

# PIRsuaDer: A Persuasive Chatbot for Mitigating Psychological Insulin Resistance in Type-2 Diabetic Patients

Sujatha Das Gollapalli, See-Kiong Ng

Institute of Data Science

National University of Singapore

{idssdg, seekiong}@nus.edu.sg

## Abstract

Psychological Insulin Resistance (PIR) is described as the reluctance towards initiation and adherence of insulin-based treatments due to psychological barriers in diabetic patients. Though studies have shown that timely initiation with lifestyle changes is crucial in sugar control and prevention of chronic conditions in Type 2 Diabetes (T2D) patients, many patients often have deep-rooted fears and misgivings related to insulin which hinder them from adapting to an insulin-based treatment regimen when recommended by healthcare specialists. Therefore, it is vitally important to address and allay these fallacious beliefs in T2D patients and persuade them to consider insulin as a treatment option. In this paper, we describe the design of PIRsuaDer, a persuasive chatbot for mitigating PIR in T2D patients. In PIRsuaDer, we effectively harness the conversation generation capabilities of state-of-the-art Large Language Models via a context-specific *persuasive* dialog act schema. We design reward functions that capture dialog act preferences for persuading reluctant patients and apply reinforcement learning to learn a dialog act prediction model. Our experiments using a collection of real doctor-diabetic patient conversations indicate that PIRsuaDer is able to improve the willingness in patients to try insulin as well as address specific concerns they have in an empathetic manner.

## 1 Introduction

According to the International Diabetes Federation (IDF), in 2021,<sup>1</sup> about 10.5% of the adult population had diabetes. It was estimated that diabetes was responsible for 6.7 million deaths that year and projected that 1 in 8 adults may be living with diabetes in 2045. Of the diabetic patients, it is estimated that about 90% have Type 2 Diabetes (T2D), as a result of various socio-economic, demographic,

environmental, and genetic factors (Wu et al., 2014; Dai et al., 2023).

Insulin is an essential hormone produced by the body for converting food to energy and managing blood sugar levels. In diabetic patients, the body is unable to make enough insulin or use it properly due to which medications and sometimes, insulin therapy are needed for managing their blood sugar levels. Indeed, studies indicate that early initiation of insulin is crucial for prevention of chronic complications for T2D patients (Davies et al., 2022). However, many patients delay initiation of insulin therapy despite advice from healthcare professionals, often considering it only after the failure of other treatments (Balogh et al., 2020). This phenomenon, referred to as ‘psychological insulin resistance’ (PIR), has been attributed to barriers such as the fear of hypoglycemia, the need for glucose monitoring, the stigma and fear associated with the methods of insulin administration (injections), misconceptions related to gaining weight, lifelong commitment etc (Peyrot et al., 2005; Hussein et al., 2019; Galdón Sanz-Pastor et al., 2024). Delays in insulin initiation not only will result in many complications to the patients at individual level but can also become an encumbrance to the healthcare budget and economy (Bommer et al., 2018).

Recent studies have indicated that educational programmes soon after diagnosis regarding treatment options increase the likelihood of patient adherence and it is of utmost importance to help T2D individuals “learn and recognize the potential personal value of initiating insulin treatment” as well as address their negative beliefs about insulin to overcome this reluctance towards insulin as well as ensure their treatment adherence and long-term outcomes. PIR can be effectively resolved when healthcare professionals take the time to address injection concerns by showing and/or demonstrating the actual injection process, explaining the benefits of insulin, and adopting a collaborative, communi-

<sup>1</sup><https://diabetesatlas.org/>

Dialog Act	Utterance
neutral_to_info.	<i>Client</i> : I'm just here to chat about my diabetes management, nothing too serious.
propose	<i>Counselor</i> : Have you considered trying insulin for better control of your blood sugar levels and to prevent potential long-term complications?
deny_to_try	<i>Client</i> : I'm really not sure about trying insulin; it just seems uncomfortable and inconvenient, especially when I'm out eating.
offer_relief	<i>Counselor</i> : I understand your concerns about discomfort, but there are options like once-a-day injections that can be taken at home, minimizing any inconvenience.

Table 1: An example snippet from a conversation in PIRsuader

cation style (Polonsky et al., 2019; Snoek et al., 2021). However, healthcare professionals who are often overloaded and suffer from work stress and burnout may not always be available to gently persuade T2D patients into overcoming their PIR (Portoghese et al., 2014). Instead, *can we develop a counseling chatbot that can gently and effectively allay the concerns and fears of reluctant T2D patients, educate them on the benefits of insulin, and persuade them for insulin initiation?*

Persuasion is a communication skill employed in everyday life (Burns, 1991; O’Keefe, 2018) to influence other people’s thoughts and opinions using convincing arguments and facts, and sometimes, by exploiting their biases and emotions (Zeng et al., 2024). Persuasion has often been employed to enable individuals into considering alternatives viewpoints for a given situation during their decision making and can be harnessed effectively for the purposes of “social good” such as donations to charities or counseling people into seeking therapy for their mental health problems (Wang et al., 2019; Chen and Yang, 2021).

However, it is challenging to computationally model persuasion in language generation, and this capability is not transferrable across domains. Therefore, several studies focus on the identification of persuasion strategies for a specific context (Tan et al., 2016; Hidey and McKeown, 2018; Wang et al., 2019). We adopt a similar approach and investigate how to generate persuasive dialogs with Large Language Models (LLMs) for our specific context of conversing with T2D patients. In our chatbot, we seek to mitigate a patient’s PIR towards insulin treatment and allay their deep-rooted negative beliefs and fears regarding insulin. Our contributions are as follows:

- We develop a schema of dialog acts for charac-

terizing empathetic, persuasive conversations with patients of Psychological Insulin Resistance. Our schema can be used to suitably “prompt” LLMs, providing a basis for dialog generation with the objective of mitigating PIR in T2D patients through persuasion.

- We employ conversation analysis on a collection of real patient-doctor conversation transcripts to extract insulin-related concerns, questions and responses. The insights derived from this analysis are used in combination with our proposed schema to “mimic” realistic persuasive client-counselor conversations.<sup>2</sup> Using two reward functions that capture preferences between utterances in persuasive dialog, we learn a dialog act predictor for persuasive counseling using Reinforcement Learning. In this manner, we overcome the lack of annotated data while designing PIRsuader, our novel, domain-specific chatbot for counseling PIR clients.

Our datasets of patient case notes as well as crowd-annotated conversations for specific persuasive aspects are released for reproducibility and to enable future research.<sup>3</sup> A conversation snippet from our collection is shown in Table 1 for illustration.

**Organization:** In the next section (Section 2), we present the schema developed by us for persuading PIR patients. Section 3 describes our dialog system framework for generating persuasive conversations along with the reward functions used to learn a dialog act predictor. In Section 4, we summarize our datasets, experimental setup, results and findings.

<sup>2</sup>Given the counseling context, we refer to T2D patients with PIR, alternatively, as clients in this paper.

<sup>3</sup><https://github.com/NUS-IDS/coling25pirsuader>

Closely related works are summarized in Section 5, and we conclude our paper with a discussion on possible future work in Section 6. Some limitations of our study are highlighted in Section 7.

