021. Empathy modeling in conversational systems using reinforcement learning. *Journal of Artificial Intelligence Research*.
- Syed Mahmudul Huq, Rytis Maskeliūnas, and Robertas Damaševičius. 2022. Dialogue agents for artificial intelligence-based conversational systems for cognitively disabled: A systematic review. *Disability and Rehabilitation: Assistive Technology*, pages 1–20.
- Albert Q Jiang, Alexandre Sablayrolles, Arthur Mensch, Chris Bamford, Devendra Singh Chaplot, Diego de las Casas, Florian Bressand, Gianna Lengyel, Guillaume Lample, Lucile Saulnier, et al. 2023. Mistral 7b. *arXiv preprint arXiv:2310.06825*.
- Kathryn Johnson and Justin Jacob. 2017. Support for individuals with disabilities: Overview and recommendations. *Journal of Disability Policy Studies*, 27(4):238–245.
- M. Johnson and R. Adams. 2017. The role of politeness and empathy in conversational systems. *IEEE Transactions on Affective Computing*.
- M. Johnson and R. Smith. 1998. Assistive technology for individuals with physical disabilities: A review. *Journal of Rehabilitation Engineering*, 5(3):123–135.
- Ahmet Baki Kocaballi, Shlomo Berkovsky, Juan C Quiroz, Liliana Laranjo, Huong Ly Tong, Dana Reza-zadegan, Agustina Briatore, and Enrico Coiera. 2019. The personalization of conversational agents in health care: systematic review. *Journal of medical Internet research*, 21(11):e15360.
- S. Lee and H. Tan. 2020. Fostering supportive environments through empathetic conversations. *Journal of Artificial Intelligence Research*.
- Yuanzhi Li, Sébastien Bubeck, Ronen Eldan, Allie Del Giorno, Suriya Gunasekar, and Yin Tat Lee. 2023. Textbooks are all you need ii: **phi-1.5** technical report. *arXiv preprint arXiv:2309.05463*.
- Yinhan Liu, Myle Ott, Naman Goyal, Jingfei Du, Mandar Joshi, Danqi Chen, Omer Levy, Mike Lewis, Luke Zettlemoyer, and Veselin Stoyanov. 2019. Roberta: A robustly optimized bert pretraining approach. *arXiv preprint arXiv:1907.11692*.
- Ilya Loshchilov and Frank Hutter. 2018. Decoupled weight decay regularization. In *International Conference on Learning Representations*.
- C. F. Martinez-Cruz, J. M. Diez-Anton, and J. J. Gomez-Sanz. 2020. User model for providing personalized recommendations in web accessibility evaluation. *Applied Sciences*, 10(11):3867.
- Behzad Karami Matin, Heather J Williamson, Ali Kazemi Karyani, Satar Rezaei, Moslem Soofi, and Shahin Soltani. 2021. Barriers in access to healthcare for women with disabilities: a systematic review in qualitative studies. *BMC women’s health*, 21:1–23.
- R. R. McCrae and P. T. Jr. Costa. 1992. Personality in adulthood: A five-factor theory perspective. *Guilford Press*.
- Robert R. McCrae and Paul T. Jr. Costa. 2008. Empirical and theoretical status of the five-factor model of personality traits. *The Sage Handbook of Personality Theory and Assessment*, 1:273–294.

- Robert R. McCrae, Antonio Terracciano, Anu Realo, and Juri Allik. 2007. On the validity of culture-level personality and stereotypes scores. *Journal of Personality*, 75(5):917–947.
- Mary L. McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica*, 22(3):276–282.
- J. Miller and T. Lee. 2020. Advances in conversational ai: A review. *AI Communications*.
- K. Miller and M. Wilson. 2021. Understanding variability in user personalities during conversations. *Journal of Conversational AI*.
- Kshitij Mishra, Mauajama Firdaus, and Asif Ekbal. 2022a. Please be polite: Towards building a politeness adaptive dialogue system for goal-oriented conversations. *Neurocomputing*, 494:242–254.
- Kshitij Mishra, Mauajama Firdaus, and Asif Ekbal. 2024. Please donate to save a life: Inducing politeness to handle resistance in persuasive dialogue agents. *IEEE/ACM Transactions on Audio, Speech, and Language Processing*.
- Kshitij Mishra, Priyanshu Priya, Manisha Burja, and Asif Ekbal. 2023a. e-therapist: I suggest you to cultivate a mindset of positivity and nurture uplifting thoughts. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing*, pages 13952–13967.
- Kshitij Mishra, Priyanshu Priya, and Asif Ekbal. 2023b. Pal to lend a helping hand: Towards building an emotion adaptive polite and empathetic counseling conversational agent. In *Proceedings of the 61st Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*, pages 12254–12271.
- Kshitij Mishra, Azlaan Mustafa Samad, Palak Totala, and Asif Ekbal. 2022b. Pepds: A polite and empathetic persuasive dialogue system for charity donation. In *Proceedings of the 29th International Conference on Computational Linguistics*, pages 424–440.
- Joao Luis Zeni Montenegro, Cristiano André da Costa, and Rodrigo da Rosa Righi. 2019. Survey of conversational agents in health. *Expert Systems with Applications*, 129:56–67.
- World Health Organization. 2021. [Disability and health](#). *World Health Organization*.
- Long Ouyang, Jeffrey Wu, Xu Jiang, Diogo Almeida, Carroll Wainwright, Pamela Mishkin, Chong Zhang, Sandhini Agarwal, Katarina Slama, Alex Ray, et al. 2022. Training language models to follow instructions with human feedback. *Advances in Neural Information Processing Systems*, 35:27730–27744.
- Monika Parchomiuk. 2019. Teacher empathy and attitudes towards individuals with disabilities. *International Journal of Disability, Development and Education*, 66(1):56–69.
- Sarah Masud Preum, Sirajum Munir, Meiyi Ma, Mohammad Samin Yasar, David J Stone, Ronald Williams, Homa Alemzadeh, and John A Stankovic. 2021. A review of cognitive assistants for health-care: Trends, prospects, and future directions. *ACM Computing Surveys (CSUR)*, 53(6):1–37.
- Alec Radford, Jeffrey Wu, Rewon Child, David Luan, Dario Amodei, Ilya Sutskever, et al. 2019. Language models are unsupervised multitask learners. *OpenAI blog*, 1(8):9.
- Azlaan Mustafa Samad, Kshitij Mishra, Mauajama Firdaus, and Asif Ekbal. 2022. Empathetic persuasion: reinforcing empathy and persuasiveness in dialogue systems. In *Findings of the Association for Computational Linguistics: NAACL 2022*, pages 844–856.
- Theresa Schachner, Roman Keller, and Florian v Wengenheim. 2020. Artificial intelligence-based conversational agents for chronic conditions: systematic literature review. *Journal of medical Internet research*, 22(9):e20701.
- John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. Proximal policy optimization algorithms. *arXiv preprint arXiv:1707.06347*.
- Howard B Smith and Gail P Robinson. 1995. Mental health counseling: Past, present, and future. *Journal of Counseling & Development*, 74(2):158–162.
- Ronnie Smith and Mauro Dragone. 2023. Generalisable dialogue-based approach for active learning of activities of daily living. *ACM Transactions on Interactive Intelligent Systems*, 13(3):1–37.
- W. Tan and Y. Zhao. 2020. Reinforcement learning for adaptive politeness in conversational systems. *Neural Computing and Applications*.
- Hugo Touvron, Louis Martin, Kevin Stone, Peter Albert, Amjad Almahairi, Yasmine Babaei, Nikolay Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruiti Bhosale, et al. 2023. Llama 2: Open foundation and fine-tuned chat models. *arXiv preprint arXiv:2307.09288*.
- Lewis Tunstall, Edward Beeching, Nathan Lambert, Nazneen Rajani, Kashif Rasul, Younes Belkada, Shengyi Huang, Leandro von Werra, Clémentine Fourrier, Nathan Habib, Nathan Sarrazin, Omar Sanseviero, Alexander M. Rush, and Thomas Wolf. 2023. [Zephyr: Direct distillation of lm alignment](#).
- H. Wang and G. Zhang. 2022. Neural approaches for conversational ai: Current trends and future prospects. *Journal of Artificial Intelligence Research*.
- J. Wang and S. Li. 2018. Personalized support systems for individuals with physical disabilities. *Journal of Assistive Technologies*, 10(2):79–91.
- J. Wang and L. Zhang. 2019. Personalization in conversational systems: A review. *ACM Transactions on Interactive Intelligent Systems*.

