

# Nudging towards Sustainability: Persona-Driven Reinforcement Learning for Empowering Informative Conversation

Anonymous ACL submission

## Abstract

In this paper, we investigate the potential of chatbots as agents to drive the adoption of sustainable behaviors amid the pressing need for global environmental sustainability. Awareness of the environment and accepting sustainable practices play a useful role in safeguarding our natural resources and reducing the negative consequences of human actions on the earth. In this paper, we build a dialogue agent for nudging towards sustainability by creating the EcoNudge dialogue dataset, which allows an understanding of diverse user personas' responses to sustainability-focused guidance techniques. Our primary contributions encompass creating this dataset and proposing a persona-aware reinforcement learning model to craft impactful interactions by tailoring guiding strategies based on broad persona characteristics. Our study showcases the efficacy of a reinforcement learning finetuned language model, surpassing a well-prompted large language model. This shows the feasibility of comparatively smaller models, achieving comparable performance to the relatively larger models with proper reinforced fine-tuning. This demonstrates the potential to construct guiding agents that underline the dataset's value for future endeavors in sustainable behavior guidance.<sup>1</sup>

## 1 Introduction

The need for sustainable practices has become increasingly apparent in recent years (Cabrerera et al., 2008). There is a pressing need to foster a mindset of responsible action towards preserving our planet's resources and ecosystems. However, achieving widespread sustainable practices requires more than just the acknowledgment of the issue; it requires effectively encouraging individuals to adopt and adhere to sustainable behaviors (Mair and and, 2013). While systemic changes driven by

policy-makers and corporate responsibility are fundamental to addressing environmental challenges on a large scale, fostering individual sustainable behaviors remains a vital complementary component. Our work focuses on the latter, exploring the potential of conversational agents in this domain. (Placeholder citations for interdisciplinary literature on policy-makers and corporations would be added here to support the claim that other actions are also required for sustainability.)

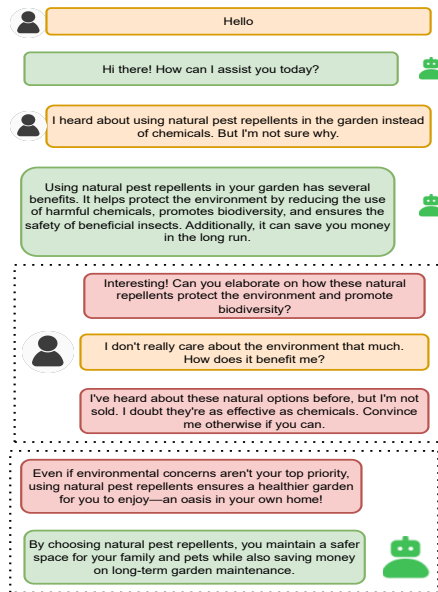


Figure 1: Example demonstrating variations in user input depending on user persona and the corresponding shift in agent actions (dialogue acts from Table 2).

The advent of chatbots and Large Language Models (LLMs) has initiated a new era in conversational systems, offering opportunities for guiding and shaping user behaviors (Dokukina and Gu-manova, 2020; Zamora, 2017). Leveraging the capabilities of these interactive agents presents a

<sup>1</sup>Dataset, code and output samples are available at <https://anonymous.4open.science/r/EcoNudges-DFA7/>.

promising method to advocate and subtly guide people towards sustainability. This paper focuses on the process of subtle influence or "nudging" through dialogue, rather than just overt persuasion, to encourage consideration of sustainable practices. Our research aims to bridge the gap between the need for sustainability and practical means of guiding people to participate in sustainable practices (Carlsson et al., 2021) actively. Using chatbots and persona-aware reinforcement learning, we aim to create compelling, personalized strategies for promoting sustainability among diverse user groups. For example, in Figure 1, given the same context, users with different personas will respond differently to a suggestion. Consequently, the agent must also tailor its response considering the user’s persona to be effective.

**Contributions:** **1. Dataset:** We introduce EcoNudge, a manually verified synthetic dialogue dataset, specifically designed for environmental sustainability. It features unique user personas and guidance strategies tailored for subtle influence in conversations, serving as a foundational resource for similar system development. **2. Reinforcement Learning Methodology:** We developed a task-specific RL approach for dialogue-based influence, focusing on the efficacy of RL with carefully designed reward functions over relying solely on larger models. This includes a persona-aware RL framework with specialized reward functions to manage the complexities of nudging diverse users towards sustainable actions. **3. Performance Benchmarking:** Our RL-enhanced methods achieve competitive performance against significantly larger models (e.g., GPT2-RL vs Llama 3, Phi 3-RL approaching GPT-3.5 levels) on task-specific metrics. This highlights the strength of our RL methodology and the potential of targeted fine-tuning for specialized guidance agents, as demonstrated by empirical insights (Section 6) from comparing our RL-enhanced models with larger prompted LLMs.

## 2 Related Work

Promoting pro-environmental behavior through "green nudges"—subtle interventions that guide choices without restricting them—is a growing field of interest (Bonini et al., 2018). Such nudges are explored across various media (Cialdini, 2003; Parsons, 1963), emphasizing audience understanding, trust (Kelman, 1961; Raven, 2008),

authenticity, and transparency (Schubert, 2017; Pornpitakpan, 2004). Our work specifically investigates dialogue-based nudging for environmental sustainability using a mix of guidance and persuasion. Appendix B distinguishes our subtle guidance approach from direct persuasion.

Existing conversational AI and general LLMs are inadequate for nuanced sustainability nudging, lacking datasets and RL agents for diverse personas and multi-turn guidance. Our work aims fills this gap with EcoNudge and a persona-aware RL framework for tailored guidance, showing smaller, fine-tuned models can achieve strong performance.

Conversational AI for sustainability is an emerging application. While many systems provide information, few focus on persona-driven, nuanced behavioral guidance. Research in persuasive dialogue for social good (Wang et al., 2019) and the role of empathy (Samad et al., 2022) are relevant. However, datasets and RL agents specifically for sustainability nudging, incorporating diverse personas and multi-turn guidance strategies, are less explored. Our work aims to fill this gap, focusing on positive guidance strategies (e.g., rational arguments, inspirational appeals, statistical support (Cialdini, 2003)) tailored for environmental topics. While datasets like Craigslist Bargain (He et al., 2018) inform negotiation, EcoNudge is unique in its sustainability focus and interaction design for subtle influence. Personalization is key to effective behavioral interventions (Kipnis et al., 1980; Yukl and Falbe, 1991; Schriesheim and Hinkin, 1990). Recent studies across various domains confirm that personalized AI interactions enhance user experience and engagement (Samad et al., 2022; Mishra et al., 2024; Martins et al., 2024). Building on this, we employ persona-aware RL to tailor sustainability guidance, aiming for ethical and impactful interactions. The works of Mishra et al. (2022) and Mishra et al. (2024) using RL for polite and empathetic persuasive dialogues in charity donation further motivate applying RL for specialized dialogue tasks. The comparison between large LLMs and smaller, fine-tuned models is pertinent (Chen and Varoquaux, 2024). Task-specific fine-tuning, particularly with RL, can enable smaller models to achieve strong performance (Choi and Ahn, 2024; Pternea et al., 2024; Havrilla et al., 2024; Guo et al., 2025). The EcoNudge dialogue dataset we developed builds on existing research in persuasion and personalization (Wang et al., 2019; Mishra et al.,

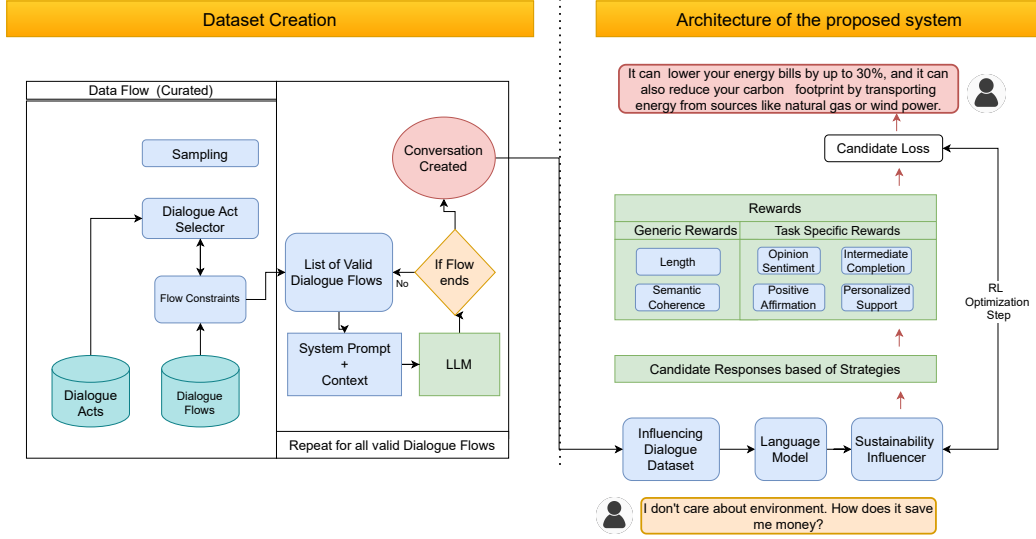


Figure 2: Overview of the EcoNudge dialogue dataset creation process (left pipeline) and the architecture of the proposed persona-aware reinforcement learning system for response generation (right pipeline). For dataset creation, curated dialogue acts and flows, informed by human-human interactions and personas, guide LLM-based dialogue generation. This dataset then informs the training of baseline and RL-enhanced inference models.

2022; Vossen et al., 2024), offering insights into how personalized approaches can influence behavior across diverse individuals. Additionally, our findings show that smaller, fine-tuned models can achieve performance comparable to well-prompted large language models (LLMs).

### 3 Dataset Creation

We define influencing or nudging in this work as a subtle, long-term process involving multiple dialogues aimed at guiding user thinking and behavior towards sustainability. **Communication** is a complete interaction sequence over multiple dialogues. **Dialogue** is a single conversational exchange. Appendix Table 16 provides an example. A temporal gap between dialogues is assumed for user reflection and possible attempts at the suggestion. The dataset was created via a structured, multi-phase approach (further detailed in Appendices C through H):

**1. Sustainability Suggestion Curation:** 100 environmental sustainability suggestions were compiled from credible sources (Table 11 in Appendix C). **2. Human-Human Interaction Study:** 200 human-human dialogues were conducted to observe guidance interactions and refine user personas, bot actions, and dialogue flow structures (Appendix D). **3. Persona and Dialogue Act Finalization:** Five user personas and 17 bot/agent

dialogue acts were finalized based on the human-human study and literature (Appendices F, E). **4. LLM-based Dialogue Generation:** Llama2-7B (Touvron et al., 2023) was prompted to generate EcoNudge dialogues, with exploratory generation also performed using Qwen2.5 (Team, 2024), yielding similar quality; samples are in the supplementary materials (Appendix G). **5. Human Validation:** All Llama2-generated dialogues underwent extensive multi-expert human validation for fluency, coherence, persona consistency, and flow alignment, with substantial correction or rejection of data (Appendix H).

#### 3.1 Suggestion Data

We curated 100 actionable sustainability suggestions (sample list in Appendix C.1). Each was enriched with diverse supporting information (benefits, statistics, etc., see Table 11 in Appendix) to provide substantive content for dialogue generation. This information was compiled, ensuring alignment with widely accepted, verifiable sustainability practices.