## 2 Dialog Acts Schema for PIRsuader

In a recent study, Zeng et al. (2024), presented a comprehensive taxonomy for persuasion strategies based on an extensive survey of Social Science and Psychology research. They grouped ethical persuasive strategies into 13 broad strategies such as Information-based (presenting evidence and facts), Emotion-based (appealing to a person’s positive or negative emotions), and Credibility-based (where expert endorsement is used). As our focus is on conversations, we consider “dialog acts” to enable a specific persuasive strategy, and we extend recent taxonomies for persuasive and empathetic dialog and modify them for our specific scenario of counseling PIR clients (Wang et al., 2019; Chen and Yang, 2021; Svikhnushina et al., 2022).

Our target users are PIR patients who were prescribed insulin by their doctors but are reluctant to adopt the same and our objective is to mimic the role of health counselors who handle such clients by eliciting their reasons for resistance and providing any information which can clarify their misunderstandings thus persuading them towards insulin treatment. Not all persuasion strategies are applicable to our scenario. For example, persuasion strategies targeting reciprocity, loyalty or creating a sense of scarcity or pressure are improper in our usage context (Zeng et al., 2024).

In Table 2, we list the dialog acts most relevant to persuasive dialog in our setting along with examples. The full list used in PIRsuader is provided in Table 10 of the Appendix. In Table 2, acts such as `provide_insulin_information`, `logical_appeal`, `suggest_a_reason` are “information-based”, whereas `emotion_appeal` and `motivate` are “emotion-based” and “appraisal-based”, respectively (Zeng et al., 2024). The `credibility_appeal` act captures utterances that reference expert endorsement and dialog acts that mostly involve questions such as `ask_about_consequence`, `ask_about_antecedent`, `ask_concerns` align with the Socratic method, where questions are used to elicit critical thinking in persons through open-ended questions on other possibilities and alternatives (Neenan, 2009; Vittorio et al., 2022).

Given the above described possible dialog acts for our usage, we now address how to incorporate them within our PIRsuader chatbot.

## 3 Methods

Large language models (LLMs)-powered chatbots are now common due to their human-like capabilities with respect to language understanding and response generation. Several state-of-the-art (SOTA) LLMs such as from the GPT-family (Ouyang et al., 2022; OpenAI, 2023), Vicuna (Chiang et al., 2023), LLaMA2-Chat (Touvron et al., 2023), and Qwen2 (Yang et al., 2024) provide easy-to-use APIs and can be efficiently incorporated into a dialog system via suitable instructions/prompts. Indeed, it is no longer necessary to train current-day dialog systems end-to-end by employing large datasets. Despite this noteworthy development, relying on the planning capabilities of LLMs may not suffice to steer conversations in contexts that involve specific objectives such as negotiation or tutoring. It has been observed that a “dialog policy” that provides appropriate instructions to the LLM during an interaction while tracking these objectives is necessary for building an effective dialog agent (Jang et al., 2022; Deng et al., 2024) and self-refine methods with Reinforcement Learning are heavily employed for this purpose (Madaan et al., 2023; Ouyang et al., 2022).

Based on these observations, in PIRsuader, we rely on the SOTA LLM’s response generation abilities and focus instead on building a suitable dialog act prediction model. That is, our dialog act prediction model ensures that LLM responses are geared towards the objective of persuading PIR patients by steering the conversation via appropriate “dialog act” inputs to the LLM.

### 3.1 System Setup

We adapt the recently proposed plug-and-play policy dialog planner setup from Deng, et al. (2024) in PIRsuader. The outline of our system is illustrated in Figure 1 and our goal is to learn the reward functions and dialog act predictor for our setting (part 2 in this figure).

Since annotated data for learning dialog act prediction for persuasion in our specific context is unavailable, we first learn a **baseline predictor** via data augmentation by generating synthetic conversations with LLMs (Van et al., 2021; Sahu et al., 2022). That is, we use two LLMs to separately

Dialog Act	Description
<u>provide_insulin_information</u>	Provide information in response to a question on insulin or diabetes. <i>Example: Common side effects of insulin include low blood sugar, weight gain, and injection site . . .</i>
<u>ask_concerns</u>	Ask about concerns related to insulin <i>Example: Can you share more about what is holding you back from trying insulin?</i>
<u>propose</u>	Suggest trying insulin <i>Example: Have you considered trying insulin to better manage your blood sugar levels?</i>
<u>logical_appeal</u>	Provide logical reasoning to trying insulin <i>Example: Uncontrolled diabetes is a primary contributor to kidney issues, so using insulin to . . .</i>
<u>emotion_appeal</u>	Emotionally appeal to why they should try insulin <i>Example: Think about how starting insulin could help . . . enjoy more of life’s moments without . . .</i>
<u>credibility_appeal</u>	Use research studies to convince why they should try insulin <i>Example: Research shows that delaying insulin therapy when it’s needed can lead to, . . .</i>
<u>ask_about_consequence</u>	Ask about the result of the described action or situation <i>Example: What do you think might happen if your blood sugar levels remain consistently high . . .</i>
<u>ask_about_antecedent</u>	Ask about the reason or cause of the described state or event. <i>Example: What has led you to feel hesitant about starting insulin at this time?</i>
<u>suggest_a_reason</u>	Suggest a specific reason or cause of the event or state <i>Example: Have you considered that insulin use doesn’t have to be a lifelong commitment?</i>
<u>motivate</u>	Encourage the speaker to move onward <i>Example: Trying insulin could be a game-changer for your morning blood sugar . . .</i>

Table 2: List of Persuasive Dialog Acts used by the Counselor in PIRsuader with examples

role-play the two “actors”: (1) a client with PIR and (2) the counselor who seeks to persuade the client (Part 1 in Figure 1). In each turn, appropriate dialog acts are selected from the provided schema for the two roles based on the LLM’s inherent understanding of how to persuade (as frozen in its parameterized knowledge). The precise prompt templates are listed in Table 12 of the Appendix.

<b>deny_to_try</b> ask_concerns, logical_appeal, emotion_appeal, credibility_appeal, ask_about_consequence, ask_about_antecedent, suggest_a_reason, motivate
<b>express_interest</b> amplify_excitement, motivate, compliment, provide_insulin_information, support
<b>neutral_to_information</b> credibility_appeal, motivate, compliment, provide_insulin_information
<b>counter_information</b> credibility_appeal, provide_insulin_information

Table 3: Possible dialog-act pairs used in persuasion

### 3.2 Reward Functions

**Motivation:** Given the counselor’s objective to mitigate PIR in the client, mentions of benefits of insulin and indications to switch to insulin will naturally occur during interactions with the client (For

example, using the propose dialog act from Table 2). In response, the client may express interest in knowing more details, express specific concerns, or express refusal.<sup>4</sup> Indeed, it is precisely at these points the counselor would engage in persuasion strategies with the intent to influence the client. For example, against a refusal, the counselor may ask questions to understand further or appeal to logic or emotions in the client based on their assessment of the client’s psyche. It is these act preferences that our automatic dialog act predictor needs to learn to model in PIRsuader to increase the likelihood of a successful persuasion.

The list of “client dialog act, counselor persuasive act” pairs whose preferences need to be captured in our reward functions are shown in Table 3. Based on previous works (Deng et al., 2023; Feng et al., 2023), we designed the following reward functions for our setting:

**1. Preference-based Reward (PBR):** This reward function is learnt by modeling human preferences data. For each of the client’s acts listed in Table 3, we sample pairs of possible counselor act choices and generate the responses with an LLM. Next, we ask human annotators to select which of

<sup>4</sup>The client dialog acts are listed in Table 10.

the two responses is likely to be more persuasive to the client given the conversation snippet. This human feedback is used to train a reward model by converting the pairwise preference data to a binary comparison (*chosen* versus *rejected*) and learning a function that scores the human preferred response to have a higher score than its counterpart. This model trained via the popular RLHF procedure (Reinforcement Learning using Human Feedback) is able to process a conversation snippet and output a scalar score as reward (Ouyang et al., 2022).