- Q. Wang and S. Li. 2020. Incorporating personality traits in conversational systems. *ACM Transactions on Interactive Intelligent Systems*.
- Nirmalie Wiratunga, Kay Cooper, Anjana Wijekoon, Chamath Palihawadana, Vanessa Mendham, Ehud Reiter, and Kyle Martin. 2020. Fitchat: conversational artificial intelligence interventions for encouraging physical activity in older adults. *arXiv preprint arXiv:2004.14067*.
- Qingyang Wu, Yichi Zhang, Yu Li, and Zhou Yu. 2021. Alternating recurrent dialog model with large-scale pre-trained language models. In *Proceedings of the 16th Conference of the European Chapter of the Association for Computational Linguistics: Main Volume*, pages 1292–1301.
- Q. Zhang and W. Liu. 2021. Recent advances in conversational systems: A survey. *Journal of Intelligent Information Systems*.
- Tianyi Zhang, Varsha Kishore, Felix Wu, Kilian Q Weinberger, and Yoav Artzi. 2019. Bertscore: Evaluating text generation with bert. *arXiv preprint arXiv:1904.09675*.



## A Appendix

### A.1 Dataset Details

#### A.1.1 Topics and Associated Physical Disabilities

1. **Mobility Aids:** Mobility aids, including wheelchairs, crutches, and prosthetics, offer vital support for diverse conditions like spinal cord injuries, amputations, and neurological disorders. Proper guidance ensures effective selection and maintenance, benefiting individuals with arthritis, degenerative conditions, and temporary injuries. These aids enhance independence and quality of life across various mobility challenges.
2. **Home Modifications:** Home modifications, like ramps and grab bars, enhance accessibility and comfort for individuals with mobility impairments, including wheelchair users and the elderly. Tailored adjustments address specific needs related to musculoskeletal and neurological disorders, sensory impairments, balance issues, amputations, and injuries. These modifications support rehabilitation, safety, and independent living, ensuring individuals can age in place with dignity and ease.
3. **Physical Therapy Exercises:** Physical therapy enhances mobility, strength, and flexibility for individuals with diverse disabilities. Tailored routines address musculoskeletal conditions, neurological disorders, and spinal cord injuries, preventing complications. They aid amputations, orthopedic injuries, cerebral palsy, and muscular dystrophy, promoting muscle tone, mobility, and balance. Post-surgical rehabilitation and overall fitness are also supported.
4. **Assistive Technology:** Assistive technology offers tools like speech recognition and screen readers, aiding communication and access to digital content for individuals with disabilities. It addresses speech impairments, visual and motor disabilities, and cognitive impairments, adapting to degenerative conditions and aiding in rehabilitation post-trauma. Aging adults benefit from its support for age-related impairments, fostering inclusivity and independence across diverse disability types.
5. **Pain Management:** Users receive guidance on pain management for physical disabilities, including medication options and relaxation techniques. Strategies address conditions like musculoskeletal issues, neurological disorders, and spinal cord injuries. Amputations may lead to phantom limb pain, while orthopedic injuries require postoperative care. The system offers holistic approaches, including medication management and stress reduction techniques, to alleviate chronic pain and discomfort.
6. **Activities of Daily Living (ADLs):** The system provides tailored strategies for ADLs, addressing mobility impairments, musculoskeletal conditions, and neurological disorders. It offers adaptive techniques for spinal cord injuries and amputations and temporary assistance during orthopedic injury recovery. Additionally, it aids older adults and individuals with cerebral palsy, balance/gait disorders, visual impairments, or hearing impairments in maintaining independence in daily activities.
7. **Emotional Support:** For those with physical disabilities, coping with emotional challenges is vital. The system offers guidance, coping strategies, and mental health resources. Individuals with mobility impairments, spinal cord injuries, musculoskeletal conditions, and neurological disorders may find support for navigating emotional adjustments and managing chronic pain. Amputations, traumatic injuries, and aging-related disabilities also benefit from emotional support, addressing issues like social isolation and caregiver stress.
8. **Employment and Education:** Guidance on opportunities and accommodations for mobility, visual, and hearing impairments. Neurological conditions may require flexible schedules, while spinal cord injuries and amputations need accessible transport and tools. Orthopedic disabilities benefit from ergonomic setups, and cerebral palsy, muscular dystrophy, and chronic illnesses may require specialized support. Learning disabilities need extended testing time, and mental health conditions necessitate holistic care.