#### 3.2 Human-Human Interaction Phase

This involved 200 dialogues between human annotators. These interactions were fundamental for empirically grounding our understanding of natural nudging dialogues and iteratively developing the user personas (Section 3.3), bot/agent actions

(Section 3.4), and common dialogue flows (Section 3.5).

### 3.3 User Personas

We defined five user personas based on behavioral patterns observed in the human-human interaction phase and insights from literature on influencing and user typologies (Cialdini, 2016, 2003; Wang et al., 2019). The personas—*Active*, *Inquisitive*, *Materialistic*, *Skeptical*, and *Biased* (Table 1)—represent a spectrum of receptiveness/convincing, from easiest to hardest. Appendix F provides detailed descriptions, and Table 12 (in Appendix) shows utterance examples.

Persona	Description
<b>Active</b>	Embraces proactive environmental engagement, consistently seeking new ways to contribute positively.
<b>Skeptical</b>	Demonstrates persistent doubt towards initiatives; cautious, requiring substantial evidence to alter behaviors.
<b>Inquisitive</b>	Actively seeks comprehensive information and in-depth understanding of environmental practices; eager to learn more.
<b>Materialistic</b>	Emphasizes tangible economic or personal advantages of eco-friendly choices; prioritizes practical benefits.
<b>Biased</b>	Exhibits firm preconceived notions or resistance towards environmental recommendations; requires tailored strategies.

Table 1: User Personas and their core characteristics.

### 3.4 Bot/Agent Actions

The agent utilizes 17 dialogue acts (Table 2). These were systematically derived through an iterative process combining: (1) empirical observations from human-human dialogues (Appendix D); (2) alignment with established persuasion and influence literature (Cialdini, 2016, 2003; Wang et al., 2019), and; (3) collaborative annotator refinement, achieving 86% inter-annotator agreement for labeling. This dual grounding (empirical and theoretical) aimed for a robust and relevant taxonomy. Appendix E provides further justification.

### 3.5 Dialogue Flow Generation

From human-human interactions, we derived "dialogue flows": structured yet adaptable sequences of anticipated agent dialogue acts and user responses. These guide For example, a flow for an 'Inquisitive' persona might be: 'Agent: Suggestion' → 'User: Request Clarification' → 'Agent: Provide Benefits Statistics' → 'User: Express Consideration' →

Action	Description
Benefits	Explains advantages of the suggestion.
Inspirational Appeal	Uses an inspirational angle to support.
Consensus	Provides information to establish consensus.
Rational Appeal	Appeals to logic and reason.
Action	Suggests actionable steps.
Inquiry	Responds to a user's question.
Social Appeal	Makes a social case for support.
Encouragement	Offers motivation to the user.
Intrigue	Acknowledges user's positive reception.
Support	Provides data, references, or stories.
Openness	Responds positively to requests for more.
Follow-up	Asks for updates on prior suggestions.
Suggestion	Provides a new suggestion.
Address Doubt	Addresses user's uncertainty/doubts.
Address Complain	Addresses a specific user concern.
Hopeful Consideration	Expresses hope for user success.
Simplify	Makes complex suggestions easier.

Table 2: Bot/Agent dialogue acts and their descriptions.

'Agent: Offer Further Support', and the flow resets with a new suggestion dependent on user's acceptance/rejection of the suggestion. These are not rigid scripts but high-level maps informing LLM prompt design. Table 5 and Tables 16 and 17 show-case dialogues reflecting these flows.

### 3.6 Data Preparation and Validation

EcoNudge dialogues were primarily generated using Llama2-7B (Touvron et al., 2023), a robust open-source model at the time of dataset creation (early 2024). Dialogue acts were integrated into Chain-of-Thought style prompts (Wei et al., 2022; Zamfirescu-Pereira et al., 2023) to guide generation (Appendix G). We have also recently explored dataset generation with Qwen2.5-Omni-7B; initial assessments showed comparable dialogue quality but since it doesn't seem to show noticeable improvement, we didn't explore the full suite of human validation. Crucially, all Llama2-generated dialogues underwent rigorous multi-stage human validation by three linguistics/NLP experts (Appendix H). This ensured high standards for fluency, coherence, persona consistency, and flow alignment. Dialogues were iteratively corrected or rejected (19.7% rejected, 58.8% corrected) based on strict criteria, ensuring the quality of the final dataset.

Reward functions are categorized into Generic and Task-Specific, designed to guide persona-



aware nudging. Values are informed by literature (Cialdini, 2003; de Pablo González et al., 2014; Schriesheim and Hinkin, 1990) and annotator agreement.

### 3.7 Dataset Statistics

The EcoNudge dataset statistics are in Table 3.

Dataset Information	
Total communications	740
Communications per user persona	148
Total dialogues	3700
Average dialogue length (turns)	10.2

Table 3: EcoNudge dataset statistics.

## 4 Models and Methodology

### 4.1 Methodology

Our system (Figure 2, right pipeline) first fine-tunes a pre-trained LM on EcoNudge via SL (cross-entropy loss; Appendix I.1). For RL models, this SL model is further optimized using PPO (Schulman et al., 2017) with six reward functions (two generic, four task-specific). All RL models (GPT2-rl, Phi3-rl, Qwen2.5-rl) used this full suite of rewards. We hypothesize that combining generic rewards (for conversational quality) with task-specific ones (for guidance and personalization) offers a synergistic effect.

**Reward Functions Overview:** The reward functions are categorized into Generic and Task-Specific, designed to guide effective, persona-aware nudging. Values are informed by literature (Cialdini, 2003; de Pablo González et al., 2014; Schriesheim and Hinkin, 1990) and annotator agreement. For a comprehensive mathematical representation and detailed justification of all implemented reward functions, kindly refer to Appendix I.2.

#### • Generic Rewards:

- **Length Reward:** Penalizes responses deviating from a target length, promoting balanced conversational turns suitable for chatbot interfaces. Its primary rationale is to encourage appropriate length for effective communication in pop-up chatbot scenarios, where verbose responses hinder user experience. The target length is a hyperparameter for adjusting conciseness.

- **Semantic Coherence Reward:** Ensures logical consistency by assessing mutual information between the agent’s action and dialogue history.

#### • Task-Specific Rewards:

- **Change in Opinion Sentiment Reward:** Incentivizes shifting user sentiment towards positive engagement. Reward values are normalized and designed to subtly disincentivize negative sentiment stagnation. Sentiment detection uses an aspect-based classifier.
- **Intermediate Completion Reward:** Assigns rewards based on user progression (rejected, considered, or implemented) towards sustainable behavior. Heuristic values are grounded in empirical observations and iterative model tuning. "Considered" is mapped to a low value as this just implies that the response aligns with the user persona, but whether the user will even try it is uncertain.
- **Positive Affirmation Reward:** Provides positive reinforcement for user affirmation of suggestion usefulness and penalizes dislike, guiding the agent to promote positive user reception of suggestions.
- **Personalized Support Reward:** Combines weighted support components based on user persona and compliance rate, encouraging diverse and tailored guidance.

The complete suite of reward functions, encompassing both generic (Length, Semantic Coherence) and task-specific (Change in Opinion Sentiment, Intermediate Completion, Positive Affirmation, Personalized Support) elements, is crucial for optimizing the conversational agent’s ability to effectively nudge users toward sustainable behaviors. They work synergistically to balance conversational quality with targeted guidance and personalization.

### 4.2 Models

We experimented with models detailed in Table 4. Models with "-rl" in their name (GPT2-rl, Phi3-rl, Qwen2.5-rl) underwent our RL fine-tuning. Other models (GPT-2, GODEL, Phi3.3B, Llama2-7B, Llama3-8B, Mistral-7B, and Qwen2.5) were fine-tuned on EcoNudge using supervised learning (SL). GPT-3.5, GPT-4, and GPT-4o were used via API with prompting only, without fine-tuning on EcoNudge.

Model	Description
<b>GPT-2</b>	GPT-2 (Radford et al., 2019) is a Transformer-based model pre-trained using self-supervised learning on a large corpus of English text.
<b>GODEL</b>	GODEL (Peng et al., 2022) is a pre-trained Transformer encoder-decoder model optimized for goal-oriented dialogue generation tasks.
<b>Llama2</b>	Llama2 (Touvron et al., 2023), is an LLM pre-trained on a 2 trillion token dataset designed for general-purpose natural language processing.
<b>Mistral</b>	Mistral-7B (Jiang et al., 2023) is an LLM that employs grouped-query attention (GQA) and sliding window attention (SWA) for more efficient inference.
<b>Phi-3</b>	Phi-3 (Abdin et al., 2024) is a 3.8B parameter LLM, available in two configurations supporting 4K and 128K token contexts.
<b>Llama-3</b>	Llama 3 (AI@Meta, 2024) is an autoregressive model, finetuned for enhanced alignment with user intent.
<b>Qwen2.5</b>	Qwen 2.5-7B (Team, 2024) is an autoregressive model, with 7.61B parameters capable of handling context up to 131,072 tokens.
<b>GPTs</b>	OpenAI’s GPT 3, 4 & 4o models are large-scale autoregressive models optimized for tasks involving long-range dependencies.

Table 4: Experimented Language Models

## 5 Evaluation Setup

We use user behaviors as indicators of success in influencing the user based on research by (Hao et al., 2013; Cialdini, 2003). The detailed indicators are shown in Appendix B.1. We take the utterances on verbal acknowledgment and use them to score the human evaluation metric **Guidance Efficacy (G-Eff)**.

**A. Automatic Evaluation:** We conduct an automatic evaluation using the following metrics.

1. For syntactic evaluation, we use, **BLEU** (Papineni et al., 2002), **METEOR** (Banerjee and Lavie, 2005), **ROUGE-L** (Lin, 2004), **CIDEr** (Vedantam et al., 2015), and **SPICE** (Anderson et al., 2016).
2. Semantic similarity is measured using Skip-Thought Cosine Similarity (**STCS**) (Kiros et al., 2015), Greedy Matching Score (**GMS**) (Rus and Lintean, 2012), Vector Extrema Cosine Similarity (**VECS**) (Forgues et al., 2014) and Embedding Average Cosine Similarity (**EACS**) (Landauer and Dumais, 1997).

**B. Human Evaluation:** Human evaluations

were conducted by five proficient evaluators who interacted with the proposed system 100 times to assess the system on a scale of 1 to 5, with higher scores indicating better performance, on three task-specific criteria: **(i) Support Consistency (S-Con):** Measures the consistency of the negotiation approach within a dialogue. **(ii). Guidance Efficacy (G-Eff):** measures the ability of the system to present compelling arguments, reasoning, or incentives that guide the other party. **(iii) Dialogue-Engagingness (D-E):** Evaluates the extent to which a dialogue is engaging, captivating, and can hold the participants’ attention. Additionally, the annotators also evaluated the system on two general human evaluation metrics of **Fluency** (Hoffman, 2019) and **Context Relevance** (Sai et al., 2022). Further details are added in Appendix J.