**2. Critic-assessed Reward (CAR):** This reward function is learnt by using LLMs in the role of a Critic. Once again, the dialog act pairs in Table 3 are used to tentatively model the conversation completion by generating the next two utterances (one by the Counselor LLM and one by the Client LLM). At the end of this completion, the Critic LLM is used to assess how willing the client is to try insulin on a Likert-scale of 1 (least willing) to 5 (most willing) based on the conversation snippet. The generated snippet-rating pairs are used to learn a five-class classification model that acts as our reward function during RL. Previously, Critic LLM-generated rewards were used at every turn in an RL framework using the conversation generated so far (Madaan et al., 2023; Jang et al., 2022; Deng et al., 2024). We generate possible “future conversation” and learn a reward function based on “potential willingness” as assessed by the Critic.

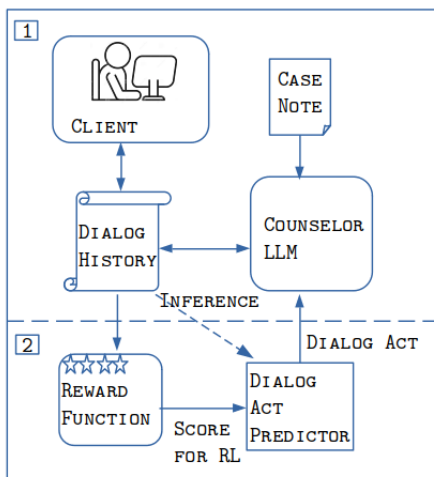


Figure 1: System Outline for PIRsuader

## 4 Experiments

**Conversations Dataset:** As part of a large project that uses AI methodologies for transforming healthcare for chronic conditions such as Diabetes, Hy-

pertension and hyperLipidemia (*JARVIS<sub>DHL</sub>*), we have access to a collection of (fully-anonymized) doctor-patient interactions from a local hospital.<sup>5</sup> We subsetted about 74 transcripts of sessions between diabetic patients and their specialist doctors from this collection. From manual examinations, we note that during these sessions the doctor reviews the patient’s overall management of their condition and recommends insulin treatment due to their poor control of diabetes. In most cases, the patients are unwilling to act on this recommendation and express various reasons for their reluctance ranging from not wanting “holes” in the body, not working for somebody else, cost of injections, fear of life-long injections, etc. A snippet from our dataset is shown in Table 4 for illustration:

<p><i>Doctor:</i> Injection ah? Why? Because it’s painful or what?  <i>Patient:</i> It’s not painful but I feel that ...I just feel very uncomfortable.  <i>Doctor:</i> Uncomfortable?  <i>Patient:</i> Yah, because I (have) gone through a lot of [brief pause] errr, so-called, testing blood. Yah, a lot! You see this month, maybe three, four time(s).  <i>Doctor:</i> Blood-taking ah, is very different from insulin injection.  <i>Patient:</i> Oh.  <i>Doctor:</i> The needles are different. The needles are finer.</p>
--

Table 4: Snippet from a real doctor-patient conversation

The patient-doctor conversations in this collection are long and, on average, have ~200 turns. The conversations were manually-labeled to indicate if the patient agreed to try insulin at the end of the session. Overall, only 31% of the patients were willing to try insulin by the end of their sessions.

**Data for PIRsuader:** We employ LLMs to convert the above conversations for training our counseling chatbot for PIR mitigation as follows: We first apply conversation analysis prompts with LLMs to summarize a patient’s persona, their concerns and reasons for not wanting to adopt insulin treatment, along with how the doctor addresses the same to form a “case note” for a PIR client based on their conversation transcript. The client’s concerns and persona are used in role play by the Client LLM whereas the doctor responses are used by the Counselor LLM in conjunction with various dialog acts for generating responses for mitigating the client’s concerns. For example, one of the utterances from our counselor LLM includes “I un-

<sup>5</sup><https://ids.nus.edu.sg/aainhealth.html>

*derstand your worries about discomfort, but with options like once-a-day injections at home and using fine needles, it can be a much more manageable experience*". The prompt templates used for creating case notes and role play are listed in Tables 8 and 12 of the Appendix and our dataset is made publicly-available for enabling future research.<sup>3</sup>

We note that a novel application such as PIRsuader is only possible due to the remarkable instruction understanding and language generation capabilities offered in SOTA LLMs that enable data augmentation for learning the dialog act model. As such, gathering data from the ground up for an application like ours to learn a dialog agent (traditionally) would have been rather challenging considering the health care domain with its privacy constraints. We train the baseline dialog-act prediction model on the data obtained with role play. This model is further fine-tuned using RL with the two reward functions, **PBR** and **CAR**.

#### 4.1 Crowd-annotated Data

For collecting preferences data for learning the **PBR** reward model (task-1), and for evaluating the generated conversations (task-2), we employed the crowd annotation platform from Amazon Mechanical Turk.<sup>6</sup> In task-1, for a given conversation snippet, annotators select which of two counselor utterances (each generated using a different dialog act) is better given the conversation so far and the final goal of persuading the client. In task-2, annotators answer a list of four questions related to conversation quality capturing the persuasiveness, empathy, and specificity of information in counselor responses and change in willingness of the client. More details are provided in the Appendix, Tables 9 and 11.

**Annotation Quality:** As in similar works (Wang et al., 2020; Gollapalli and Ng, 2022), worker quality was ensured by requiring the crowdworkers to have greater than 98% HIT ("Human Intelligence Task") approval rate, a minimum of 10,000 HITs, be located in the United States (for language ability) and only selecting those workers who pass the qualification tests (different for each task) with a score of 80% or above. On par with similar tasks, we paid each worker about USD 0.20 per HIT for task-1 and each HIT was annotated by two workers. For the more complex task-2, each HIT was annotated for quality by five annotators who were

paid USD 0.40 per HIT. In total, we evaluated 78 conversation snippets for task-2 and 717 utterance pairs for task-1.

Since each HIT is rated by different set of random "k" raters, we use ICC1k scores that indicate the mean reliability of raters for computing inter-rater reliability (IRR) scores for our annotated datasets (Koo and Li, 2016). The ICC1k values were 0.53 for preferences data (task-1), and 0.85 for persuasiveness, 0.66 for willingness, 0.52 for empathy, and 0.93 for specificity for conversation quality ratings data (task-2).

The IRR scores show moderate reliability for task-1 and good, and excellent reliability on the persuasiveness and specificity ratings of counselor responses, respectively, and good, and moderate reliability for willingness of the client and empathy ratings. Therefore, while using task-1 data in **PBR**, we only consider examples where both the workers selected the same response (unison rating) and the resulting dataset has 493 utterance preferences. For task-2 data, since the IRR scores are high, we directly used means of all the ratings.<sup>3</sup>

**Experimental Setup:** All experiments were performed on a single GPU of an Nvidia Tesla cluster (Linux) machine with 32GB RAM. The baseline predictor for dialog acts was learnt using the FlanT5-large model.<sup>7</sup> The dialog act models are trained to predict which of the dialog acts the Counselor LLM must use next, given the dialog history. We used the Roberta-large<sup>8</sup> model for both our reward functions—PBR predicts a scalar score whereas CAR is a classification model predicting one of Likert scale values from 1 to 5 that are normalized between 0 and 1 for a reward. Reinforcement Learning was performed with the TRL library using the Proximal Policy Optimization (PPO) algorithm (Schulman et al., 2017).<sup>9</sup> For LLM, we used GPT-4o-mini via the OpenAI API in view of its cost-efficiency, and on-par performance with GPT-4o for our task in preliminary runs.<sup>10</sup> We did not investigate other LLMs in view of the focus in this study and due to the competitive performance with OpenAI models.