9. **Social Interaction:** Provide tips for meaningful connections and overcoming barriers; mobility impairments address accessibility in venues and transportation. Visual impairments include communication techniques while hearing impairments need strategies for effective engagement. Neurological disorders, amputations, and orthopedic disabilities may require support in social contexts, along with cerebral palsy and muscular dystrophy. Chronic pain, aging-related disabilities, and mental health conditions receive integrated support for overall well-being.
10. **Fitness and Recreation:** The system recommends adaptive sports for various disabilities, like wheelchair basketball, goalball, and deaf volleyball. Activities include adaptive skiing, wheelchair rugby, and adapted dance. Adaptive yoga aids neurological disorders and balance issues, while gentle yoga helps manage chronic pain. Aging-related disabilities benefit from seated exercise programs.
11. **Peer Support Groups:** The system connects individuals with physical disabilities to peer support groups, fostering discussions on accessibility, adaptive living, and emotional well-being. Participants share experiences and advice on mobility aids, communication strategies, and coping mechanisms. Topics cover diverse conditions like spinal cord injuries, visual impairments, and chronic pain, offering insights into prosthetic options, symptom management, and lifestyle adjustments.
12. **Parenting with Disabilities:** The system supports parents with disabilities, offering adaptive tools and community guidance. Topics include mobility, vision, and hearing impairments, speech challenges, and neurological conditions like multiple sclerosis or cerebral palsy. Parents receive advice on safe environments, communication, and daily tasks. The system addresses spinal cord injuries, amputations, orthopedic disabilities, and chronic pain, ensuring effective caregiving despite disabilities.
13. **Transitions and Life Changes:** The system assists users in navigating life transitions, including moving to accessible homes, adapting to changes in disability status, and transitioning through various life stages. Tailored guidance is offered for mobility, vision, hearing, speech impairments, neurological disorders, spinal cord injuries, amputations, orthopedic disabilities, cerebral palsy, muscular dystrophy, balance, gait disorders, and chronic pain.

### **A.1.2 Persona Combination**

1. High Openness (O), High Conscientiousness (C), High Extraversion (E), High Agreeableness (A), Low Neuroticism (N): They are imaginative, organized, sociable, empathetic, emotionally stable, and resilient to stress. This person thrives in diverse settings and values creativity, structure, and positive relationships while staying composed and adaptable to challenges.
2. Low Openness (O), High Conscientiousness (C), High Extraversion (E), High Agreeableness (A), Low Neuroticism (N): This person is likely to be practical, organized, outgoing, empathetic, and emotionally stable. They value structure, enjoy social interactions, prioritize harmony in relationships, and handle stress effectively.
3. High Openness (O), Low Conscientiousness (C), High Extraversion (E), High Agreeableness (A), Low Neuroticism (N): This individual is characterized by a vivid imagination and a fondness for novel experiences (High O), combined with a laid-back and spontaneous approach to life (Low C). Their outgoing and compassionate nature (High E, High A) is complemented by emotional stability (Low N), contributing to a harmonious and socially engaging personality.
4. Low Openness (O), Low Conscientiousness (C), High Extraversion (E), High Agreeableness (A), Low Neuroticism (N): This individual tends to favor routine and tradition over novel experiences (Low O) and may display a relaxed and easygoing attitude towards responsibilities (Low C). Their sociable and amiable nature (High E, High A) is coupled with emotional resilience (Low N), contributing to a stable and affable personality.

| <b>Issues</b>                     | <b>Physical Disability</b>   |
|-----------------------------------|--|
| Home Modifications                | Mobility Impairments, Wheelchair Users, Limited Mobility Due to Age, Musculoskeletal Disorders, Neurological Disorders, Sensory Impairments, Balance and Gait Disorders, Amputations, Injuries and Accidents, Elderly Population.  |
| Physical Therapy Exercises        | Musculoskeletal Conditions, Neurological Disorders, Spinal Cord Injuries, Amputations, Orthopedic Injuries, Cerebral Palsy, Muscular Dystrophy, Balance and Gait Disorders, Aging-Related Mobility Issues, Post-Surgical Rehabilitation, General Physical Fitness.   |
| Assistive Technology              | Speech Impairments, Deafness or Hearing Impairments, Visual Impairments, Motor Disabilities, Cognitive Impairments, Multiple Disabilities, Degenerative Conditions, Traumatic Injuries, Aging-Related Disabilities, and Communication Disorders.   |
| Pain Management                   | Musculoskeletal Conditions, Neurological Disorders, Spinal Cord Injuries, Amputations, Orthopedic Injuries, Cerebral Palsy, Degenerative Conditions, Postural Issues, Complex Pain Syndromes, Aging-Related Issues, Medication Management, Relaxation and Stress Reduction.                                |
| Activities of Daily Living (ADLs) | Mobility Impairments, Musculoskeletal Conditions, Neurological Disorders, Spinal Cord Injuries, Amputations, Orthopedic Injuries, Aging-Related Mobility Issues, Cerebral Palsy, Muscular Dystrophy, Balance and Gait Disorders, Visual Impairments, Deafness or Hearing Impairments.                      |
| Emotional Support                 | Mobility Impairments, Spinal Cord Injuries, Musculoskeletal Conditions, Neurological Disorders, Amputations, Chronic Pain, Degenerative Conditions, Traumatic Injuries, Aging-Related Disabilities, Social Isolation, Caregiver Stress, General Mental Health.   |
| Employment and Education          | Mobility Impairments, Visual Impairments, Hearing Impairments, Communication Disorders, Neurological Disorders, Spinal Cord Injuries, Amputations, Orthopedic Disabilities, Cerebral Palsy, Muscular Dystrophy, Learning Disabilities, Chronic Illnesses, Mental Health Conditions.                        |
| Social Interaction                | Mobility Impairments, Visual Impairments, Hearing Impairments, Speech Impairments, Neurological Disorders, Spinal Cord Injuries, Amputations, Orthopedic Disabilities, Cerebral Palsy, Muscular Dystrophy, Chronic Pain, Aging-Related Disabilities, Mental Health Conditions.                             |
| Fitness and Recreation            | Mobility Impairments, Visual Impairments, Hearing Impairments, Upper Limb Amputations, Lower Limb Amputations, Orthopedic Disabilities, Spinal Cord Injuries, Cerebral Palsy, Muscular Dystrophy, Neurological Disorders, Balance and Gait Disorders, Chronic Pain, Aging-Related Disabilities.            |
| Peer Support Groups               | Mobility Impairments, Visual Impairments, Hearing Impairments, Speech Impairments, Neurological Disorders, Spinal Cord Injuries, Amputations, Orthopedic Disabilities, Cerebral Palsy, Muscular Dystrophy, Balance and Gait Disorders, Chronic Pain, Aging-Related Disabilities, Mental Health Conditions. |
| Parenting with Disabilities       | Mobility Impairments, Visual Impairments, Hearing Impairments, Speech Impairments, Neurological Disorders, Spinal Cord Injuries, Amputations, Orthopedic Disabilities, Cerebral Palsy, Muscular Dystrophy, Balance and Gait Disorders, Chronic Pain.   |
| Transitions and Life Changes      | Mobility Impairments, Visual Impairments, Hearing Impairments, Speech Impairments, Neurological Disorders, Spinal Cord Injuries, Amputations, Orthopedic Disabilities, Cerebral Palsy, Muscular Dystrophy, Balance and Gait Disorders, Chronic Pain.   |