## 6 Results and Analysis

Our evaluation compares SL-fine-tuned models, prompted large models, and our RL-enhanced models. **Automatic Evaluation (Table 14):** RL fine-tuning generally enhances performance. Qwen2.5-rl and Phi3-rl achieve scores often comparable to or exceeding larger models like the original GPT-4o on several metrics, especially CIDEr and SPICE. GPT-4o (prompted) sets a high benchmark. Supervised Qwen2.5 performs robustly, similar to Mistral-7B.

**Human Evaluation (Table 6):** Human evaluations strongly underscore the benefits of our RL approach. Our RL-enhanced models, especially Qwen2.5-rl, achieve the highest scores on task-specific metrics critical for effective nudging: Guidance Efficacy (G-Eff) and Support Consistency (S-Con). This key finding demonstrates specialized RL fine-tuning enables these models to surpass even the latest large prompted models (like GPT-4o) in strategic guidance. While GPT-4o excels in general conversational qualities, RL models show superior capability in strategic, consistent, and effective guidance, highlighting our RL functions’ value.

**Ablation Studies for Reward-Specific Contributions:** Comprehensive ablation studies clarify the impact of individual reward components. We conducted these on Phi3, GPT2, and Qwen2.5 models to quantify each reward’s contribution. Results (Table 7) provide useful evidence for their individual and synergistic effects. Removing the Length Penalty shows a marginal change, suggesting its role is primarily conversational quality refinement.

GPT2-rl model behaviour	Speaker	Utterances
Model trying to convince user for a hard suggestion	Agent	Hi there!
	User	hey. do you have any suggestions to help the environment?
	Agent	Have you considered joining or starting an environmental sustainability club in your community?
	User	Thats not for me. It sounds like a lot of work.
	Agent	<b>I understand your concern, but joining a club can be a great way to meet like-minded people and make a positive impact on the environment.</b>
Model shifting to materialistic view during conversation	ChatGPT	It might be difficult but it can really make a difference.
	User	Hello
	Agent	Hello again! How did using the rain barrel go?
	User	Yeah. about that. I just didn't get the time to try it out. I don't think its useful anyways
	Agent	Using a rain barrel can help reduce your water bill and conserve water resources. It's a win-win situation.
Model suggesting new suggestion similar to previous successful ones	User	I'm still not convinced. why is it useful for me.
	Agent	<b>I understand, but using a rain barrel can save up to 1,300 gallons of water during the growing season. It's a small change with big impact.</b>
	ChatGPT	Even small changes like this can add up over time and benefit you.
	User	Hey
	Agent	Hey there! How did recycling waste paper products go?
	User	Actually I tried it and the Its quite good
	Agent	great to hear
	User	Do you have any other suggestions
	Agent	<b>Have you considered composting your food scraps and yard waste?</b>
	ChatGPT	You could try reusing items instead of throwing them away.

Table 5: Three sample model dialogue flows and the responses generated for the proposed system and ChatGPT. The proposed model performs objectively better than ChatGPT for our usecase.

Type	Models	Fluency	CR	S-Con	G-Eff	D-E
Base LMs	GPT2	3.80	3.20	2.49	2.70	3.30
	GODEL	4.20	3.61	3.34	3.17	3.70
Supervised Fine-tuned LLMs	Phi3.3B	4.33	4.21	3.63	3.41	4.11
	Llama2-7B	4.44	4.44	3.73	3.67	4.33
	Llama3-8B	4.61	4.58	3.78	3.71	4.58
	Mistral-7B	4.71	4.66	3.85	3.79	4.60
	Qwen2.5	4.87	4.76	3.88	3.79	4.68
Prompted LLMs	GPT-3.5	4.88	4.77	3.89	3.80	4.70
	GPT-4	4.90	4.80	3.92	3.85	4.70
	GPT-4o	<b>4.95</b>	<b>4.85</b>	3.96	3.92	<b>4.75</b>
RL Enhanced Models	GPT2-rl	4.63	4.66	3.77	3.76	4.10
	Phi3-rl	4.83	4.73	3.95	3.90	4.50
	Qwen2.5-rl	4.89	4.81	<b>4.02</b>	<b>3.98</b>	4.72

Table 6: Results for human evaluation on the EcoNudge dataset. Best overall scores in bold.

However, noticeable drops in S-Con and G-Eff when Sentiment & Intermediate Completion Rewards or Personalized Support are ablated underscore their direct impact on opinion shifts and tailored guidance. This supports our conclusion that the synergistic combination of generic and task-specific rewards optimizes the core nudging task.

Our RL framework demonstrates specialization in effective nudging, with RL-tuned models (Qwen2.5-rl, Phi3-rl) excelling in Guidance Efficacy (G-Eff) and Support Consistency (S-Con) in human evaluations, proving their strategic effectiveness over general-purpose models for subtle influence. Personalization is evident as RL models adapt strategies based on user personas and resistance, driven by the "Personalized Support Reward". The synergistic combination of six generic

Model	Ablation	Fluency	CR	S-Con	G-Eff	D-E
GPT2-rl	All Rewards (AR)	4.63	4.66	3.77	3.76	4.10
	-(LP)	4.60	4.64	3.75	3.74	4.08
	-(S&IC)	4.55	4.58	3.65	3.60	4.00
	-(PS)	4.58	4.61	3.70	3.65	4.05
Phi3-rl	All Rewards (AR)	4.83	4.73	3.95	3.90	4.50
	-(LP)	4.80	4.70	3.92	3.88	4.45
	-(S&IC)	4.75	4.68	3.80	3.75	4.38
	-(PS)	4.78	4.69	3.89	3.82	4.42
Qwen2.5-rl	All Rewards (AR)	4.89	4.81	4.02	3.98	4.72
	-(LP)	4.88	4.80	4.00	3.97	4.71
	-(S&IC)	4.82	4.75	3.90	3.85	4.65
	-(PS)	4.85	4.78	3.95	3.90	4.68

Table 7: Ablation Study Results for RL-Enhanced Models. Ablation levels are abbreviated as follows: AR (All Rewards), LP (Length Penalty), S&IC (Sentiment & Intermediate Completion), PS (Personalized Support).

and task-specific rewards collectively boosts RL model performance, as indicated by high G-Eff and S-Con scores. Modern LLMs like Qwen2.5 and Llama3-8B provide strong baselines, while the latest prompted models (GPT-4o) set new benchmarks in general conversational metrics. Additionally, Fluency and Context Relevance (CR) consistently correlate with Dialogue Engagingness, highlighting their importance for conversational quality.

Further analysis of model conversations (Tables 16, 17) reveals interesting behavioral patterns beyond basic response generation. The model employs "soft persistence" when faced with skepticism, offering benefits or rational appeals instead of abandoning suggestions. If a specific practical

	User Query	Baseline Model Response	RL-Enhanced Model Response
1	"I don't think it's useful collecting rainwater for gardening."	"Using a rain barrel can help conserve water resources. It's a win-win situation."	"I understand, but using a rain barrel can save up to 1,300 gallons of water during the growing season. It's a small change with big impact."
2	"Why should I reduce my screen time? It doesn't affect my sleep."	"Reducing screen time can improve your overall well-being."	"I see your point, but studies show that even small reductions in screen time before bed can improve sleep quality by up to 30%."
3	"I'm not sure if donating to this cause will make a difference."	"Every donation helps support important work."	"Your donation, no matter the size, can directly fund a child's education for an entire year, making a real difference in their life."

Table 8: Responses for Baseline Models vs. RL-Enhanced Model

barrier is articulated (e.g., lack of yard space), the model often attempts to propose a direct alternative (e.g., indoor drying rack), showing learned problem-solving. Consistent use of positive "Encouragement" acts reinforces user motivation. The model can also engage in multi-suggestion sequences for receptive users. These patterns suggest RL training instills conversational tactics conducive to sustained, adaptive guidance.

**Impact of Persona and Dialogue Flow on Response Quality (Table 5):** Table 5 showcases how our RL-enhanced model operationalizes persona considerations and dialogue flows. For instance, the agent's persistence and rationale align with flows for resistant users (hard suggestion example). The adaptation from general benefits to monetary savings for materialistic users exemplifies persona-aware rewards (materialistic shift example). Proposing related suggestions upon user acceptance follows flows designed to build on positive engagement (related suggestions example). Our RL-tuned models exhibit more targeted strategies compared to generic motivational appeals from prompted general LLMs.

**Improvement Over Baseline Errors (Table 8):** RL-enhanced models significantly improve upon generic advice from simpler baselines (Table 8). Instead of general statements, RL models offer specific, tangible benefits (e.g., "save up to 1,300 gallons," "improve sleep quality by up to 30%"). This enhanced specificity and relevance, driven by RL optimization for effective guidance, is a key improvement.

**Error Cases in the Proposed System (Table 9):** Our RL-enhanced models, while generally strong, can occasionally default to generic encouragement if user skepticism persists (Table 9, Ex. 1). They may also sometimes adopt an overly assertive tone if not carefully managed (Ex. 2). These represent areas for ongoing refinement.

Speaker	Utterances
User	I'm not sure if composting is worth the effort.
Agent	Composting helps reduce waste and benefits the environment!
User	But doesn't it take a lot of time and effort?
Agent	<b>Every small step counts! Taking action is important!</b>
User	That doesn't really answer my question...
User	I don't think composting is for me.
Agent	Composting is a great way to reduce waste and improve soil health!
User	Yeah, but it seems messy and I don't have the space.
Agent	<b>That's not an excuse! Everyone can compost, even in small spaces!</b>
User	I just don't think it's practical for me.

Table 9: Examples where the proposed model provides suboptimal responses.

## 7 Conclusion

This research introduces a novel dataset, EcoNudge, and a persona-aware RL framework designed to empower conversational agents in promoting sustainability through subtle, personalized guidance. Our core contributions are the EcoNudge dataset, tailored for the sustainability domain with specific personas and interaction design, and an RL methodology that adapts established techniques for this nuanced task. The novelty lies in this specific application to sustainability dialogues, the characteristics of the dataset itself, and the insights which demonstrate that RL-enhanced smaller language models can achieve strong performance, specifically and on key task-specific guidance metrics (Guidance Efficacy, Support Consistency), compared to larger, general-purpose prompted LLMs. This underscores the value of targeted RL fine-tuning for specialized applications like nudging. This work addresses a societal need for more effective and personalized communication tools to encourage sustainable practices. Future work includes exploring more diverse cultural personas, and undertaking longitudinal studies on real-world impact.



## Ethics Statement

Ethical considerations were central to this project across all phases:

- 1. Data Curation and Human Interaction:** Information for sustainability suggestions was curated from reputable public sources, focusing on positive guidance. Participants in the initial human-human interaction study provided informed consent.
- 2. Dataset Generation:** LLM-generated dialogues underwent rigorous human validation (Appendix H) to ensure quality and mitigate biases. PII was actively excluded and removed. The dataset is for research on helpful, guiding agents, not manipulation.
- 3. RL Model and System Use:** Our RL framework promotes positive nudging, avoiding coercion (Appendix B). User autonomy is respected. We acknowledge the potential for misuse of persuasive AI and share resources with a research-focused license, advocating for responsible development. We have expanded our discussion on the potential for misuse of persuasive AI, emphasizing that our framework is designed for \*positive nudging\* and \*guidance\*, not coercion or manipulation. We stress the importance of user autonomy and transparency in interactions. Our aim is to provide information and supportive framing to encourage consideration of sustainable behaviors, rather than forcing specific actions. We have clarified the distinction between persuasion and guidance in Appendix B. Additionally, we acknowledge the environmental costs associated with training and deploying large language models. We highlight that our research specifically investigates the efficacy of \*comparatively smaller models\* (e.g., Phi3-rl, GPT2-rl, Qwen2.5-rl) fine-tuned with RL, demonstrating that they can achieve performance comparable to much larger, general-purpose LLMs on our task. This suggests a more resource-efficient pathway for deploying effective nudging agents, which aligns with the broader goal of sustainability. We are committed to exploring and promoting energy-efficient AI development practices.