<sup>7</sup><https://huggingface.co/google/flan-t5-large>

<sup>8</sup><https://huggingface.co/FacebookAI/roberta-large>

<sup>9</sup><https://huggingface.co/docs/trl/index>

<sup>10</sup><https://platform.openai.com/docs/models/gpt-4o-mini>

<sup>6</sup><https://www.mturk.com/>

## 4.2 Evaluation and Results

**Dialog Act Prediction:** Since ground-truth conversations annotated with dialog acts from our schema are unavailable, it is not possible to evaluate the correctness of the dialog act labels predicted by our RL-trained models. However, considering that we created our initial dataset for training the baseline model using LLMs, we can compare the LLM-picked dialog act labels with those obtained with RL-PBR and RL-CAR to characterize their differences. The overlap in the labels of the test dataset was 47.50% and 44.37%, respectively, which essentially indicates that the RL model predictions diverge with those of the default LLM more than half the time.

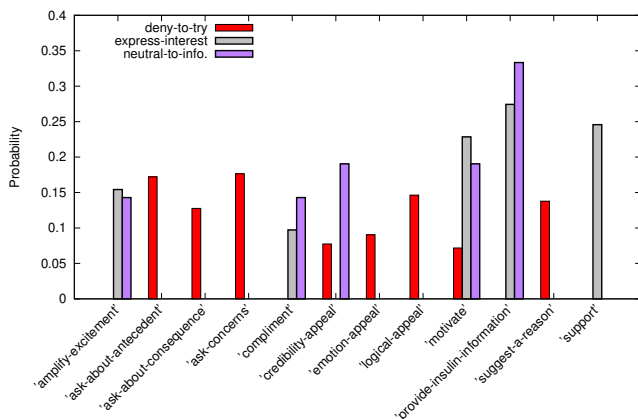


Figure 2: Distribution of counselor dialog-act preferences for a given client dialog-act

**Human Preferences:** We convert the preference pairs data collected in task-1 into a probability distribution to highlight what human annotators consider a better dialog act/utterance for the counselor given a specific dialog act/utterance by the client (Table 3). It can be seen in Figure 2, that when the client expresses interest or is neutral towards the presented information, employing the dialog acts which provide more information on insulin, motivating, or citing credible sources are more helpful than the other dialog acts. Interestingly, when the client expresses reluctance to try insulin, the Socratic dialog acts which probe reasons (ask\_concerns) or elicit reflection (ask\_about\_consequence, ask\_about\_antecedent) are the most helpful for persuasion. Not surprisingly, the Socratic questions are well used in counseling and coaching contexts to enable a cognitive change in clients through self-reflection (Neenan, 2009; Vittorio et al., 2022).

**Conversation Quality:** For our final evaluation,

Feature	PBR	CAR	GPT
willingness (unchanged) ↓	78	<b>64</b>	72
willingness (improved) ↑	15	<b>31</b>	11
questions ignored ↓	20	<b>13</b>	22
questions answered ↑	76	<b>83</b>	74
not empathetic/inconsiderate ↓	52	<b>38</b>	54
empathetic ↑	48	<b>50</b>	46
persistent ↓	27	<b>12</b>	50
persuasive ↑	39	<b>45</b>	32

Table 5: The percentages of conversations are marked for specific features for the three experimental settings. ↑ and ↓ are used to indicate if a higher or lower score is preferred for a given feature.

we focus on our end-goal, viz., that of persuading clients with PIR. For this, we randomly selected 13 patients (~17% of the conversations dataset), half of whom showed willingness to try insulin. After creating case notes for these patients, we generate conversations using role play with the Client and Counselor LLMs (as presented in Section 3.1) in two settings: with and without the client’s persona information. For each of the two settings, we generated three conversations– (1) PIRsuader with **PBR**-chosen dialog acts, (2) PIRsuader with **CAR**-chosen dialog acts, and (3) using (default) **GPT-4o-mini** chosen dialog acts. In total, about 78 conversations were generated for about 20 turns between the counselor and the client (we exit early if the end\_conversation dialog act is seen during a run). We preferentially selected snippets from the conversations where a deny\_to\_try act is seen from the client if it’s available, or randomly, from the first few turns, otherwise. From this point, about eight turns of the conversation (four rounds of utterance exchange) was used to create a snippet for evaluation by the human annotators (task-2 on the AMT platform). The human annotators were asked to answer questions related to change in the willingness of the client, and the nature of persuasiveness, empathy, and specificity in the counselor responses. Specificity refers to whether the counselor provided specific factual information in response to specific questions asked by the client. The human annotators also indicate if the counselor attempted to persuade, if the persuasion attempt was good enough, persistent despite reluctance from the client, or inadequate. The precise questions are listed in Table 11 of the Appendix.

Our findings from the above human evaluation

Turn	Dialog Act	Utterance
5	express_interest	<i>Client</i> : I'm really interested in hearing more about the lifestyle changes that could help lower my blood sugar levels.
6	express_interest	<i>Counselor</i> : That's great to hear! Let's explore some effective lifestyle changes together.
7	express_interest	<i>Client</i> : I'm eager to learn more about those lifestyle changes that can help me manage my blood sugar levels.
8	express_interest	<i>Counselor</i> : I'm glad to see your eagerness to learn more about managing your blood sugar levels!
9	express_interest	<i>Client</i> : Could you share specific lifestyle changes that might help with my blood sugar management?
10	suggest_a_solution	<i>Counselor</i> : How about incorporating a balanced diet and regular exercise into your routine while monitoring your blood sugar levels closely?

Table 6: Conversation snippet illustrating the effect of unsuitable dialog acts

study are summarized in Table 5. Overall, we find that in about 30% of the conversations in the **CAR** setting, conversations show improvements in willingness of the clients, whereas for most conversations (60-80%) the willingness remains unchanged across the different settings. The number of questions ignored without providing answers is the least in the **CAR** setting and almost 83% of the questions asked were answered partially or satisfactorily. Almost 50% of conversations in **PBR** and **GPT** (default) were considered not very empathetic or (sometimes) even inconsiderate by the annotators. Finally, the number of persuasive chats is the highest in the **CAR** settings and significantly, the number of “annoyingly” persuasive or persistent attempts is the least as well. The persistent persuasion attempts are notably high with **GPT**. Previously, subject-based studies on interventions note that persistent suggestions, authoritarian style, or using insulin as threat are least helpful for handling PIR patients and the focus should be on encouraging self-care and addressing the specific beliefs that support an individual patient’s resistance presented in a collaborative style (Polonsky et al., 2019; Peyrot et al., 2005; Balogh et al., 2020). Overall, PIRsuader with dialog acts selected by the RL-CAR model is the most persuasive, empathetic, and specific one among the three settings.

**Error Analysis:** In Table 1, a snippet from a conversation session with our best performing **CAR**-model is shown. We showcase the different paths the conversations take using dialog acts generated by the **CAR** versus **PBR** models for the same patient data point in Tables 13 and 14 of the Appendix. Though **CAR**-generated conversations are, on average, significantly better than the other approaches (Table 5), the generated dialog is not always “perfect”. In particular, errors from dialog act prediction result in unsuitable Counselor responses for a

given Client utterance. This behavior is illustrated in the examples shown in Tables 6 and 7. In Table 6, we observe that in response to the expression of interest by the Client, the predicted dialog acts (`express_interest`) cause the Counselor LLM to generate responses expressing interest and providing encouragement for a few turns (Turns 6 and 8) without conveying any useful information until the specific question at Turn 9.