Table 4: Topics and their respective physical disabilities of PERPDSCD dataset

5. High Openness (O), High Conscientiousness (C), Low Extraversion (E), High Agreeableness (A), Low Neuroticism (N): Marked by a curiosity for new ideas and a structured, goal-oriented approach to tasks (High O, High C), this individual tends to be reserved and introspective, leaning towards solitary activities (Low E). Their compassionate and cooperative demeanor (High A) aligns with emotional stability (Low N), forming a conscientious and empathetic personality.
6. Low Openness (O), High Conscientiousness (C), Low Extraversion (E), High Agreeableness (A), Low Neuroticism (N): This individual values routine and practicality (Low O, High C) and tends to be reserved, preferring quieter settings over social gatherings (Low E). Their agreeable and cooperative nature (High A) pairs with emotional stability (Low N), contributing to a dependable and calm personality.
7. Easily bored by routine, this individual thrives on creativity and exploration (High O, Low C), preferring solitary pursuits over social gatherings (Low E). Their compassionate and accommodating disposition (High A) and emotional resilience (Low N) foster a harmonious and introspective personality.
8. Low Openness (O), Low Conscientiousness (C), Low Extraversion (E), High Agreeableness (A),

Low Neuroticism (N): With a preference for familiarity and stability (Low O, Low C), this individual tends to be introverted and reserved (Low E), yet they possess a kind and accommodating nature (High A) alongside emotional resilience (Low N), fostering a gentle and steady personality.

9. High Openness (O), High Conscientiousness (C), High Extraversion (E), Low Agreeableness (A), Low Neuroticism (N): This individual is characterized by a love for new ideas and experiences (High O), combined with a strong work ethic and organizational skills (High C, High E). However, their assertive and independent nature (Low A) may lead to a more challenging interpersonal dynamic, complemented by emotional stability (Low N).
10. Low Openness (O), High Conscientiousness (C), High Extraversion (E), Low Agreeableness (A), Low Neuroticism (N): This individual leans towards practicality and tradition (Low O, High C), thriving in social situations with their outgoing and assertive nature (High E). However, their lower agreeableness (Low A) may indicate a more direct and assertive communication style, while emotional stability (Low N) contributes to a generally resilient demeanor.
11. High Openness (O), Low Conscientiousness (C), High Extraversion (E), Low Agreeableness (A), Low Neuroticism (N): This individual embraces novelty and creativity (High O) but may struggle with organization and follow-through (Low C), preferring lively social settings (High E) despite being less agreeable (Low A). Their emotional stability (Low N) suggests a resilient nature amidst challenges.
12. Low Openness (O), Low Conscientiousness (C), High Extraversion (E), Low Agreeableness (A), Low Neuroticism (N): This person enjoys socializing and seeks stimulation (High E) but may struggle with structure and planning (Low C), showing limited interest in exploring new ideas or experiences (Low O) and maintaining agreeable interactions (Low A). Their emotional stability (Low N) may contribute to a generally calm demeanor.
13. Low Conscientiousness (C), Low Extraversion (E), Low Agreeableness (A), Low Neuroticism (N), Low Openness(O): This individual may display a reserved and introverted demeanor (Low E) with a tendency to avoid conflict (Low A), yet they might lack structure and discipline in their approach to tasks (Low C). Their emotional stability (Low N) may contribute to a generally composed nature, though they may struggle with embracing new ideas or experiences (Low O).
14. High Conscientiousness (C), Low Extraversion (E), Low Agreeableness (A), Low Neuroticism (N), High Openness(O): This highly conscientious individual is organized and disciplined (High C) but tends to be reserved and introverted (Low E), potentially prioritizing independent pursuits over social interactions. Their openness to new ideas and experiences (High O) contrasts with lower agreeableness (Low A), and emotional stability (Low N) contributes to a composed and adaptable nature.
15. Low Extraversion (E), Low Agreeableness (A), High Conscientiousness (C), Low Openness(O), Low Neuroticism (N): This person tends to be introverted and reserved (Low E) with a preference for independence over socializing (Low A), demonstrating a strong sense of organization and reliability (High C). Their lower openness to new experiences (Low O) suggests a preference for familiarity, while their emotional stability (Low N) fosters a calm and composed demeanor.
16. High Neuroticism (N), High Conscientiousness (C), High Extraversion (E), High Agreeableness (A), High Openness(O): This person exhibits heightened emotional sensitivity and reactivity (High N) alongside a strong work ethic and organizational skills (High C). Their sociable and agreeable nature (High E, High A) complements a curiosity for new ideas and experiences (High O), creating a well-rounded and adaptable personality.
17. High Neuroticism (N), Low Conscientiousness (C), High Extraversion (E), High Agreeableness (A), Low Openness(O): This individual tends to experience heightened emotional volatility (High N) and



may struggle with organization and discipline (Low C), yet they possess a sociable and outgoing nature (High E) coupled with a compassionate and cooperative demeanor (High A). Their inclination towards familiarity over novelty (Low O) suggests a preference for routine and tradition.

18. High Neuroticism (N), High Conscientiousness (C), Low Extraversion (E), High Agreeableness (A), Low Openness(O): Emotionally sensitive yet reliably organized (High N, High C), this individual leans towards introspection over socializing (Low E), yet demonstrates warmth and cooperation (High A). Their preference for the familiar (Low O) underscores their stable and practical approach to life.
19. High Neuroticism (N), Low Conscientiousness (C), Low Extraversion (E), High Agreeableness (A), Low Openness(O): Inclined towards emotional sensitivity and occasional anxiety (High N), this person may struggle with structured routines (Low C) and prefers quieter settings (Low E). Yet, they radiate warmth and cooperation (High A), although they may shy away from novel experiences (Low O).