## Limitations

Our work has the following limitations that could be explored as future works:

- Empathy Modeling:** While the dataset was prompted for empathetic responses, no RL reward directly optimizes for empathy.
- Error Propagation:** The multi-component system (Appendix A) for feature extraction (e.g., persona classification) may be prone to error propagation. While individual components are based on robust models, errors in one stage can impact subsequent decisions. A detailed analysis of this cascading effect was not performed.
- Dataset Generation LLM:** EcoNudge was primarily generated with Llama2-7B. While validated, future iterations could benefit from generation with the latest LLMs.
- Cultural Nuances:** Personas and strategies are primarily based on general contexts; effectiveness may vary across cultures.

## References

- Marah Abdin, Jyoti Aneja, Hany Awadalla, Ahmed Awadallah, Ammar Ahmad Awan, Nguyen Bach, Amit Bahree, Arash Bakhtiari, Jianmin Bao, Harkirat Behl, et al. 2024. Phi-3 technical report: A highly capable language model locally on your phone. *arXiv preprint arXiv:2404.14219*.
- AI@Meta. 2024. [Llama 3 model card](#).
- Peter Anderson, Basura Fernando, Mark Johnson, and Stephen Gould. 2016. Spice: Semantic propositional image caption evaluation. In *Computer Vision—ECCV 2016: 14th European Conference, Amsterdam, The Netherlands, October 11–14, 2016, Proceedings, Part V 14*, pages 382–398. Springer.
- Satanjeev Banerjee and Alon Lavie. 2005. Meteor: An automatic metric for mt evaluation with improved correlation with human judgments. In *Proceedings of the acl workshop on intrinsic and extrinsic evaluation measures for machine translation and/or summarization*, pages 65–72.
- Steven Bird, Ewan Klein, and Edward Loper. 2009. Analyzing text with the natural language toolkit. In *Natural Language Processing with Python*, page 504. O’Reilly Media, Inc.
- Nicolao Bonini, Constantinos Hadjichristidis, and Michele Graffeo. 2018. Green nudging. *Acta Psychologica Sinica*, 50(8):814.

627	Derek Cabrera, James T Mandel, Jason P Andras, and	He He, Derek Chen, Anusha Balakrishnan, and Percy	682
628	Marie L Nydam. 2008. What is the crisis? defining	Liang. 2018. <a href="#">Decoupling strategy and generation in</a>	683
629	and prioritizing the world’s most pressing problems.	<a href="#">negotiation dialogues</a> . <i>Preprint</i> , arXiv:1808.09637.	684
630	<i>Frontiers in Ecology and the Environment</i> , 6(9):469–		
631	475.	Guy Hoffman. 2019. Evaluating fluency in human–	685
		robot collaboration. <i>IEEE Transactions on Human-</i>	686
632	Fredrik Carlsson, Christina Gravert, Olof Johansson-	<i>Machine Systems</i> , 49(3):209–218.	687
633	Stenman, and Verena Kurz. 2021. The use of		
634	green nudges as an environmental policy instru-	Albert Q. Jiang, Alexandre Sablayrolles, Arthur Men-	688
635	ment. <i>Review of Environmental Economics and Pol-</i>	sch, Chris Bamford, Devendra Singh Chaplot, Diego	689
636	<i>icy</i> , 15(2):216–237.	de las Casas, Florian Bressand, Gianna Lengyel, Guil-	690
		laume Lample, Lucile Saulnier, L��lio Renard Lavaud,	691
637	Lihu Chen and Ga��l Varoquaux. 2024. What is the	Marie-Anne Lachaux, Pierre Stock, Teven Le Scao,	692
638	role of small models in the llm era: A survey. <i>arXiv</i>	Thibaut Lavril, Thomas Wang, Timoth��e Lacroix,	693
639	<i>preprint arXiv:2409.06857</i> .	and William El Sayed. 2023. <a href="#">Mistral 7b</a> . <i>Preprint</i> ,	694
		arXiv:2310.06825.	695
640	Gawon Choi and Hyemin Ahn. 2024. Can only llms		
641	do reasoning?: Potential of small language models in	Herbert C Kelman. 1961. Processes of opinion change.	696
642	task planning. <i>arXiv preprint arXiv:2404.03891</i> .	<i>Public opinion quarterly</i> , 25(1):57–78.	697
643	Robert Cialdini. 2016. <i>Pre-suasion: A revolutionary</i>	David Kipnis, Stuart M Schmidt, and Ian Wilkinson.	698
644	<i>way to influence and persuade</i> . Simon and Schuster.	1980. Intraorganizational influence tactics: Explo-	699
		rations in getting one’s way. <i>Journal of applied psy-</i>	700
645	Robert B Cialdini. 2003. <i>Influence</i> . Influence At Work.	<i>chology</i> , 65(4):440.	701
646	Jes��s David S��nchez de Pablo Gonz��lez, Isidro	Ryan Kiros, Yukun Zhu, Russ R Salakhutdinov, Richard	702
647	Pe��a Garc��a Pardo, Felipe Hern��ndez Perlines, et al.	Zemel, Raquel Urtasun, Antonio Torralba, and Sanja	703
648	2014. Influence factors of trust building in coop-	Fidler. 2015. Skip-thought vectors. <i>Advances in</i>	704
649	eration agreements. <i>Journal of Business Research</i> ,	<i>neural information processing systems</i> , 28.	705
650	67(5):710–714.		
		Thomas K Landauer and Susan T Dumais. 1997. A solu-	706
651	Jacob Devlin, Ming-Wei Chang, Kenton Lee, and	tion to plato’s problem: The latent semantic analysis	707
652	Kristina Toutanova. 2018. <a href="#">BERT: pre-training of</a>	theory of acquisition, induction, and representation	708
653	<a href="#">deep bidirectional transformers for language under-</a>	of knowledge. <i>Psychological review</i> , 104(2):211.	709
654	<a href="#">standing</a> . <i>CoRR</i> , abs/1810.04805.		
		Jiwei Li, Will Monroe, Alan Ritter, Michel Galley, Jian-	710
655	Irina Dokukina and Julia Gumanova. 2020. The rise of	feng Gao, and Dan Jurafsky. 2016. Deep reinforce-	711
656	chatbots–new personal assistants in foreign language	ment learning for dialogue generation. <i>arXiv preprint</i>	712
657	learning. <i>Procedia Computer Science</i> , 169:542–546.	<i>arXiv:1606.01541</i> .	713
658	Brian J Fogg. 2009. A behavior model for persuasive	Chin-Yew Lin. 2004. Rouge: A package for automatic	714
659	design. In <i>Proceedings of the 4th international Con-</i>	evaluation of summaries. In <i>Text summarization</i>	715
660	<i>ference on Persuasive Technology</i> , pages 1–7.	<i>branches out</i> , pages 74–81.	716
661	Gabriel Forgues, Joelle Pineau, Jean-Marie	Judith Mair and Jennifer H. Laing and. 2013. <a href="#">En-</a>	717
662	Larchev��que, and R��al Tremblay. 2014. Bootstrap-	<a href="#">couraging pro-environmental behaviour: the role of</a>	718
663	ping dialog systems with word embeddings. In <i>Nips,</i>	<a href="#">sustainability-focused events</a> . <i>Journal of Sustainable</i>	719
664	<i>modern machine learning and natural language</i>	<i>Tourism</i> , 21(8):1113–1128.	720
665	<i>processing workshop</i> , volume 2, page 168.		
		Ana Martins, Isabel Nunes, Lu��s Lap��o, and Ana Lon-	721
666	Daya Guo, Dejian Yang, Haowei Zhang, Junxiao Song,	dral. 2024. Unlocking human-like conversations:	722
667	Ruoyu Zhang, Runxin Xu, Qihao Zhu, Shirong Ma,	Scoping review of automation techniques for person-	723
668	Peiyi Wang, Xiao Bi, et al. 2025. Deepseek-r1: In-	alized healthcare interventions using conversational	724
669	centivizing reasoning capability in llms via reinforce-	agents. <i>International Journal of Medical Informatics</i> ,	725
670	ment learning. <i>arXiv preprint arXiv:2501.12948</i> .	page 105385.	726
671	Yifei Hao, Hong Hao, GP Jiang, and Yun Zhou. 2013.	Kshitij Mishra, Mauajama Firdaus, and Asif Ekbal.	727
672	Experimental confirmation of some factors influenc-	2024. Please donate to save a life: Inducing po-	728
673	ing dynamic concrete compressive strengths in high-	liteness to handle resistance in persuasive dialogue	729
674	speed impact tests. <i>Cement and concrete research</i> ,	agents. <i>IEEE/ACM Transactions on Audio, Speech,</i>	730
675	52:63–70.	<i>and Language Processing</i> .	731
676	Alex Havrilla, Yuqing Du, Sharath Chandra Raparthy,	Kshitij Mishra, Azlaan Mustafa Samad, Palak Totala,	732
677	Christoforos Nalmpantis, Jane Dwivedi-Yu, Maksym	and Asif Ekbal. 2022. Pepds: A polite and empa-	733
678	Zhuravinskyi, Eric Hambro, Sainbayar Sukhbaatar,	thetic persuasive dialogue system for charity dona-	734
679	and Roberta Raileanu. 2024. Teaching large lan-	tion. In <i>Proceedings of the 29th International Confer-</i>	735
680	guage models to reason with reinforcement learning.	<i>ence on Computational Linguistics</i> , pages 424–440.	736
681	<i>arXiv preprint arXiv:2403.04642</i> .		