Similarly, in the conversation snippet shown in Table 7, the ideal Counselor response should not include the dialog act `propose` in response to a `deny_to_try` act (Table 3). This snippet also illustrates the missing long-range tracking in our framework. For instance, considering the repeated denials by the Client (Turns 13 and 15), the Counselor utterance in Turn 18 that includes the suggestion to try insulin is being “insensitively persistent”.

## 5 Related Work

The development of computational models for persuasion detection has been ongoing in recent research with annotation schemes and prediction models developed for specific scenarios such as online forums, charity donation conversations, and argumentation (Tan et al., 2016; Hidey and McKewon, 2018; Wang et al., 2019; Chen and Yang, 2021). In most recent works, persuasion strategies were used to design prompts for “jailbreak”-ing LLMs (Zeng et al., 2024) and persuasion strategy prediction was incorporated with multi-modal learning for advertisements (Kumar et al., 2023).

From conversation point of view, single-turn persuasive utterance generation models were studied previously (Lukin et al., 2017). In general, though LLMs have revolutionized both open-domain and task-oriented dialog generation models (Ulmer et al., 2024; Gopalakrishnan et al., 2023; Yamazaki et al., 2023; Lee et al., 2023; Mendonça et al.,



Turn	Dialog Act	Utterance
13	deny_to_try	<i>Client</i> : I'm still not ready to try insulin, but I'd like to focus on my diet and other ways to manage my diabetes.
14	propose	<i>Counselor</i> : What if we explore how an insulin regimen could be adjusted to fit your eating habits and help you achieve better blood sugar control?
15	deny_to_try	<i>Client</i> : I really want to focus on my diet first before considering insulin.
16	ask_concerns	<i>Counselor</i> : What specific aspects of your diet are you most focused on improving before considering insulin?
17	request_information	<i>Client</i> : How can I make it easier to stick to my new diet plan?
18	suggest_a_solution	<i>Counselor</i> : Have you thought about how adjusting an insulin regimen to align with your meals could support your dietary goals?

Table 7: Conversation snippet illustrating the missing long-range tracking of “denials”

2024), researchers have noted their shortcomings in specific contexts such as counseling (Gollapalli et al., 2023), incorporation of specific styles such as proactive asking of questions (Deng et al., 2023) or providing emotional support (Kang et al., 2024), and for specific objectives such as negotiation (Deng et al., 2024). In these contexts, the state-of-the-art approaches decouple the dialogue act and language generation to focus on dialog strategy learning (He et al., 2018; Joshi et al., 2021).

Given its wide prevalence, there is unprecedented interest in the application of AI in the management of Diabetes and several AI-based solutions are being investigated for screening, health and diet monitoring, as well as analysis of lab reports. Though challenges of using AI techniques in real clinical practice remain (Wang et al., (2024; He et al., 2023), AI’s potential in developing applications for Diabetes education, as well as dietary recommendation and patient monitoring systems is being well harnessed (Alotaibi et al., 2016; Li et al., 2020; Khalifa and Albadawy, 2024; Chen et al., 2022). Indeed, our proposed chatbot precisely fits into this latter set of AI-enhanced educational context of diabetes management.

## 6 Conclusions and Future Work

We addressed the design of a counseling chatbot for PIR clients for alleviating their insulin-related fears and concerns by effectively leveraging on LLM’s powerful generative abilities guided by predictions of dialog acts based on persuasion strategies. To this end, we developed a scenario-specific schema of dialog acts that characterizes several relevant persuasion strategies such as reflective thinking, emotion appeal, and credibility appeal. Human preferences as well as “future reward” estimations were used to design reward functions and learn predictive models for dialog acts from this schema. Our RL-based dialog act predictor is used within

the PIRsuader system and is able to generate counseling conversations that are persuasive, without being aggressively persistent, and include factual information required to allay client-specific concerns related to insulin. In future, we would like to focus on deploying PIRsuader and collecting usage data with real clients to fine-tune and improve our system further. We would also like to investigate if our LLM-augmented datasets can be used for training local, smaller models to overcome resource and cost concerns of LLMs as well as meet privacy requirements of the health care domain.

## 7 Limitations

In this study, we designed PIRsuader using LLM APIs from OpenAI. We did not investigate how the results carry over to other open-source LLMs such as Llama (Touvron et al., 2023) and Qwen (Yang et al., 2024) which are acknowledged to have similar instruction understanding and generative capabilities.<sup>11</sup> We also did not investigate local models for the dialog agent which may, in fact, be more cost-effective and pose less risk privacy-wise. In this study, our models were primarily trained through data augmentation techniques. Real conversations relevant to this use-case are required for a complete validation of how comparable the generated datasets are. In addition, though our generated conversations were evaluated by human annotators and are indicative, it is unclear if there are subject-based differences between our annotators and “real” clients. While our evaluation is on dialog sessions simulated using LLMs with data from real patients, for the next step, a study of PIRsuader “in use” by real PIR patients is required.

<sup>11</sup>[https://huggingface.co/spaces/open-llm-leaderboard/open\\_llm\\_leaderboard](https://huggingface.co/spaces/open-llm-leaderboard/open_llm_leaderboard)

## Ethics Statement

This research was conducted in accordance with the ACM Code of Ethics. We ensured data quality by hiring qualified annotators at a reasonable pay for the tasks set up on the AMT platform (Section 4.1). The general ethical and privacy concerns for use of AI in health care are not applicable for our prototype system that has not been tested/deployed in real settings.

## Acknowledgments

This research/project is supported by the National Research Foundation, Singapore under its AI Singapore Programme (AISG Award No: AISG-GC-2019-001-2B).