### A.1.3 Data Quality Control

We recruit 10 human participants to conduct manual checks to ensure the conversations' internal coherence, content consistency, and naturalness. The team rated the dialogues on a Likert scale from 1 to 5, adhering to predefined guidelines covering grammatical correctness, thematic consistency, language appropriateness, user profile consistency, and clinical sensitivity. After experts review of 5% of the data, they provided feedback in the form of guidelines as given below:

- Participants were tasked with identifying grammatical errors, subject-verb agreement issues, and improper word usage within the conversations.
- They checked for thematic coherence and logical flow, aiming to maintain consistency and avoid abrupt topic shifts or dialogue discontinuity.
- Evaluating the appropriateness of language used in the conversations, particularly in terms of natural tone and flow, formality, and cultural sensitivity, was emphasized.
- Ensure that user attributes and characteristics remain consistent throughout the conversation to maintain coherence and believability.
- Participants were requested to pay special attention to the portrayal of clinical interactions, aiming to enhance politeness and empathy.

Following these guidelines, participants cross-verified their given scores for the remaining 95% of the dataset and corrected them where necessary. Dialogues with updated scores of 1 were discarded, while those with scores of 2 and 3 underwent modifications similar to the first phase. An inter-evaluator Kappa agreement ratio of 80.3%, 81.2%, and 82.5% for internal coherence, content consistency, and naturalness, respectively, was observed among all participants.

| Statistics                              | PerPDSCD |
|---|----------|
| # of Conversations created              | 18974    |
| # of Conversations scored 1 (Discarded) | 948      |
| # of Conversations scored 2 (Modified)  | 2278     |
| # of Conversations scored 3 (Modified)  | 3415     |
| # of Conversations scored 4             | 5502     |
| # of Conversations scored 5             | 6831     |
| # of total conversations                | 18026    |

Table 5: Data quality control Statistics of PERPDSCD

### A.1.4 Dataset Annotation details

We engaged the same team of 10 participants as annotators. The annotation procedure is performed in two distinct phases. In the first phase, the team manually annotated 30% of the dataset, prioritizing the recognition of politeness and empathy labels. We provided illustrative examples for each level to ensure annotators shared a common understanding and could manually label the necessary politeness and empathy indicators. In the second phase, we adopted a streamlined method using two pre-trained RoBERTa large models. These models were fine-tuned to create classifiers specifically for identifying politeness and empathy labels. The process is as follows:

- Preparation: We trained two RoBERTa large models to recognize the politeness and empathy labels of the given utterances.
- Implementation: With the models ready, we applied them to the remaining 85% of the dataset.
- Prediction: Each utterance from the dataset was passed through the corresponding classifier. The classifiers then predicted whether the utterance displayed the appropriate label.
- Efficiency: By utilizing pre-trained models, we enhanced efficiency and scalability, reducing the manual effort in annotating the dataset while ensuring high accuracy across a large volume of data.

We were able to effectively annotate using these classifiers, making it scalable and accurate. After the automated annotation, we conducted a critical step: a second manual verification round by annotators to guarantee accuracy and dependability in the annotations. To assess consistency and reliability, we calculated multi-rater Kappa agreement (McHugh, 2012). In the first phase, agreement ratios of 82.7% and 80.8% for politeness and empathy respectively are observed. Whereas, in the second phase, 86.3% and 88.1% are found for politeness and empathy, respectively. We include a sample dialogue with example utterances showcasing various politeness and empathy labels, as referred to in Table 7.

| Metrics                           | Train  | Validation | Test  |
|-----------------------------------|--------|------------|-------|
| # of Utterances polite            | 213717 | 32004      | 25825 |
| # of Utterances impolite          | 43889  | 6893       | 5245  |
| # of Utterances neutral (polite)  | 56430  | 10339      | 8878  |
| # of Utterances empathetic        | 222581 | 35451      | 29457 |
| # of Utterances non-empathetic    | 37619  | 6400       | 4035  |
| # of Utterances neutral (empathy) | 53295  | 7386       | 6861  |

Table 6: Dataset annotation statistics of PERPDSCD.

| Annotation labels | Examples  |
|-------------------|---|
| Polite            | Thank you for sharing your concerns. Let’s work together to find the best solution.                     |
| Impolite          | I don’t have time for this. Just follow the instructions and you’ll be fine.                            |
| Neutral (Polite)  | I understand. Let’s explore different options and see what works best for you.                          |
| Empathetic        | I can only imagine how challenging this must be for you. I’m here to support you every step of the way. |
| Non-Empathetic    | You need to toughen up. Everyone has their struggles.   |
| Neutral (Empathy) | I see where you’re coming from. Let’s find a solution that suits your needs and comfort level.          |

Table 7: Example utterances of PERPDSCD with politeness and empathy labels

## A.2 Implementation Details

The fine-tuning process for all classifiers involves the utilization of the RoBERTa-large model Liu et al. (2019). Additionally, the language models, GPT2-large (Radford et al., 2019), ARDM (Wu et al., 2021), Zephyr-7B (Tunstall et al., 2023), Phi-1.5 (Li et al., 2023), PDSS are trained using a cross-entropy loss.

MULTI TURN CONVERSATION  
GENERATION PROMPT

Generate a conversation between a person with a physical disability and their physical disability support doctor/counselor. ⊗

Generate a multi-turn conversation between a person with a physical disability and their physical disability support doctor/counselor.

The physical disability person has below given traits: *Topic- {Topic}, disability- {disability}, gender- {gender}, age- {age}, persona combination- {personatrait}*

Very important points to consider are:

0. Make sure the patient's first utterance starts by greeting the doctor and saying his problem.

1. Generated conversations must follow alternative interplay between Patient and Therapist.

2. Keep responses short, 90 words at most.

3. Make sure there are at least 20 turns in the conversation.

4. please don't repeat the conversation; make it unique and follow the context throughout.

5. Make sure the given patient traits and all the given combination of persona traits are reflected in the generated conversation ✔

Figure 2: Prompt for multi-turn dialogue generation.