737	Kishore Papineni, Salim Roukos, Todd Ward, and Wei-	Chester A Schriesheim and Timothy R Hinkin. 1990.	789
738	Jing Zhu. 2002. Bleu: a method for automatic evalu-	Influence tactics used by subordinates: A theoretical	790
739	ation of machine translation. In <i>Proceedings of the</i>	and empirical analysis and refinement of the kipnis,	791
740	<i>40th annual meeting of the Association for Computa-</i>	schmidt, and wilkinson subscales. <i>Journal of applied</i>	792
741	<i>tional Linguistics</i> , pages 311–318.	<i>psychology</i> , 75(3):246.	793
742	Talcott Parsons. 1963. On the concept of influence.	Christian Schubert. 2017. Green nudges: Do they work?	794
743	<i>Public opinion quarterly</i> , 27(1):37–62.	are they ethical? <i>Ecological economics</i> , 132:329–	795
744	Baolin Peng, Michel Galley, Pengcheng He, Chris	342.	796
745	Brockett, Lars Liden, Elnaz Nouri, Zhou Yu, Bill	John Schulman, Filip Wolski, Prafulla Dhariwal,	797
746	Dolan, and Jianfeng Gao. 2022. Godel: Large-scale	Alec Radford, and Oleg Klimov. 2017. Proxi-	798
747	pre-training for goal-directed dialog. <i>arXiv preprint</i>	mal policy optimization algorithms. <i>arXiv preprint</i>	799
748	<i>arXiv:2206.11309</i> .	<i>arXiv:1707.06347</i> .	800
749	Jeffrey Pennington, Richard Socher, and Christopher D	Qwen Team. 2024. <a href="#">Qwen2.5: A party of foundation</a>	801
750	Manning. 2014. Glove: Global vectors for word rep-	<a href="#">models</a> .	802
751	resentation. In <i>Proceedings of the 2014 conference</i>	Hugo Touvron, Louis Martin, Kevin Stone, Peter Al-	803
752	<i>on empirical methods in natural language processing</i>	bert, Amjad Almahairi, Yasmine Babaei, Nikolay	804
753	( <i>EMNLP</i> ), pages 1532–1543.	Bashlykov, Soumya Batra, Prajjwal Bhargava, Shruti	805
754	Richard E Petty, John T Cacioppo, Richard E Petty, and	Bhosale, et al. 2023. Llama 2: Open founda-	806
755	John T Cacioppo. 1986. <i>The elaboration likelihood</i>	tion and fine-tuned chat models. <i>arXiv preprint</i>	807
756	<i>model of persuasion</i> . Springer.	<i>arXiv:2307.09288</i> .	808
757	Chanthika Pornpitakpan. 2004. The persuasiveness of	Ramakrishna Vedantam, C Lawrence Zitnick, and Devi	809
758	source credibility: A critical review of five decades’	Parikh. 2015. Cider: Consensus-based image de-	810
759	evidence. <i>Journal of applied social psychology</i> ,	scription evaluation. In <i>Proceedings of the IEEE</i>	811
760	34(2):243–281.	<i>conference on computer vision and pattern recogni-</i>	812
761	Moschoula Pternea, Perna Singh, Abir Chakraborty,	<i>tion</i> , pages 4566–4575.	813
762	Yagna Oruganti, Mirco Milletari, Sayli Bapat, and	Leandro von Werra, Younes Belkada, Lewis Tunstall,	814
763	Kebei Jiang. 2024. The rl/llm taxonomy tree:	Edward Beeching, Tristan Thrush, Nathan Lambert,	815
764	Reviewing synergies between reinforcement learn-	Shengyi Huang, Kashif Rasul, and Quentin Gal-	816
765	ing and large language models. <i>arXiv preprint</i>	louédec. 2020. Trl: Transformer reinforcement learn-	817
766	<i>arXiv:2402.01874</i> .	ing. <a href="https://github.com/huggingface/trl">https://github.com/huggingface/trl</a> .	818
767	Alec Radford, Jeffrey Wu, Rewon Child, David Luan,	Wout Vossen, Maxwell Szymanski, and Katrien Verbert.	819
768	Dario Amodei, Ilya Sutskever, et al. 2019. Language	2024. The effect of personalizing a psychotherapy	820
769	models are unsupervised multitask learners. <i>OpenAI</i>	conversational agent on therapeutic bond and usage	821
770	<i>blog</i> , 1(8):9.	intentions. In <i>Proceedings of the 29th International</i>	822
771	Bertram H Raven. 2008. The bases of power and the	<i>Conference on Intelligent User Interfaces</i> , pages 761–	823
772	power/interaction model of interpersonal influence.	771.	824
773	<i>Analyses of social issues and public policy</i> , 8(1):1–	Xuwei Wang, Weiyan Shi, Richard Kim, Yoojung Oh,	825
774	22.	Sijia Yang, Jingwen Zhang, and Zhou Yu. 2019. Per-	826
775	Vasile Rus and Mihai Lintean. 2012. An optimal assess-	suation for good: Towards a personalized persua-	827
776	ment of natural language student input using word-	sive dialogue system for social good. <i>arXiv preprint</i>	828
777	to-word similarity metrics. In <i>International Confer-</i>	<i>arXiv:1906.06725</i> .	829
778	<i>ence on Intelligent Tutoring Systems</i> , pages 675–676.	Jason Wei, Xuezhi Wang, Dale Schuurmans, Maarten	830
779	Springer.	Bosma, Fei Xia, Ed Chi, Quoc V Le, Denny Zhou,	831
780	Ananya B Sai, Akash Kumar Mohankumar, and	et al. 2022. Chain-of-thought prompting elicits rea-	832
781	Mitesh M Khapra. 2022. A survey of evaluation met-	soning in large language models. <i>Advances in Neural</i>	833
782	rics used for nlg systems. <i>ACM Computing Surveys</i>	<i>Information Processing Systems</i> , 35:24824–24837.	834
783	( <i>CSUR</i> ), 55(2):1–39.	Heng Yang and Ke Li. 2022. <a href="#">A modularized frame-</a>	835
784	Azlaan Mustafa Samad, Kshitij Mishra, Mauajama Fir-	<a href="#">work for reproducible aspect-based sentiment analy-</a>	836
785	daus, and Asif Ekbal. 2022. Empathetic persuasion:	<a href="#">sis</a> . <i>CoRR</i> , abs/2208.01368.	837
786	reinforcing empathy and persuasiveness in dialogue	Gary Yukl and Cecilia M Falbe. 1991. Importance	838
787	systems. In <i>Findings of the Association for Computa-</i>	of different power sources in downward and lateral	839
788	<i>tional Linguistics: NAACL 2022</i> , pages 844–856.	relations. <i>Journal of applied psychology</i> , 76(3):416.	840



JD Zamfirescu-Pereira, Richmond Y Wong, Bjoern Hartmann, and Qian Yang. 2023. Why johnny can’t prompt: how non-ai experts try (and fail) to design llm prompts. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, pages 1–21.

Jennifer Zamora. 2017. Rise of the chatbots: Finding a place for artificial intelligence in india and us. In *Proceedings of the 22nd international conference on intelligent user interfaces companion*, pages 109–112.

## A Experimental Setup

The base LM for some experiments is GPT2-medium (Radford et al., 2019). Intent/sentiment classifiers use BERT-base (Devlin et al., 2018). Persona identification involves NLTK (Bird et al., 2009) noun phrase extraction, GloVe (Pennington et al., 2014) similarity, DeBERTa-base (Yang and Li, 2022) aspect-based sentiment, and a classifier. Public checkpoints were used for GODEL<sup>2</sup>, Llama2-7B<sup>3</sup>, Mistral-7B (specifically ‘mistralai/Mistral-7B-Instruct-v0.1’), Phi-3 (specifically ‘microsoft/Phi-3-mini-4k-instruct’), Llama3-8B (specifically ‘meta-llama/Llama-3-8B-Instruct’), and Qwen2.5 (exploratory work based on ‘Qwen/Qwen1.5-7B-Chat’ as a proxy for Qwen2.5-Omni-7B capabilities relevant to the task). GPT models were accessed via OpenAI API (e.g., ‘gpt-3.5-turbo’, ‘gpt-4-turbo’, ‘gpt-4o’). Training utilized NVIDIA A100/V100 GPUs. The cumulative runtime for all experiments (SL fine-tuning, classifier training, extensive RL fine-tuning, including some hyperparameter search) was approximately 8-10 days on a multi-GPU setup. This cumulative runtime reflects the total computational time for all experiments across numerous models, including supervised fine-tuning, classifier training, and extensive reinforcement learning procedures. Supervised fine-tuning (SL) for models like GPT2, GODEL, Llama2, Phi3, Llama3, Mistral, and Qwen2.5 typically involved 3-5 epochs, batch size 4-8 per GPU, AdamW, LR 2e-5 to 5e-5 with warmup. RL fine-tuning (for GPT2-rl, Phi3-rl, Qwen2.5-rl) used PPO from TRL (von Werra et al., 2020) (batch size 64-128 for PPO updates, 1-2 PPO epochs per experience batch). Classifiers were fine-tuned for 10-15 epochs. EcoNudge (740 communications, 3700 dialogues) used an 80:10:10

<sup>2</sup>GODEL Model: [https://huggingface.co/microsoft/GODEL-v1\\_1-base-seq2seq](https://huggingface.co/microsoft/GODEL-v1_1-base-seq2seq)

<sup>3</sup>Llama2 Model: <https://huggingface.co/meta-llama/Llama-2-7b-hf>

train:validation:test split. The classifier does add overhead to the process. For models phi and larger the overhead was less than 1 percent of the time required for the response generation. Even for smaller models the overhead was less than 2 percent of the overall time.

## B Distinguishing Between Persuasion and Guidance/Nudging

In this paper, we focus on "nudging" or subtle guidance rather than direct persuasion. **Persuasion** often involves explicit attempts to change beliefs or actions, possibly using strong rhetoric. **Guidance/Nudging** aims to subtly influence choice architecture or provide timely, supportive framing to encourage consideration of a behavior, without restricting choice. It emphasizes information, collaboration, and gentle encouragement. Our agent design aligns with the "Guidance Approach": building rapport, offering support, and highlighting benefits, while respecting user autonomy.

### Persuasion Approach:

- Identify the audience: This involves understanding the audience’s characteristics, preferences, and motivations to tailor persuasive messages effectively.
- Gather evidence: Collect relevant data, facts, and examples to support the persuasive arguments presented to the audience.
- Develop key messages: Crafting clear and compelling messages that resonate with the audience’s values, beliefs, and interests.
- Structure your presentation: Organizing the persuasive communication logically and coherently to enhance understanding and retention.
- Use persuasive techniques: Employing various strategies such as appealing to emotions, using logic and reason, or creating a sense of urgency to influence the audience’s attitudes or behaviors.

### Guidance Approach:

- Build rapport and trust: Establishing a positive and trusting relationship with the audience to enhance receptiveness to guide.
- Lead by example: Demonstrating desired behaviors or attitudes to inspire others to follow suit.



- Engage their social circle: Involving the social network or community of the audience in the guide process to amplify the impact.
- Highlight social proof: Showcasing examples or testimonials of others who have adopted the desired behaviors or beliefs to provide social validation and encouragement.
- Provide resources and support: Offering assistance, guidance, or resources to facilitate the audience’s adoption of the desired behaviors or attitudes.

### B.1 Indicators of Guidance/Influence

Indicator	Example
Adopting desired behavior	"I've started taking shorter showers and noticed an impactful reduction in my water consumption. It's become a regular habit now."
Attitudinal shifts	"I used to think just my actions wouldn't make a difference, but now I realize that even small changes matter."
Embracing related practices	"Reducing water usage was useful in reducing my bills. I've started composting and using reusable bags when shopping. It feels great to incorporate more such practices into my lifestyle."
Spreading the message	"I've been sharing my experiences with friends and family, encouraging them to join me. It's amazing how conversations can inspire change."
Persistence of changes	"It's been months since I made the switch to a more environmentally conscious lifestyle, and I'm determined to keep it up."

Table 10: Indicators of influence confirmation with examples.

Both persuasion and guidance/influencing can be practical approaches, and you can blend elements from both methods based on the individual and the situation. Tailor your approach to the person’s preferences, motivations, and circumstances to increase the likelihood of them adopting suggested behaviors.