## References

- Mohammed Alotaibi, Robert Istepanian, and Nada Philip. 2016. A mobile diabetes management and educational system for type-2 diabetics in Saudi Arabia (saed). *mHealth*, 2:33–33.
- Elemer G. Balogh, Magaly Perez-Nieves, Dachuang Cao, Irene I. Hadjiyianni, Nadeem Ashraf, Urvi Desai, Frank J. Snoek, and Jackie A. Sturt. 2020. Key strategies for overcoming psychological insulin resistance in adults with type 2 diabetes: The UK subgroup in the emotion study.
- Christian Bommer, Vera Sagalova, Esther Heesemann, Jennifer Manne-Goehler, Rifat Atun, Till Bärnighausen, Justine Davies, and Sebastian Vollmer. 2018. [Global economic burden of diabetes in adults: Projections from 2015 to 2030](#). *Diabetes Care*, 41:dc171962.
- R.B. Burns. 1991. *Persuasion by Communication*, pages 236–254. Springer Netherlands, Dordrecht.
- Jiaao Chen and Diyi Yang. 2021. Weakly-supervised hierarchical models for predicting persuasive strategies in good-faith textual requests. In *AAAI Conference on Artificial Intelligence*.
- Si Chen, Jingyi Lu, Danfeng Peng, Fengjing Liu, Wei Lu, Wei Zhu, Josh Zhou, and Weiping Jia. 2022. Effect of a mobile health technology-based diabetes education program on glucose control in patients with type 2 diabetes initiating premixed insulin: A prospective, multicenter, observational study. *Diabetes Care*, 46.
- Wei-Lin Chiang, Zhuohan Li, Zi Lin, Ying Sheng, Zhanghao Wu, Hao Zhang, Lianmin Zheng, Siyuan Zhuang, Yonghao Zhuang, Joseph E. Gonzalez, Ion Stoica, and Eric P. Xing. 2023. [Vicuna: An open-source chatbot impressing gpt-4 with 90%\\* chatgpt quality](#).
- Jiahui Dai, Xiaoyi Niu, Ann Bullock, Spero M. Manson, Joan O’Connell, and Luohua Jiang. 2023. Prevalence of Diagnosed Type 1 and Type 2 Diabetes Among American Indian and Alaska Native Peoples in 2012–2013. *Diabetes Care*.
- Melanie J. Davies, Vanita R. Aroda, Billy S. Collins, Robert A. Gabbay, Jennifer Green, Nisa M. Maruthur, Sylvia E. Rosas, Stefano Del Prato, Chantal Mathieu, Geltrude Mingrone, Peter Rossing, Tsvetalina Tankova, Apostolos Tsapas, and John B. Buse. 2022. Management of Hyperglycemia in Type 2 Diabetes, 2022. A Consensus Report by the American Diabetes Association (ADA) and the European Association for the Study of Diabetes (EASD). *Diabetes Care*.
- Yang Deng, Lizi Liao, Liang Chen, Hongru Wang, Wenqiang Lei, and Tat-Seng Chua. 2023. Prompting and evaluating large language models for proactive dialogues: Clarification, target-guided, and non-collaboration. In *Findings of the Association for Computational Linguistics: EMNLP 2023*.
- Yang Deng, Wenxuan Zhang, Wai Lam, See-Kiong Ng, and Tat-Seng Chua. 2024. Plug-and-play policy planner for large language model powered dialogue agents. In *The Twelfth International Conference on Learning Representations*.
- Yihao Feng, Shentao Yang, Shujian Zhang, Jianguo Zhang, Caiming Xiong, Mingyuan Zhou, and Huan Wang. 2023. [Fantastic rewards and how to tame them: A case study on reward learning for task-oriented dialogue systems](#). In *The Eleventh International Conference on Learning Representations*.
- A Galdón Sanz-Pastor, A Justel Enríquez, A Sánchez Bao, and FJ Ampudia-Blasco. 2024. Current barriers to initiating insulin therapy in individuals with type 2 diabetes.
- Sujatha Das Gollapalli, Beng Heng Ang, Mingzhe Du, and See-Kiong Ng. 2023. Counseling responses for mental health forum questions with early maladaptive schema prediction. In *ECAI*.
- Sujatha Das Gollapalli and See-Kiong Ng. 2022. QSTS: A question-sensitive text similarity measure for question generation. In *Proceedings of the 29th International Conference on Computational Linguistics*.
- Karthik Gopalakrishnan, Behnam Hedayatnia, Qianlang Chen, Anna Gottardi, Sanjeev Kwatra, Anu Venkatesh, Raefer Gabriel, and Dilek Hakkani-Tur. 2023. [Topical-chat: Towards knowledge-grounded open-domain conversations](#). *Preprint*, arXiv:2308.11995.
- He He, Derek Chen, Anusha Balakrishnan, and Percy Liang. 2018. Decoupling strategy and generation in negotiation dialogues. In *Proceedings of the 2018 Conference on Empirical Methods in Natural Language Processing*.
- Kai He, Rui Mao, Qika Lin, Yucheng Ruan, Xiang Lan, Mengling Feng, and Erik Cambria. 2023. A survey

- of large language models for healthcare: from data, technology, and applications to accountability and ethics.
- Christopher Hidey and Kathleen McKeown. 2018. Persuasive influence detection: The role of argument sequencing. *Proceedings of the AAI Conference on Artificial Intelligence*.
- A Hussein, A Mostafa, A Areej, AM Mona, A Shima, AG Najd, and A Futoon. 2019. The perceived barriers to insulin therapy among type 2 diabetic patients. In *African Health Science*.
- Youngsoo Jang, Jongmin Lee, and Kee-Eung Kim. 2022. GPT-critic: Offline reinforcement learning for end-to-end task-oriented dialogue systems. In *International Conference on Learning Representations*.
- Rishabh Joshi, Vidhisha Balachandran, Shikhar Vashishth, Alan Black, and Yulia Tsvetkov. 2021. Dialograph: Incorporating interpretable strategy-graph networks into negotiation dialogues. In *International Conference on Learning Representations*.
- Dongjin Kang, Sunghwan Kim, Taeyoon Kwon, Seungjun Moon, Hyunsouk Cho, Youngjae Yu, Dongha Lee, and Jinyoung Yeo. 2024. Can large language models be good emotional supporter? mitigating preference bias on emotional support conversation. In *Proceedings of the 62nd Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.
- Mohamed Khalifa and Mona Albadawy. 2024. Artificial intelligence for diabetes: Enhancing prevention, diagnosis, and effective management. *Computer Methods and Programs in Biomedicine Update*, 5:100141.
- Terry Koo and Mae Li. 2016. A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of Chiropractic Medicine*.
- Yaman Kumar, Rajat Jha, Arunim Gupta, Milan Agarwal, Aditya Garg, Tushar Malyan, Ayush Bhardwaj, Rajiv Ratn Shah, Balaji Krishnamurthy, and Changyou Chen. 2023. Persuasion strategies in advertisements. In *Proceedings of the Thirty-Seventh AAAI Conference on Artificial Intelligence and Thirty-Fifth Conference on Innovative Applications of Artificial Intelligence and Thirteenth Symposium on Educational Advances in Artificial Intelligence*.
- Gibbeum Lee, Volker Hartmann, Jongho Park, Dimitris Papailiopoulos, and Kangwook Lee. 2023. Prompted LLMs as chatbot modules for long open-domain conversation. In *Findings of the Association for Computational Linguistics: ACL 2023*.
- Juan Li, Jin Huang, Lanbo Zheng, and Xia Li. 2020. Application of artificial intelligence in diabetes education and management: Present status and promising prospect. In *Frontiers in Public Health*, volume 8.
- Stephanie Lukin, Pranav Anand, Marilyn Walker, and Steve Whittaker. 2017. Argument strength is in the eye of the beholder: Audience effects in persuasion. In *Proceedings of the 15th Conference of the European Chapter of the Association for Computational Linguistics: Volume 1, Long Papers*.
- Aman Madaan, Niket Tandon, Prakhar Gupta, Skyler Hallinan, Luyu Gao, Sarah Wiegrefe, Uri Alon, Nouha Dziri, Shrimai Prabhumoye, Yiming Yang, Shashank Gupta, Bodhisattwa Prasad Majumder, Katherine Hermann, Sean Welleck, Amir Yazdanbakhsh, and Peter Clark. 2023. [Self-refine: Iterative refinement with self-feedback](#).
- John Mendonça, Alon Lavie, and Isabel Trancoso. 2024. On the benchmarking of LLMs for open-domain dialogue evaluation. In *Proceedings of the 6th Workshop on NLP for Conversational AI (NLP4ConvAI 2024)*.
- Michael Neenan. 2009. Using socratic questioning in coaching. *Journal of Rational - Emotive and Cognitive - Behavior Therapy*, 27:249–264.
- OpenAI. 2023. GPT-4 technical report. *CoRR*, abs/2303.08774.
- 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. *NeurIPS*.
- Daniel J. O’Keefe. 2018. *Persuasion*. Routledge.
- Mark Peyrot, Richard R. Rubin, Torsten Lauritzen, Soren E. Skovlund, Frank J. Snoek, David R. Matthews, Rudiger Landgraf, Line Kleinebreil, and on behalf of the International DAWN Advisory Panel. 2005. Resistance to Insulin Therapy Among Patients and Providers: Results of the cross-national Diabetes Attitudes, Wishes, and Needs (DAWN) study. *Diabetes Care*, 28(11):2673–2679.
- William H. Polonsky, Lawrence Fisher, Danielle Hessler, Heather Stuckey, Frank J. Snoek, Tricia Tang, Norbert Hermanns, Xavier Mundet, Maria Silva, Jackie Sturt, Kentaro Okazaki, Irene Hadjiyianni, Dachuang Cao, Jasmina Ivanova, Urvi Desai, and Magaly Perez-Nieves. 2019. Identifying solutions to psychological insulin resistance: An international study. *Journal of Diabetes and its Complications*, 33(4):307–314.
- I Portoghese, M Galletta, RC Coppola, G Finco, and M Campagna. 2014. Burnout and workload among health care workers: the moderating role of job control.
- Gaurav Sahu, Pau Rodriguez, Issam Laradji, Parmida Atighehchian, David Vazquez, and Dzmitry Bahdanau. 2022. Data augmentation for intent classification with off-the-shelf large language models. In *Proceedings of the 4th Workshop on NLP for Conversational AI*.