|   | Seed Utterances  |
|---|--|
| Topic: Physical therapy exercises<br>Disability: Amputations<br>Gender: Male<br>Age: Middle-Aged<br>Persona: Low Openness (O), Low Conscientiousness (C), Low Extraversion (E), High Agreeableness (A), Low Neuroticism (N)     | Patient: Hi, Doctor. I hope you're doing well. I've been struggling with my physical therapy exercises after the amputation. It's challenging, and I'm not sure how to stay motivated.<br>Doctor: Hello there. Thank you for reaching out. It's common to feel overwhelmed with physical therapy, especially after such a significant change. Let's work together to find strategies to make it more manageable. What specific difficulties are you facing?<br>Patient: Honestly, I find it hard to stick to the exercises. It's like I lose interest or forget about them altogether.   |
| Topic: Social interaction<br>Disability: Mobility Impairments<br>Gender: Female<br>Age: Older<br>Persona: High Openness (O), Low Conscientiousness (C), High Extraversion (E), High Agreeableness (A), Low Neuroticism (N)      | Patient: Hello, Doctor. How are you today? I've been feeling a bit isolated lately due to my mobility impairment. Social interactions seem more challenging than ever.<br>Doctor: Hi there. I'm doing well, thank you. I'm sorry to hear you're feeling isolated. It's understandable given the circumstances. Let's explore ways to improve your social interactions. What difficulties are you experiencing specifically?<br>Patient: I feel like I'm missing out on social events and gatherings because of my mobility issues. It's frustrating not being able to participate fully. |
| Topic: Activities of Daily Living<br>Disability: Visual Impairments<br>Gender: Male<br>Age: Younger<br>Persona: High Openness (O), Low Conscientiousness (C), High Extraversion (E), Low Agreeableness (A), Low Neuroticism (N) | Patient: Hi, Doctor. I hope you're well. I've been struggling with my daily activities since my visual impairment. It's been tough, and I could use some guidance.<br>Doctor: Hello! I'm here to help. It's understandable to face challenges with daily activities after a visual impairment. Let's discuss what specific tasks you find difficult and explore solutions together.<br>Patient: I find it hard to navigate around my house and perform tasks like cooking and cleaning. It's frustrating, and I feel like I'm constantly dependent on others.                            |

Table 8: Example seed utterances of PERPDSCD

For **ABLE**, training is conducted with  $batch\_size = 8$ ,  $seed\_value = 10$ ,  $human\_reward = 10$ ,  $max\_candidate\_length = 50$ ,  $clip\_ratio = 0.2$ ,  $discount\_factor = 0.95$ ,  $number\_of\_steps = 32000$ ,  $steps\_per\_update = 640$  and AdamW optimizer Loshchilov and Hutter (2018) with a learning rate of  $\alpha = 1e^{-05}$ ,  $\epsilon = 0.2$  and  $epochs = 20$ .

### A.2.1 Hardware Configuration

The experimental setup encompasses the subsequent device configurations:

1. GPU: A100-PCIE-40GB
2. CUDA Support: CUDA 11.x (or later)

**DATA PREPARATION -  
A SAMPLE DIALOGUE**

**PROMPT**

Generate a multi-turn conversation between a person with a physical disability and their physical disability support doctor/counselor.

The physical disability person has below given traits:

**Topic** - {Peer Support Groups},

**Disability** - {Amputations},

**Gender** - {Male}

**Age** - {Older},

**Persona combination** - {High Openness (O), Low Conscientiousness (C), High Extraversion (E), High Agreeableness (A), Low Neuroticism (N)}

Very important points to consider are:

0. Make sure the patient's first utterance starts by greeting the doctor and saying his problem.
1. Generated conversations must follow alternative interplay between Patient and Therapist.
2. Keep responses short, 90 words at most.
3. Make sure there are at least 20 turns in the conversation.
4. please don't repeat the conversation; make it unique and follow the context throughout.
5. Make sure the given patient traits and all the given combination of persona traits are reflected in the generated conversation

**SEED UTTERANCES**

**Patient:** Hello, Doctor. I hope you're well. Lately, I've been struggling with adjusting to life after my amputation. The changes are overwhelming, and I thought talking to someone might help.

**Doctor:** Hello! I appreciate you reaching out. It's completely normal to feel overwhelmed after such a significant change. Let's talk about it. How have you been coping with the adjustments?

**Patient:** Honestly, it's been tough. I'm an older guy, and the peer support groups I've tried don't seem to click with me. I feel a bit isolated.

**GPT 3.5**

**GENERATED SAMPLE DIALOGUE**

**Doctor:** I understand, and I appreciate your honesty. Adjusting to new circumstances takes time. Let's explore different support groups that align better with your interests. What activities or hobbies do you enjoy?

**Patient:** I used to love woodworking before the amputation. But now, it feels like I've lost that part of me.

**Doctor:** Woodworking is a fantastic skill! Perhaps we can find a support group focusing on adaptive techniques for hobbies like woodworking. Would that interest you?

**Patient:** That sounds intriguing. I'd love to explore that. It might help me reconnect with a part of my life I had lost.

Figure 3: A sample dialogue generation of our dataset **PerPDSCD** using our method.



## SAMPLE CONVERSATION

### PATIENT PROFILE

**Topic:** Transitions and life changes

**Disability:** Orthopedic Disabilities

**Gender:** Female

**Age:** Elder

**Persona:** High Neuroticism (N), Low Conscientiousness (C), High Extraversion (E), High Agreeableness (A), Low Openness(O)

### CONVERSATION

**Patient:** Hello, doctor. I hope you're doing well. I wanted to discuss an issue I've been facing lately.

**Doctor:** Hello, Mrs. Johnson. Of course, I'm here to help. What seems to be the problem? [Polite] [Empathetic]

**Patient:** Well, I've been struggling with transitions and life changes. It's been quite overwhelming for me.

**Doctor:** I understand. Transitions can be challenging, especially for individuals with orthopedic disabilities. Can you tell me more about what you're experiencing? [Polite] [Empathetic]

**Patient:** Whenever there's a change in my routine or environment, I become extremely anxious and find it difficult to adapt. It's affecting my daily life.

**Doctor:** I see. It's important to address these concerns. We can work together to develop strategies to help you cope with transitions more effectively. Have you tried any techniques in the past? [Polite] [Empathetic]

**Patient:** I've tried deep breathing exercises, but they haven't been very effective. I feel like I need more guidance.

**Doctor:** That's understandable. Deep breathing can be helpful, but it might not be enough for everyone. Let's explore other techniques that might suit you better. How about creating a structured schedule to ease the transition process? [Polite] [Neutral]

**Patient:** That sounds like a good idea. Having a clear plan might reduce my anxiety. I'll give it a try.

**Doctor:** Great! We can also work on setting realistic goals and breaking them down into smaller, manageable steps. This way, you can gradually adapt to changes without feeling overwhelmed. [Polite] [Empathetic]

**Patient:** I like the idea of breaking things down. It seems less daunting that way. Thank you for suggesting it

**Doctor:** You're welcome, Mrs. Johnson. Remember, I'm here to support you throughout this process. Is there anything else you'd like to discuss? [Polite] [Empathetic]

**Patient:** Actually, I've been feeling quite isolated lately. I find it hard to socialize due to my disability, and it's been affecting my mood.

**Doctor:** I understand how important social connections are for overall well-being. Let's explore ways to help you engage with others and find a supportive community. Have you considered joining any local disability support groups? [Polite] [Empathetic]

**Patient:** I haven't thought about it, but it might be worth a try. I'll look into it and see if there's any group that suits my interests.