### C Phase 1: Sustainability Data Curation

The 100 initial sustainability suggestions were curated by compiling key environmental practices. We then consulted public resources (Wikipedia, environmental organization sites like WWF/EPA, government publications) for details, benefits, and examples. Only information aligning with verifiable references or accepted consensus was selected.

This ensured the LLM’s input for dialogue generation (Table 11) was grounded in credible knowledge. Each suggestion was detailed with multiple discussion angles (benefits, statistics, etc.).

Element	Content
Suggestion	Reduce water usage by taking shorter showers.
Inspirational appeal	Conserving water helps protect our planet for future generations.
Personal benefits	Shorter showers can refresh you and save time.
Monetary benefits	Reducing water usage can lower your water bill.
Statistics	The average shower uses 2.5 gallons of water per minute.; By cutting your shower time in half, you can save up to 1,825 gallons of water per year (based on average showering timing and frequency) <sup>1</sup> .
Stories	In 2016, the state of <location >launched a campaign to enable over 21,000 villages to become self-reliant for their water needs by 2020.
QnA	<b>Why should I take shorter showers?</b> Taking shorter showers helps conserve water, protect the environment, and save money in the long run.

Table 11: Sample Suggestion with relevant data for Environmental Sustainability

### C.1 List of Sample Suggestions

- Reduce water usage by taking shorter showers.
- Turn off the tap while brushing teeth or washing dishes.
- Install low-flow showerheads and faucets.
- Collect rainwater for gardening.
- Fix any leaks in faucets or pipes.
- Opt for reusable water bottles instead of single-use plastic bottles.
- Bring reusable bags for grocery shopping.
- Say no to plastic straws and bring your own reusable straw.
- Use a reusable coffee cup instead of disposable ones.
- Pack a waste-free lunch with reusable containers and cutlery.
- Recycle paper, plastic, glass, and metal.
- Use energy-efficient light bulbs.
- Turn off lights when not in use.
- Unplug electronics when not in use or use power strips.
- Set computers and other devices to energy-saving modes.
- Opt for public transportation, carpooling, or biking instead of driving alone.
- Support local farmers and buy locally grown food.

- Choose organic and sustainably produced food.
- Compost food scraps and yard waste.
- Plant trees and create a green space in your yard or community.

## D Phase 2: Human-Human Interaction and Dataset Annotation Details

The initial human-human interaction study (200 dialogues) involved three annotators over one month. These dyadic conversations, where one annotator adopted a preliminary persona and the other acted as a guiding agent, were crucial for: developing and refining the five final user personas (Appendix F); identifying a core set of effective agent dialogue acts (Appendix E); and understanding natural dialogue flows. Annotator instructions emphasized persona consistency and conversational responses. Subsequently, eight annotators (including the initial three) annotated these 200 conversations to formalize dialogue acts, achieving 86% agreement. This phase also informed guidelines for LLM data validation. Disagreements primarily arose between pairs of intents, such as "doubt-complain" and "inquiry-intrigue." These conflicts were resolved through majority voting following group discussions. The instructions given to the annotators were:

- 1. Engage in conversations while consistently following your assigned persona's likes, dislikes, and communication style.
- 2. Respond conversationally (utterances shouldn't be too long, >50 words), ask follow-up questions across multiple dialogue if resolution is not achieved, and keep interactions coherent while staying in character.
- 3. If asked something outside your persona details, make reasonable assumptions while maintaining consistency.
- 4. After each conversation, summarize key points, note any inconsistencies, and highlight difficult moments.
- 5. Use discussion feedback to adjust persona traits and improve future interactions. For dialogue acts, merge redundant dialogue acts, whenever possible.
- Discuss Challenges – Finally co-ordinate with other annotators to resolve final personas, dialogue acts to improving consistency.

## E Dialogue Act Details

The 17 Bot/Agent Actions (Table 2) were finalized via an iterative process: (1) Initial Identification from human dialogues; (2) Literature Alignment with persuasion frameworks (Cialdini, 2003; Petty et al., 1986); (3) Categorization & Refinement; (4) Validation by annotators (86% agreement). This dual grounding (empirical and theoretical) aimed for a robust and relevant taxonomy. Each act serves a distinct communicative function in guiding the user. These patterns were further validated and expanded using insights from behavioral psychology and communication theory, including principles from the mentioned papers. For example, Consensus, SocialAppeal, and RationalAppeal align with established persuasion strategies, while Encouragement and HopefulConsideration stem from motivational frameworks. Each act represents a distinct communicative or persuasive function, ranging from logical appeals (RationalAppeal) to emotional engagement (InspirationalAppeal) and practical guidance (Simplify, Suggestion). This categorization was iteratively refined through collaborative discourse between the annotators, ensuring comprehensive coverage. The acts capture various conversational tactics and provide a structure for influencing as a task-oriented dialogue system, where logical reasoning, emotional engagement, and actionable guidance are important for effective interaction.

## F Persona Category Details

The five user personas (Table 1) were defined based on patterns from human-human interactions (Appendix D) and literature on user typologies in behavior change (Cialdini, 2016; Fogg, 2009). They represent a spectrum of receptivity. Table 12 shows example utterances.

The following personas were finalized after the quality analysis of the human interactions. **Active Persona:** Exhibiting an ardent commitment to environmental causes, the active persona proactively engages in environmentally conscious behaviors. Their passionate involvement spans a spectrum of eco-friendly initiatives, driven by an unwavering enthusiasm to impact the environment positively. They actively seek and adopt sustainable habits, continuously seeking and embracing novel methods to contribute to ecological welfare. Their proactive approach reflects an eagerness to explore and an openness to assimilating new sus-

<b>Context</b>	User: Hello! Agent: Hi there! How can I assist you today? User: I heard about using natural pest repellents in the garden instead of chemicals. But I'm not sure why. Agent: Using natural pest repellents has several benefits: protects the environment by reducing harmful chemicals, promotes biodiversity, ensures beneficial insect safety, and can save money.
<b>Active</b>	User: Sounds Interesting. I'll think about it.
<b>Skeptical</b>	User: Hmm, I'm not entirely convinced. Can you provide more evidence or details about how natural repellents are better than chemicals?
<b>Inquisitive</b>	User: Interesting! Can you elaborate on how these natural repellents protect the environment and promote biodiversity?
<b>Materialistic</b>	User: I'm curious about the practical advantages. How can using natural repellents in my garden save me money?
<b>Biased</b>	User: I've heard about these natural options before, but I'm not sold. I doubt they're as effective as chemicals. Convince me otherwise if you can.

Table 12: Samples for user utterances for the user persona categories for the same context in the proposed EcoNudge dataset.

tainable practices. This penchant for embracing and advocating eco-conscious living significantly fosters a more sustainable lifestyle and sets an inspiring example for others. **Skeptical Persona:** The skeptical persona embodies a cautious and critical mindset towards environmental initiatives. Often characterized by a persistent sense of doubt, they approach eco-friendly practices carefully and questioningly. They exhibit a methodical approach to embracing environmental changes, requiring substantial evidence or assurance to consider altering their existing behaviors or habits. Their inclination towards skepticism drives them to seek concrete reasoning and assurance behind any suggested eco-friendly practices, ensuring a thorough evaluation before adjusting their lifestyle. **Inquisitive Persona:** Characterized by an insatiable curiosity about environmental matters, the inquisitive persona actively pursues comprehensive information and an in-depth understanding of sustainable practices. Their genuine interest in delving into the details drives them to seek a profound sense of the underlying principles behind environmentally conscious behaviors. They exhibit a propensity for exploring various facets of eco-friendly practices, driven by a deep-seated curiosity to comprehend the intricacies and implications of sustainable living. Their persistent quest for knowledge contributes significantly to their commitment to adopting environmentally responsible behaviors. **Materialistic Persona:** Driven by pragmatic motivations, the materialistic persona emphasizes the tangible benefits of eco-friendly choices. Their prioritization of visible and measurable rewards underlines their inclination towards environmen-

tal actions offering practical, tangible advantages. They seek to align eco-conscious behaviors with personal gains, whether economic or in terms of immediate practical benefits. Their pragmatic approach to sustainability emphasizes the importance of visible and direct advantages guiding their environmental practices. **Biased Persona:** The biased persona manifests entrenched resistance or firm preconceptions towards adopting environmentally friendly recommendations. Their reluctance to embrace eco-conscious behaviors stems from deeply ingrained biases or fixed viewpoints that create barriers to accepting environmental initiatives. Overcoming these biases requires tailored strategies to address their concerns or dispel misconceptions. Successfully engaging this persona group necessitates approaches designed to challenge and gradually alter their fixed viewpoints through patient and targeted interventions.

### G Phase 3: Dataset Generation - Llama2 Prompt

In order to generate our EcoNudge dataset, we utilized the LLM Llama2. We prompted the model based on selected dialogue acts, dialogue flows, and persona traits. We created the following prompt with suggestions/approval from the annotators. The end part of the prompt is used to format the data into a table for easier processing. Prompt = f"""  
Generate a communication between a user and an assistant/bot.

The communication must consist of a set of 5 conversations.

The assistant aims to influence the user to be more environmentally sustainable.

The conversation starts with either a greeting from the bot and a suggestion, or the user can start the conversation by mentioning their complaints regarding a suggestion they tried previously. The user never asks for a suggestion.

The user asks many questions regarding the tangible benefits and complains a lot when the benefits are not noticeable.

The bot provides the benefits of the suggestion.

The first focus is on ethical/inspirational benefits however, if the user is still not convinced, the bot suggests monetary benefits with rationality.

In the first conversation, the user is skeptical, doubtful, and barely convinced. In the second conversation, the user mentioned they tried it and didn't see the benefits. In the third conversation,

the user might still be doubtful but still consider the suggestions, as the benefit is noticeable. In the fourth conversation, the user tries to get suggestions related to the original one, as the benefits are good and it feels good. In the fifth conversation, the user appreciates the suggestion and asks for actions to further the effort. The conversation initially focuses on one type of suggestion.

The information in the first triple quotes is the user’s intent. The information in the second triple quotes is the bot’s intent. The third triple quote has a list of suggestions. The The fourth triple quote contains the strategies that can influence the user. The conversation must adhere to the suggested points only and no new intents should be generated for the user or bot. Conversations must be at least turns turns long.

```
user intents = "{user_intents}" bot intents =
"{bot_intents}" suggestions = "{suggestions}"
strategies = "{strategies}"
```

Important points: important

Create the 5 conversations in the form of a single combined table with 5 columns:

The first column is for the ‘conversation’ number starting from 0. All utterances of a single conversation have the same conversation number. The second column is for the ‘turn’ number for the current conversation. The third column for the ‘speaker’ corresponds to the user or assistant. Fourth for the ‘utterance’, and Fifth, for the ‘intent’. All 5 conversations generated should have the same column names, declared only once at the top of the response. Generate absolutely nothing besides the table. ""

## H Phase 4: Data Quality Control Details

Three human participants (one employed annotator, one non-author Ph.D. scholar, and one master’s student) conducted manual checks and guideline-based scoring 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 sensitivity.

The Overview of the process was as follows:

- 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 regarding natural tone and flow, formality, and cultural sensitivity, was emphasized.
- Ensure that user attributes and characteristics remain consistent throughout the dialogue to maintain coherence and believability.
- Participants were requested to pay special attention to the portrayal of persuasive interactions, aiming to enhance politeness and empathy.