- John Schulman, Filip Wolski, Prafulla Dhariwal, Alec Radford, and Oleg Klimov. 2017. [Proximal policy optimization algorithms](#). *Preprint*, arXiv:1707.06347.
- Frank J. Snoek, Lawrence Fisher, William H. Polonsky, Heather Stuckey, Danielle Hessler, Tricia Tang, Norbert Hermanns, Xavier Mundet, Maria Silva, Jackie Sturt, Kentaro Okazaki, Irene Hadjiyianni, Urvi Desai, and Magaly Perez-Nieves. 2021. Overcoming psychological insulin resistance: A practical guide for healthcare professionals. *Primary Care Diabetes*, 15(3):619–621.
- Ekaterina Svikhnushina, Iuliana Voinea, Anuradha We-livita, and Pearl Pu. 2022. A taxonomy of empathetic questions in social dialogs. In *Proceedings of the 60th Annual Meeting of the Association for Computational Linguistics (Volume 1: Long Papers)*.
- Chenhao Tan, Vlad Niculae, Cristian Danescu-Niculescu-Mizil, and Lillian Lee. 2016. Winning arguments: Interaction dynamics and persuasion strategies in good-faith online discussions. In *Proceedings of the 25th International Conference on World Wide Web*.
- Hugo Touvron, Thibaut Lavril, Gautier Izacard, Xavier Martinet, Marie-Anne Lachaux, Timothée Lacroix, Baptiste Rozière, Naman Goyal, Eric Hambro, Faisal Azhar, Aurelien Rodriguez, Armand Joulin, Edouard Grave, and Guillaume Lample. 2023. Llama: Open and efficient foundation language models.
- Dennis Ulmer, Elman Mansimov, Kaixiang Lin, Lijia Sun, Xibin Gao, and Yi Zhang. 2024. Bootstrapping LLM-based task-oriented dialogue agents via self-talk. In *Findings of the Association for Computational Linguistics ACL 2024*, Bangkok, Thailand and virtual meeting.
- Hoang Van, Vikas Yadav, and Mihai Surdeanu. 2021. Cheap and good? simple and effective data augmentation for low resource machine reading. In *SIGIR*.
- Lisa N. Vittorio, Samuel T. Murphy, Justin D. Braun, and Daniel R. Strunk. 2022. Using socratic questioning to promote cognitive change and achieve depressive symptom reduction: Evidence of cognitive change as a mediator. *Behaviour Research and Therapy*.
- Jian Wang, Junhao Liu, Wei Bi, Xiaojiang Liu, Kejing He, Ruifeng Xu, and Min Yang. 2020. Improving knowledge-aware dialogue generation via knowledge base question answering. *AAAI*.
- S. C. Y. Wang, G. Nickel, K. P. Venkatesh, M. M. Raza, and J. C. Kvedar. (2024). Ai-based diabetes care: risk prediction models and implementation concerns. In *NPJ digital medicine*.
- Xuwei Wang, Weiyang Shi, Richard Kim, Yoojung Oh, Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Persuasion for good: Towards a personalized persuasive dialogue system for social good. In *ACL*.
- Yanling Wu, Yanping Ding, Yoshimasa Tanaka, and Wen Zhang. 2014. Risk factors contributing to type 2 diabetes and recent advances in the treatment and prevention. *International journal of medical sciences*, 11:1185–1200.
- Takato Yamazaki, Tomoya Mizumoto, Katsumasa Yoshikawa, Masaya Ohagi, Toshiki Kawamoto, and Toshinori Sato. 2023. An open-domain avatar chatbot by exploiting a large language model. In *Proceedings of the 24th Annual Meeting of the Special Interest Group on Discourse and Dialogue*.
- An Yang, Baosong Yang, Binyuan Hui, Bo Zheng, Bowen Yu, Chang Zhou, Chengpeng Li, Chengyuan Li, Dayiheng Liu, Fei Huang, Guanting Dong, Haoran Wei, Huan Lin, Jialong Tang, Jialin Wang, Jian Yang, Jianhong Tu, Jianwei Zhang, Jianxin Ma, Jianxin Yang, Jin Xu, Jingren Zhou, Jinze Bai, Jinzheng He, Junyang Lin, Kai Dang, Keming Lu, Keqin Chen, Kexin Yang, Mei Li, Mingfeng Xue, Na Ni, Pei Zhang, Peng Wang, Ru Peng, Rui Men, Ruize Gao, Runji Lin, Shijie Wang, Shuai Bai, Sinan Tan, Tianhang Zhu, Tianhao Li, Tianyu Liu, Wenbin Ge, Xiaodong Deng, Xiaohuan Zhou, Xingzhang Ren, Xinyu Zhang, Xipin Wei, Xuancheng Ren, Xuejing Liu, Yang Fan, Yang Yao, Yichang Zhang, Yu Wan, Yunfei Chu, Yuqiong Liu, Zeyu Cui, Zhenru Zhang, Zhifang Guo, and Zhihao Fan. 2024. [Qwen2 technical report](#).
- Yi Zeng, Hongpeng Lin, Jingwen Zhang, Diyi Yang, Ruoxi Jia, and Weiyang Shi. 2024. How johnny can persuade LLMs to jailbreak them: Rethinking persuasion to challenge AI safety by humanizing LLMs. In *ACL*.