**Doctor:** That's a proactive approach, Mrs. Johnson. Being part of a support group can provide you with a sense of belonging and understanding. It might also help you navigate through life changes with the support of others who have similar experiences. [Polite] [Empathetic]

**Patient:** I can see the benefits of joining a support group. I'll definitely give it a shot. Thank you for the suggestion, doctor.

**Doctor:** You're welcome. Remember, taking small steps towards positive changes can make a big difference. Is there anything else you'd like to discuss today? [Neutral] [Neutral]

**Patient:** Not at the moment, doctor. I appreciate your time and guidance. It's reassuring to know that I have your support.

**Doctor:** I'm glad to hear that, Mrs. Johnson. Remember, I'm just a phone call away if you need any assistance or have any questions. Take care and stay positive. [Polite] [Neutral]

Figure 4: A sample conversation of our dataset PerPDSCD.

| PROMPT       | Ever since my injury, coping with transitions and life changes has been overwhelming. It feels like I've lost control, and I'm not sure how to navigate these new challenges.  |
|--------------|--|
| GPT2 - LARGE | It sounds like you're going through a tough time. Remember, change can be difficult, but you can adapt to your new circumstances.  |
| ARDM         | I'm sorry to hear about the challenges you're facing. It's natural to feel overwhelmed after such a significant change. Let's explore some coping strategies together.   |
| ZYPHER - 7B  | Coping with transitions can indeed be overwhelming, especially after experiencing an injury. It's important to take things one step at a time. How about we start by identifying some small goals to work towards?   |
| PHI 1.5      | I can imagine how difficult it must be to adjust to these new challenges. It's important to remember that change takes time, and it's okay to seek support along the way. Let's figure out some strategies that can help you navigate through this period and lets start with small steps.   |
| ABLE         | I truly hear you. Coping with transitions after an injury can indeed be incredibly challenging. It's vital to acknowledge the depth of your emotions and to know that seeking support when needed is not just okay, it's courageous. Together, let's explore some strategies to help you regain not just stability, but also a sense of control and empowerment in your life. Your journey matters greatly to me, and I'm here to support you every step of the way. |

Figure 5: Generated text for the given prompt using different models.

3. GPU Clocks: Base: 765 MHz, Boost: 1410 MHz
4. Memory Size: 40 GB
5. Memory Type: HBM2
6. Memory Clock: 1215 MHz
7. Bus Width: 5120 bits
8. Total Board Power: 250 W.

### A.3 GPT-3.5 Results and Analysis

In the **zero-shot setting**, GPT-3.5 achieves a **Persona Consistency Accuracy (PCA)** of 42.8%, indicating that persona consistency is a notable challenge for the model. Similarly, **Gender-Age Accuracy (GAA)** stands at 46.5%, reflecting moderate recognition of gender and age-related preferences in responses. On the other hand, **Politeness Accuracy (PA)** and **Empathy Accuracy (EA)** are relatively higher, reaching 74.9% and 73.4%, respectively, which indicates the model's stronger performance in producing polite and empathetic dialogues. The **response length (R-len)** of 16.23 and **non-repetitiveness (N-Rep)** of 0.16

| <b>Metric</b> | <b>Zero-Shot</b> | <b>Few-Shot</b> |
|---------------|------------------|-----------------|
| PCA (%)       | 42.8             | 49.5            |
| GAA (%)       | 46.5             | 54.3            |
| PA (%)        | 74.9             | 79.2            |
| EA (%)        | 73.4             | 78.1            |
| R-len         | 16.23            | 17.19           |
| N-Rep         | 0.16             | 0.13            |

Table 9: Zero-Shot and Few-Shot Results with GPT-3.5.

suggest that while the responses are of adequate length, the model still generates a noticeable level of repetition in its responses.

In the **few-shot setting**, GPT-3.5 demonstrates improved performance across all metrics. **PCA** increases to 49.5% and **GAA** rises to 54.3%, indicating that the model benefits significantly from the few-shot learning paradigm, leading to better persona consistency and gender-age adaptation. The **PA** increases to 79.2% and **EA** to 78.1%, showing further improvements in generating polite and empathetic responses when the model is provided with a few examples. The **response length (R-len)** increases slightly to 17.19, and the **N-Rep** decreases to 0.13, suggesting better fluency and reduced repetitiveness in the generated outputs.

Despite these improvements, certain limitations remain in GPT-3.5’s ability to handle complex conversational dynamics, particularly in disability-specific dialogues where the variations between different disability types pose challenges. The model struggles to adapt to the nuanced nature of disability-related conversations, resulting in lower persona consistency (PCA). Additionally, the model frequently exhibits confusion when handling multiple personas within the same conversation, leading it to deviate from the intended context by focusing too much on a single persona and neglecting the broader conversational flow.

#### A.4 ABLE’s Bias Check

To ensure the absence of bias in **ABLE**’s responses, we conducted further human evaluations. This section outlines the experimental procedure, followed by a detailed analysis of the obtained results.

##### A.4.1 Experimental Procedure

We engaged 15 new human evaluators to interact with **ABLE**, with each evaluator conducting 10 interactions with the system. These interactions were designed to assess the system’s responses for any signs of bias. To ensure this, we divided evaluators into two separate sets. In the first set, eight evaluators were employed, while in the second set, seven evaluators were utilized.

After completing the interactions, we obtained two sets of human evaluation results. To check the sensitivity of changes, these two sets of human evaluation results (Tables 10 and 11) are compared with the human evaluation results, depicted in Table 3 of the main paper.

| <b>Model</b> | <i>PCA</i>  | <i>GAA</i>  | <i>PA</i>   | <i>EA</i>   | <i>FY</i>   | <i>CY</i>   | <i>N<sub>rep</sub></i> |
|--------------|-------------|-------------|-------------|-------------|-------------|-------------|------------------------|
| GPT2-large   | 1.91        | 2.66        | 1.71        | 1.74        | 2.64        | 1.95        | 2.16                   |
| ARDM         | 2.40        | 3.00        | 2.69        | 2.61        | 3.80        | 2.32        | 2.44                   |
| Zephyr-7B    | 2.85        | 3.15        | 3.48        | 3.65        | 4.20        | 3.54        | 2.65                   |
| Phi-1.5      | 2.77        | 3.11        | 4.38        | 3.93        | 4.30        | 3.65        | 2.75                   |
| PDSS         | 3.02        | 3.70        | 4.50        | 4.10        | 4.15        | 3.75        | 2.95                   |
| ABLE-R       | 3.08        | 3.76        | 4.67        | 4.20        | 4.05        | 4.20        | 3.45                   |
| ABLE-TR      | 3.13        | 3.72        | 4.73        | 4.29        | 4.23        | 4.25        | 3.65                   |
| ABLE-GR      | 3.24        | 3.78        | 4.85        | 4.31        | 4.32        | 4.38        | 3.75                   |
| <b>ABLE</b>  | <b>3.44</b> | <b>4.02</b> | <b>4.97</b> | <b>4.54</b> | <b>4.40</b> | <b>4.49</b> | <b>4.03</b>            |