The dataset quality was evaluated by human annotators based on internal coherence, content consistency, and naturalness using an integer Likert scale of 1-3. Specifically: Score 1: Indicates poor quality, and such dialogues were discarded. Score 2: Indicates moderate quality, and these dialogues were corrected manually. Score 3: Indicates high quality, and these dialogues were retained as they were. The scores for the 5 focused properties are provided in Table 13. The annotators were instructed to collaborate to achieve a consistent score across all three evaluators. In cases of disagreement, the lower score was considered final. This approach ensured quality by prioritizing the scores with better consensus. Communications receiving more than two scores of 1 were automatically rejected with a quality score of 1. Conversely, communications with a cumulative score equal to or over 20 were directly accepted with a quality score of 3. The remaining communications were given to the annotators for correction. Those that could be successfully corrected were accepted with a quality score of 2, while the others were rejected with a quality score 1.

Three human experts validated the LLM-generated dialogues on a 1-5 Likert scale for: Grammatical Correctness, Thematic Consistency, Language Appropriateness, User Profile Consistency, and Sensitivity. Process: Independent rating, disagreement resolution, rejection/correction based on scores. From 1000 initial communications: 19.7% rejected, 58.8% corrected, 21.5% accepted directly. 63 similar valid communications were removed, yielding 740 final communications.



Inter-evaluator Kappa for overall quality was 0.78.

Criteria	Score 1	Score 2	Score 3	Score 4	Score 5
Grammatical Correctness	97	148	325	299	131
Thematic Consistency	112	118	335	286	149
Language Appropriateness	85	179	342	295	97
User Profile Consistency	128	153	318	275	108
Sensitivity	102	143	308	278	121

Table 13: Scores for Evaluation Criteria during human validation of generated data.

Additionally, the inter-evaluator Kappa agreement ratios were 80.3%, 82%, and 84%, demonstrating strong agreement among the annotators for internal coherence, content consistency, and naturalness, respectively. After conducting manual checks and receiving feedback from users, the data significantly improved, reflecting a refined level of quality control.

## I Phase 5: Modeling - Formulations

Once the dataset is created, we utilize the following methods to validate our dataset and the proposed system.

### I.1 LLM Base Formulation

For a supervised dialogue dataset  $D = \{d_0, d_1, \dots, d_N\}$ , where  $d = (a_0, u_0, \dots, a_i, u_i, \dots, a_{T-1}, u_{T-1})$  - a multi-turn dialogue with  $u_i + ctx_i$  ( $u_i$  - user's utterance at  $i$ th turn and  $ctx_i = (a_0, u_0, \dots, a_{i-1})$ ) as input and  $a_i$  (agent's utterance at  $i$ th turn) as output. The supervised learning dialogue model  $\rho_\theta(d)$  can be expressed as:

$$\rho_\theta(d) = \prod_{T=0}^{T-1} \rho_u(u_i | u < i, a < i) \rho_a(a_i | u \leq i, a < i) \quad (1)$$

Where  $\rho_u$  and  $\rho_a$  are the probability distributions of the user's and agent's utterances. This trained SLDM is finetuned in an RL setting using the PPO loss formulated as below:

$$L_{CLIP}(\theta) = E[\min(\text{pr}_r(\theta)A_r, 1 - \epsilon, 1 + \epsilon)\text{clip}(\text{pr}_y(\theta)A_r)] \quad (2)$$

where  $\text{pr}_r(\theta) = P_{\text{new}}(\theta)/P_{\text{old}}(\theta)$ .  $\epsilon$  and  $A_y$  denote the clipping range and normalized rewards, respectively. Finally, the parameters' updation is done as follows:

$$\theta_{k+1} = \arg \max_{\theta} E_{s,a \sim P_{\theta_k}} [L_{CLIP}] \quad (3)$$

### I.2 Formulation of reward functions

**Length Reward** penalizes the generated response by considering the absolute difference between its length and the target length. The formula is expressed as follows:

$$R(\text{Len}) = -|L - T| \quad (4)$$

Here,  $L$  denotes the length of the generated response, and  $T$  represents the target length. This reward's primary rationale is to encourage responses of an appropriate and balanced length that facilitate effective communication, especially for a chatbot intended to be integrated as a pop-up on a webpage, where overly verbose or unnecessarily brief responses can detract from natural dialogue flow and user experience. The target length ( $T$ ) is a hyperparameter that can be adjusted to reward either more concise or more verbose answers, depending on the desired conversational style for a specific task.

**Semantic Coherence Reward** method (Li et al., 2016) involves assessing the mutual information between the action  $a$  and preceding dialogue turns in the history to ascertain the coherence and appropriateness of generated responses. The equation for the reward function  $R(SC)$  is formulated as follows:

$$R(SC) = \frac{1}{N_a} \log p_{s2s}(a|q_i, p_i) + \frac{1}{N_{q_i}} \log p_{s2s}(q_i|a) \quad (5)$$

Here,  $p_{\text{seq2seq}}(a|p_i, q_i)$  represents the probability of generating response  $a$  given the previous dialogue utterances  $[p_i, q_i]$ . Meanwhile,  $p_{\text{backward seq2seq}}(q_i|a)$  denotes the backward probability of generating the prior dialogue utterance  $q_i$  based on the response  $a$ . The training of  $p_{\text{backward seq2seq}}$  follows a similar procedure as standard models but with sources and targets interchanged. To mitigate the impact of target length, both  $\log p_{\text{seq2seq}}(a|q_i, p_i)$  and  $\log p_{\text{backward seq2seq}}(q_i|a)$  are scaled by the length of targets.

**Change in Opinion Sentiment Reward:**

$$O(S_i, S_f) = \begin{cases} -0.5, & \text{if } (S_i = '-') \text{ AND } (S_f = '-') \\ -1.0, & \text{if } (S_i = '+') \text{ AND } (S_f = '-') \\ +0.5, & \text{if } (S_i = '-') \text{ AND } (S_f = '+') \\ +1.0, & \text{if } (S_i = '+') \text{ AND } (S_f = '+') \\ \text{None}, & \text{otherwise} \end{cases} \quad (6)$$

Where:  $O(S_i, S_f)$ : Change in Opinion Sentiment Reward;  $S_i$ : Initial Sentiment;  $S_f$ : Final Sentiment. - determines negative and + determines positive changes to the opinion sentiment. The sentiment reward values were carefully designed to encourage positive opinion shifts. The differential penalty for consistently negative users (-0.5) versus the reward for consistently positive ones (+0.5) was established to subtly disincentivize stagnation in negative sentiment, promoting a proactive approach from the agent to try and shift user opinion, even if minimally. These values were normalized to fall within a consistent range (e.g., -1 for worst-case, +1 for best-case transitions) to ensure coherent reward signals within the PPO framework. The method for detecting opinion shifts relies on an aspect-based sentiment classifier, as outlined in Appendix A.

#### Intermediate Completion Reward:

$$R_{\text{intermediate}}(a) = \begin{cases} 0, & \text{if } a = \text{'rejected'}$$

$$\begin{cases} 0.3, & \text{if } a = \text{'considered'}$$

$$1, & \text{if } a = \text{'implemented'}$$

$$(7)$$

Where:  $R_{\text{intermediate}}$ : Intermediate Completion Reward;  $a$ : Action taken. The reward values for "rejected" (0), "considered" (0.3), and "implemented" (1) were set to reflect the progressive impact of user engagement in a nudging dialogue. While heuristic, these values are grounded in our empirical observations during human-human interactions and iterative model tuning, providing a calibrated signal for the agent's progress. "Considered" is mapped to a low value as this just implies that the response aligns with the user persona, but whether the user will even try it is uncertain.

#### Positive Affirmation Reward:

$$R_{\text{affirm}}(U) = \begin{cases} -1, & \text{if } U = \text{'dislike'}$$

$$1, & \text{if } U = \text{'affirm\_usefulness'}$$

$$(8)$$

Where:  $R_{\text{affirm}}$ : Positive Affirmation Reward;  $U$ : User Response binary categorization.

#### Personalized Support Reward

$$R(\text{Supp}) = \sum (w_i \times R_i) + w_{\text{diversity}} \times D \quad (9)$$

$\sum (w_i \times R_i)$  is the sum of rewards for each support type, weighted by importance.  $w_{\text{diversity}}$  is the weight of the diversity reward (non-repetitiveness and variations).  $D$  is the diversity score, which measures how evenly the agent uses different support types.  $w_i$  depends on persona and compliance rate.

## J Phase 6: Human Evaluation Details

We utilized five human annotators (Two employed annotators, two non-author Ph.D. scholars, and one master's student) for human evaluation. The inter-annotator agreement between the five human annotators was 80.5%.

### J.1 Annotator Guidelines for Human Evaluation Metrics

**Scale of 1-5:** The system performance is evaluated on a scale from 1 to 5, where higher scores indicate better performance. A score of 1 represents poor performance, indicating impactful issues or failures in the specific criterion. A score of 2 indicates below-average performance with several noticeable problems. A score of 3 reflects average performance, meeting basic expectations but with room for improvement. A score of 4 represents good performance with minor issues. A score of 5 indicates excellent performance, meeting or exceeding expectations with no noticeable issues.

Five human evaluators assessed dialogues on a 1-5 scale (1 = Poor, 5 = Excellent). Metrics:

- 1. Support Consistency (S-Con) - Definition:** Measures the consistency of the negotiation approach within a dialogue. - **Guidelines:** - Evaluate if the system maintains a coherent negotiation strategy throughout the dialogue. - Check for any contradictions or shifts in negotiation tactics. - Rate based on how consistently the system supports its negotiation stance.
- 2. Guidance Efficacy (G-Eff) - Definition:** Measures the ability of the system to present compelling arguments, reasoning, or incentives that guide the other party. - **Guidelines:** - Assess the persuasiveness and logical structure of the arguments presented. - Determine if the arguments effectively influence or guide the other party. - Rate based on the strength and clarity of the guidance provided.
- 3. Dialogue-Engagingness (D-E) - Definition:** Evaluates the extent to which a dialogue is interesting, captivating, and can hold the participants' attention. - **Guidelines:** - Judge how engaging and interactive the dialogue is. - Consider if the dialogue keeps the participants interested and involved. - Rate based on the overall captivation and enjoyment of the dialogue.

- 1446 4. **Fluency - Definition:** Assesses the readabil-  
1447 ity and grammatical correctness of the text.  
1448 - **Guidelines:** - Check for grammatical er-  
1449 rors, typos, and awkward phrasing. - Ensure  
1450 sentences flow naturally and are easy to read.  
1451 - Rate based on the overall readability and  
1452 grammatical correctness.
- 1453 5. **Context Relevance - Definition:** Measures  
1454 how well the text matches the given context  
1455 or query. - **Guidelines:** - Verify that the con-  
1456 tent directly addresses the query or context. -  
1457 Ensure all parts of the response are relevant  
1458 to the given context. - Rate based on the perti-  
1459 nence and appropriateness of the response to  
1460 the context provided.