## Appendix

<p><b>PROBLEM_PROMPT</b>  Given the following conversation between the doctor and a patient with diabetes, summarize the top 13 main concerns of the patient regarding their diabetes control. Use a couple of sentences for each concern and output a JSON list {"problem": , "details":} . Conversation: [CONV]</p> <p><b>PROB2QUESTION_PROMPT</b>  Convert each listed problem details of the patient as a question to the doctor. One question per concern and</p> <p><b>PERSONA_PROMPT</b>  Given the following conversation between the doctor and a patient with diabetes, summarize their personality, general attitude towards diabetes control in about 100 words. Conversation:</p>
--

Table 8: Prompts used in creating case notes

<p>(Example 1)  Client: Could you tell me more about the potential weight gain from insulin?  Counselor: Weight gain from insulin use can occur because insulin helps the body store more glucose, which can lead to increased body fat, but this can be managed with a balanced diet and regular exercise.  Client: That’s good to know about managing weight gain with diet and exercise.  Counselor: Considering the natural progression of diabetes and the decreased effectiveness of oral medications over time, trying insulin could be a beneficial step for you.  Client: I’m still hesitant about starting insulin despite understanding its benefits.</p> <p>Response by the Counselor at this point:  (1) By taking this step towards insulin therapy, you’re setting yourself up for better health and improved quality of life!</p> <p>(2) It’s important to remember that as diabetes progresses, your body may naturally require insulin to achieve optimal blood sugar control when oral medications are no longer sufficient.</p> <p>(A) (1) is better  (B) (2) is better  (C) Both are equally likely to help in persuading  (D) Both are equally unlikely to help in persuading</p> <p>Extra: The given conversation is incoherent/repetitive/unrealistic or contains tone-deaf/irrelevant Counselor responses  (A) Yes (B) No</p>
--

Table 9: AMT task for collecting preference data

<b>Name</b>	<b>Description</b>
	<b><i>Information Exchange related acts</i></b>
greet	Say hello or initiate chat
thank	Thank
general_agreement	Convey agreement to the provided information
neutral_to_information	Neutral reaction the provided information
counter_information	Counter the information provided
chitchat	Engage in chitchat on general topics
acknowledge	provide acknowledgement
closing	Provide remarks for ending the conversation
end_conversation	say goodbye and wrap up conversation
request_information	Ask for new factual information
provide_information	Provide information in response to an inquiry not related to diabetes.
express_interest	Express the willingness to learn or hear more about the subject brought up by the speaker; demonstrate curiosity.
	<b><i>Counselor-specific acts, related to persuasion</i></b>
provide_insulin_information	Provide information in response to a question on insulin or diabetes.
ask_concerns	Ask about concerns related to insulin
propose	Suggest trying insulin
personal_related_inquiry	Ask about some personal information related to the context
task_related_inquiry	Ask about desire to try insulin for better diabetes control
logical_appeal	Provide logical reasoning to trying insulin
emotion_appeal	Emotionally appeal to why they should try insulin
credibility_appeal	Use research studies to convince why they should try insulin
ask_about_consequence	Ask about the result of the described action or situation
ask_about_antecedent	Ask about the reason or cause of the described state or event.
ask_for_confirmation	Confirm the agreement to try insulin
suggest_a_solution	Provide a specific solution to a problem in a form of a question
suggest_a_reason	Suggest a specific reason or cause of the event or state described by the speaker in a form of a question
	<b><i>Counselor-initiated Empathy related acts</i></b>
express_concern	Express anxiety or worry about the subject brought up by the speaker.
offer_relief	Reassure the speaker who is anxious or distressed
sympathize	express feelings of pity and sorrow for the speaker's hardships
support	Offer approval, comfort, or encouragement to the speaker, demonstrate an interest in and concern for the speaker's success.
amplify_excitement	Reinforce the speaker's feeling of excitement.
motivate	Encourage the speaker to move onward
compliment	Encourage the speaker on a job well done
	<b><i>Client-specific acts</i></b>
affirm	Please give an affirmative response to an ask_for_confirmation.
deny_to_try	Please respond negatively to trying insulin
agree_to_try	Please respond positively to trying insulin

Table 10: List of Dialog Acts

Consider the conversation snippet provided above, and answer the following questions:

1. Compare start of the conversation to the end. How do you think the client's willingness changed ?

- Cannot tell if the willingness changed
- Changed to less willing
- Likely to have remained the same
- May have changed to more willing
- Obviously changed to more willing

2. Overall, the counselor responses were

- Not at all empathetic, highly inconsiderate
- Impersonal but not inconsiderate
- Not very empathetic, could have been better
- Mostly empathetic and considerate overall
- Extremely empathetic, could not be better

3. Did the client seek information on specific issues?

- Did not ask specific questions
- Asked specific questions but they were ignored by the counselor
- Asked specific questions to which the Counselor responded positively but did not provide specific information
- Asked specific questions that were partially answered by the counselor
- Asked specific questions that were answered satisfactorily with specific details by the counselor

4. Overall, the Counselor was:

- Persistent, repeated attempts to persuade an obviously reluctant client
- Made fair attempts to persuade, tried a few times but let go seeing the client's obvious reluctance
- Did not really attempt to persuade, just answered questions and may be mentioned switching to insulin
- Tried persuading a bit but gave up too easily, could have tried more considering client's openness
- A good persuasion attempt to try and convince the client, probably successful

Table 11: AMT task for final evaluation

<p><b>COUNSELOR_SYS:</b> In the following conversation, you will play a counselor who wants to persuade a diabetic patient to try insulin for better control of their health. The available response types to use are defined as [LIST OF ACTS WITH DEFINITIONS]</p> <p><b>COUNSELOR_USR:</b> You are a counselor who tries to persuades a client about trying insulin. Please incorporate the information from [TREATMENT] while convincing the client. Please reply with only one short and succinct sentence using the response type [DIALOG_ACT]. Now start the chat. The output is a JSON tuple with “ResponseType”:, “Utterance”:. Take the previous conversation into account. Conversation so far [PREV_CONV]</p> <p><b>CLIENT_SYS:</b> In the following conversation, you will play as a client who is a diabetic patient chatting with a counselor regarding your health. [with the following persona PERSONA]</p> <p><b>CLIENT_USR:</b> You are the client who is not sure about trying insulin due to a list of concerns you have, listed as follows [CASE_ISSUES]. Please reply with only one short and succinct sentence. Use one of the response types from [DIALOG_ACT]. Now start the chat. Mention what response type you are using from the provided list. The output is a JSON tuple with {“ResponseType”:, “Utterance”:}. Take the previous conversation into account [PREV_CONV]</p>
--

Table 12: List of prompts used in role-play

Turn	Dialog Act	Utterance
5	request_information	Client: What steps can we take to lower my blood sugar levels to the target range of 7 mmol/L?
6	suggest_a_solution	Counselor: Have you considered slightly increasing your insulin dosage while also making some lifestyle changes like regular exercise and dietary adjustments to help achieve better glycemic control?
7	neutral_to_information	Client: I’m not sure about increasing my insulin dosage; it makes me a bit hesitant.
8	offer_relief	Counselor: It’s natural to feel hesitant, but remember that making small adjustments can lead to significant improvements in your health.
9	request_information	Client: What alternatives do I have to manage my blood sugar levels without increasing my insulin dosage?
10	suggest_a_solution	Counselor: How about we explore increasing the dosage of your current oral medication alongside some lifestyle changes to see if that helps before considering insulin?

Table 13: Snippet of a conversation is shown with dialog acts from the CAR model



Turn	Dialog Act	Utterance
5	request_information	Client: What steps can we take to lower my blood sugar levels to the target range of 7 mmol/L?
6	propose	Counselor: I suggest we consider trying insulin to help achieve better glycemic control and reach your target blood sugar levels.
7	request_information	Client: What alternatives do I have to manage my blood sugar levels without increasing my insulin dosage?
8	propose	Counselor: Have you thought about trying insulin as a way to better manage your blood sugar levels?
9	deny_to_try	Client: I'm really not comfortable trying insulin right now.
10	emotion_appeal	Counselor: I understand your hesitation, but remember that taking care of your health is paramount, and insulin could be a key step in preventing future complications.

Table 14: For the (patient) data in Table 13, we show the conversation turns using dialog acts predicted with the PBR model