Table 10: Results of human evaluation for Set 1.

| <b>Model</b> | <i>PCA</i>  | <i>GAA</i>  | <i>PA</i>   | <i>EA</i>   | <i>FY</i>   | <i>CY</i>   | <i>N<sub>rep</sub></i> |
|--------------|-------------|-------------|-------------|-------------|-------------|-------------|------------------------|
| GPT2-large   | 1.87        | 2.59        | 1.68        | 1.68        | 2.70        | 1.98        | 2.23                   |
| ARDM         | 2.35        | 2.91        | 2.61        | 2.50        | 3.89        | 2.40        | 2.36                   |
| Zephyr-7B    | 2.78        | 3.08        | 3.38        | 3.56        | 4.15        | 3.44        | 2.55                   |
| Phi-1.5      | 2.83        | 3.19        | 4.48        | 3.83        | 4.25        | 3.75        | 2.85                   |
| PDSS         | 3.10        | 3.78        | 4.58        | 4.01        | 4.04        | 3.85        | 3.05                   |
| ABLE-R       | 3.00        | 3.65        | 4.58        | 4.12        | 4.10        | 4.08        | 3.36                   |
| ABLE-TR      | 3.20        | 3.77        | 4.63        | 4.20        | 4.13        | 4.18        | 3.56                   |
| ABLE-GR      | 3.27        | 3.85        | 4.75        | 4.30        | 4.21        | 4.28        | 3.76                   |
| <b>ABLE</b>  | <b>3.39</b> | <b>4.00</b> | <b>4.88</b> | <b>4.45</b> | <b>4.31</b> | <b>4.41</b> | <b>3.96</b>            |

Table 11: Results of human evaluations for Set 2.

#### A.4.2 Results Analysis

Upon comparing the two variations of human evaluation results with the original results presented in Table 3, we observe minor fluctuations in the metrics across different models. These fluctuations fall within a range of +0.05 to -0.05, indicating slight variability in the evaluation.

In Set 1 (Table 10), we notice marginal increases or decreases in some metrics for certain models compared to the original evaluation. For example, the *PCA* score for ABLE increased by 0.01, while the *GAA* score increased by 0.05. Similarly, in Set 2 (Table 11), there are fluctuations in the metrics, with some models showing slightly higher or lower scores compared to the original evaluation.

Overall, these minor variations suggest that the changes made to the human evaluation results have not significantly altered the assessments of ABLE’s performance. The consistency in the observed patterns across different variations provides additional confidence in the reliability of the evaluations and indicates the robustness of ABLE’s responses against biases.

## B Frequently Asked Questions

- **1. How does ABLE address the limitations of existing support systems for individuals with physical disabilities, and can it effectively adapt to varying user needs and preferences?**

**Answer:** ABLE recognizes the shortcomings of conventional support systems by prioritizing personalization and empathy. Unlike generic responses, ABLE tailors its interactions to individual user characteristics, preferences, and needs. By incorporating politeness and empathy cues, ABLE fosters effective communication and rapport, overcoming the impersonal nature of many existing systems. Its adaptability lies in utilizing a large-scale persona-tailored dataset (PERPDSCD) and a reinforcement learning framework. With the help of user personality traits, politeness, and empathy information, ABLE learns to generate responses that align with individual profiles. Additionally, its novel reward function, employing four reward models, guides ABLE in tailoring responses based on appropriate politeness and empathy levels. This adaptability ensures that ABLE can cater to the diverse needs and preferences of users with physical disabilities.

- **2. How did you ensure that the PerPDSCD dataset captures a comprehensive range of scenarios and issues related to physical disabilities?**

**Answer:** The creation of the PerPDSCD dataset involved a structured approach guided by clear objectives aimed at capturing diverse scenarios relevant to individuals with physical disabilities. We crafted prompts outlining guidelines for generating multi-turn conversations covering topics, such as Mobility Aids, Home Modifications, Physical Therapy Exercises, Assistive Technology, and more. Additionally, we integrated seed utterances provided by human experts to initiate conversations that address specific challenges faced by individuals with physical disabilities. Through iterative feedback and refinement, we ensured that the dataset encompasses a comprehensive range of scenarios and issues related to physical disabilities.

- **3. How did you ensure the authenticity and relevance of the dialogues in the PerPDSCD dataset?**

**Answer:** The authenticity and relevance of dialogues in the PerPDSCD dataset were assured through robust quality control measures. This involved manual checks by human participants, expert reviews by medical health experts, and continuous improvement measures at every stage. Human experts crafted seed utterances based on real dialogues and WHO guidelines, guiding the conversation generation process. Dialogues were generated using the GPT-3.5 model, with iterative feedback and refinement to enhance authenticity. Additionally, dialogues underwent automated quality checks to ensure coherence, content consistency, and naturalness, further enhancing the dataset's authenticity and relevance.

- **4. How do the novel rewards designed in the ABLE framework contribute to guiding the learning process and promoting desirable response generation behaviors?**

**Answer:** The novel rewards designed in the ABLE framework guide the learning process and make sure the generation of responses is aligned with user characteristics and desired interaction qualities. Rewards such as Persona-Consistency and Gender-Age-Consistency encourage the model to generate responses consistent with user attributes, promoting personalized interactions. Politeness Correctness and Empathy Correctness rewards reinforce the importance of politeness and empathy in responses, fostering supportive and respectful communication. Additionally, rewards like naturalness and conversation coherence promote linguistic fluency and coherent conversation flow, enhancing the overall quality of interactions. By incorporating these rewards, the ABLE framework facilitates adaptive and empathetic support tailored to individual user needs.

- **5. How does using automatic evaluation metrics and the two-phase human evaluation process enhance the reliability and comprehensiveness of assessing ABLE's performance?**

**Answer:** By using both automatic evaluation metrics and the two-phase human evaluation



process, the assessment of ABLE's performance becomes more robust and thorough. Human evaluators provide subjective insights into interaction quality, system fluency, consistency, and non-repetitiveness. At the same time, automatic metrics offer objective measures of persona accuracy, gender-age accuracy, politeness accuracy, and empathy accuracy. The two-phase human evaluation process ensures consistency and reliability through expert validation and iterative refinement of evaluation criteria. This combined approach provides a comprehensive understanding of ABLE's effectiveness, balancing subjective user experience with quantitative measures, thereby enhancing the reliability and validity of the evaluation results.