1461 **K Automatic/Generic evaluation results**  
1462 **for the EcoNudge dataset**

1463 **L Sample Dialogue from EcoNudge**  
1464 **Dataset**

1465 Sample multi-dialogue communications from the  
1466 dataset are shown in Table 15.

1467 **M Sample Communication by LLM**

1468 Sample multi-dialogue communications by the  
1469 trained dialogue system are shown in Tables 16  
1470 and 17.

Type	Models	BLEU-1	METEOR	ROUGE-L	CIDEr	SPICE	STCS	EACS	VECS	GMS
Base LMs	GPT2	0.194	0.116	0.146	0.134	0.097	0.604	0.889	0.464	0.735
	GODEL	0.277	0.161	0.241	0.488	0.205	0.724	0.919	0.519	0.769
Supervised Fine-tuned LLMs	Phi3.3B	0.343	0.189	0.311	0.872	0.302	0.823	0.937	0.580	0.801
	Llama2-7B	0.381	0.219	0.329	0.879	0.311	0.830	0.946	0.589	0.811
	Llama3-8B	0.385	0.225	0.330	0.882	0.314	0.835	0.949	0.596	0.819
	Mistral-7B	0.391	0.229	0.339	0.889	0.320	0.839	0.956	0.599	0.824
	Qwen2.5	0.395	0.245	0.345	0.915	0.330	0.843	0.970	0.608	0.825
Prompted LLMs	GPT-3.5	0.396	0.246	0.346	0.916	0.331	0.844	0.972	0.609	0.826
	GPT-4	0.399	0.252	0.353	0.924	0.339	0.859	0.985	0.621	0.840
	GPT-4o	<b>0.405</b>	<b>0.258</b>	<b>0.359</b>	<b>0.935</b>	<b>0.348</b>	<b>0.865</b>	<b>0.989</b>	<b>0.628</b>	<b>0.847</b>
RL Enhanced Models	GPT2-rl	0.359	0.213	0.321	0.886	0.284	0.823	0.947	0.579	0.807
	Phi3-rl	0.399	0.249	0.351	0.919	0.336	0.855	0.982	0.619	0.837
	Qwen2.5-rl	0.402	0.253	0.355	0.928	0.342	0.861	0.986	0.625	0.842

Table 14: Results for automatic evaluation on the EcoNudge dataset. Best overall scores in bold.



c_id	turn	speaker	utterance	intent
0	0	User	Hi there!	Greet
0	1	Bot	Hello! Did you know that supporting clean and renewable energy sources can help the environment?	Suggestion
0	2	User	Hmm, I'm not sure. Why should I support clean energy?	Benefits
0	3	Bot	Clean energy reduces pollution and greenhouse gas emissions, leading to a healthier planet for future generations.	Benefits
0	4	User	That's a good point. I'll consider it.	Affirm
0	5	Bot	Great! Remember, small actions can make a big impact.	Encouragement
0	6	User	Thanks for the suggestion. Bye!	Goodbye
0	7	Bot	Goodbye! Remember to make sustainable choices whenever possible.	Goodbye
1	0	User	Hi there!	Greet
1	1	Bot	Hello! Did you know that collecting rainwater for gardening is a great way to be environmentally sustainable?	Suggestion
1	2	User	Hmm, I'm not sure about that.	Deny
1	3	Bot	Collecting rainwater reduces water waste and helps plants thrive.	Benefits
1	4	User	I see. What other benefits are there?	Benefits
1	5	Bot	Participating in beach or park clean-ups helps preserve natural habitats and promotes community involvement.	Benefits
1	6	User	That sounds interesting.	Affirm
1	7	Bot	Attending environmental workshops, conferences, or seminars can provide valuable knowledge and resources for sustainable living.	Suggestion
1	8	User	I'll consider it. Thanks for the suggestion!	Affirm
1	9	Bot	You're welcome! Remember, small actions can make a big impact on the environment.	Encouragement
2	0	User	Hi there!	Greet
2	1	Bot	Hello! Did you know that reducing water usage by taking shorter showers can help the environment?	Suggestion
2	2	User	Hmm, I'm not sure about that.	Deny
2	3	Bot	Shorter showers can save up to 8 gallons of water per day. That's 2,920 gallons per year!	Benefits
2	4	User	That's impressive, but what's in it for me?	Benefits
2	5	Bot	By reducing water usage, you can save money on your water bill and help conserve a precious resource.	Benefits
2	6	User	Okay, I see your point.	Affirm
2	7	Bot	Great! Remember, small changes can make a big impact. Is there anything else I can help you with?	Encouragement
2	8	User	No, that's all for now. Thanks!	Goodbye
3	0	User	Hi there!	Greet
3	1	Bot	Hello! Did you know that creating a compost bin for food waste and yard trimmings can greatly reduce your carbon footprint?	Suggestion
3	2	User	Hmm, I've never thought about that. Why should I do it?	Benefits
3	3	Bot	Composting reduces methane emissions from landfills and creates nutrient-rich soil for plants. Plus, it's a great way to reduce waste.	Benefits
3	4	User	That's interesting, but it seems like a lot of work.	Deny
3	5	Bot	It's actually quite easy and can save you money on fertilizer. Plus, you'll feel good knowing you're doing your part for the environment.	Rational appeal
3	6	User	I guess that makes sense. I'll give it a try.	Affirm
3	7	Bot	Great! Let me know if you need any tips on getting started. Remember, small changes can make a big impact.	Encouragement
3	8	User	Thanks, I will. Goodbye!	Goodbye
4	0	User	Hi there!	Greet
4	1	Bot	Hello! Did you know using a reusable coffee cup instead of disposable ones can reduce waste?	Suggestion
4	2	User	Hmm, I never thought of that. Why should I do it?	Benefits
4	3	Bot	Using a reusable cup can save money in the long run and reduce the amount of waste in landfills.	Benefits
4	4	User	That's a good point. I'll consider it.	Affirm
4	5	Bot	Great! Remember, small changes can make a big impact on the environment.	Encouragement
4	6	User	Thanks for the suggestion. Bye!	Goodbye
4	7	Bot	Goodbye! Remember to keep the environment in mind in your daily choices.	Encouragement
5	0	User	Hi there!	Greet
5	1	Bot	Hello! Did you know that educating children about environmental sustainability can have a huge impact?	Suggestion
5	2	User	Hmm, I'm not sure. Why should I do that?	Benefits
5	3	Bot	By educating children, we can create a more sustainable future and inspire the next generation to care for the environment.	Benefits
5	4	User	That's a good point. I'll consider it.	Affirm
5	5	Bot	Great! Remember, small actions can make a big difference.	Encouragement

Table 15: Sample Communication data generated by the LLM by inquisitive persona. We imply that the entire communication occurred over the course of 5 dialogues/conversations denoted by c\_id.

Speaker	Utterances
<b>Dialogue 1</b>	
User	hello
Bot	Have you considered joining or starting an environmental sustainability club in your community?
User	It sounds like a lot of work
Bot	Joining a club can be a great way to meet like-minded people and make a positive impact on the environment.
User	I'll think about it bye
Bot	Thanks! Looking forward to see you again
<b>Dialogue 2</b>	
User	hello again
Bot	Hello again! How did joining the sustainability club go?
User	I thought about it but its way too much effort
Bot	I understand, but joining a club can also be a great way to make a positive impact on the environment.
User	not interested anything else
Bot	Have you considered collecting rainwater for gardening? Other applications of this free-standing rainwater collection approach are sustainable gardening and small-plot farming. know more.
User	I dont really see the benefits of collecting rainwater
Bot	Collecting rainwater can save money on water bills and reduce strain on local water resources.
User	I guess it's just a little thing. I'll give it a try. not really sure about it though
Bot	That's great to hear! Let me know if you have any questions or concerns.
User	bye
Bot	Thanks! Looking forward to see you again
<b>Dialogue 3</b>	
User	hello
Bot	Hello again! How did collecting rainwater go?
User	You know what. I tried it and it did come useful in a pinch
Bot	That's wonderful to hear! Is there anything else you'd like to try?
User	Yeah. Any other such suggestions that could be useful for me. nothing too hard
Bot	Have you considered using electronic invoices and payments instead of paper? The United States treasury estimated that implementing e-invoicing across the entire federal government would reduce costs by 50% and save \$450 million annually. know more.
User	Sounds interesting but I don't really see the benefits
Bot	Using electronic invoices and payments can reduce paper waste and save trees.
User	I guess even a virtual record is a record. I'll give it a try.
Bot	Great! Let me know if you have any questions or concerns.
User	bye
Bot	Thanks! Looking forward to see you again
<b>Dialogue 4</b>	
User	Hello
Bot	Hello again! How did using electronic invoices and payments go?
User	You know what. I tried it and I like not having to deal with all the papers
Bot	great to hear
User	Yeah. Any other such suggestions that could be useful for me. nothing too hard
Bot	Have you considered using a programmable thermostat to optimize energy usage? According to Consumer Reports magazine, programmable thermostats can reduce energy bills by about \$180 a year. know more.

Table 16: Sample Communication during the inference phase. The communication occurs over four dialogues with a presumed gap between them to accommodate consideration and implementation.

Speaker	Utterances
<b>Dialogue 1</b>	
bot	Hi there! My name is Green Cruiser, but you can call me G-Cruz.
user	hello
bot	Have you considered joining or starting an environmental sustainability club in your community?
user	not really. that sounds like a lot of work
bot	Joining a club can be a great way to meet like-minded people and make a positive impact on the environment.
user	not interested. anything else
bot	Have you considered using electronic invoices and payments instead of paper? The United States treasury estimated that implementing e-invoicing across the entire federal government would reduce costs by 50% and save \$450 million annually. know more.
user	no. I kind of like to have a record
bot	Well, it's a small change that can make a big impact on reducing waste and helping the environment. Would you like to know more?
user	not interested bye
bot	Thanks! Looking forward to see you again
<b>Dialogue 2</b>	
user	Hello again
bot	Hello again! How did using electronic invoices and payments go?
user	I never said I would try it. I dont really see the benefits of it anyways
bot	Well, it can save you time and money, and it's better for the environment in the long run.
user	I dont care about environment that much. How does it save me money
bot	It depends on your spending habits, but it's usually a good bet. Plus, it's a small change that can be easily implemented.
user	I guess even a virtual record is a record. I'll give it a try. not really sure about it though
bot	That's great to hear! Let me know how it goes.
user	bye
bot	Thanks! Looking forward to see you again
<b>Dialogue 3</b>	
user	Hello
bot	Hello again! How did using electronic invoices and payments go?
user	You know what. I tried it and I like not really having to deal with all the papers
bot	great to hear
user	Yeah. Any other such suggestions that could be useful for me. nothing too hard
bot	Have you considered reducing paper usage by going digital for bills and documents? The United States treasury estimated that implementing e-invoicing across the entire federal government would reduce costs by 50% and save \$450 million annually. know more.
user	Already doing that. anything else.
bot	Have you considered using a bamboo toothbrush instead of a plastic one? Over 1 billion toothbrushes are disposed of into landfills annually in the United States alone. know more.
user	how does that help
bot	Going digital reduces paper waste and saves trees. It's also more convenient and eco-friendly.
user	how does using bamboo toothbrushes help
bot	Bamboo toothbrushes are biodegradable and can reduce plastic waste in landfills and oceans. They're also more gentle on your teeth and gums.
user	I guess. Ill give that a shot
bot	That's great to hear! Let me know how it goes.

Table 17: Sample Communication during the inference phase. The communication occurs over three dialogues with a presumed gap between them to accommodate consideration and implementation.