# **ORACLE**: **O**perationalizing social **R**epair **A**nd self-monitoring in **C**ol**L**aborative interactive **E**nvironments

Alafate Abulimiti

## Host Institution and Proposed Advisors

This post-doctoral program will be supervised by DR. Catherine Pelachaud hosted by the ISIR Institute at Sorbonne University and also by PR. Agnes Helme-Guizon hosted by the CERAG at Grenoble Alpes University.

## Context

The field of Human-Computer Interaction (HCI), particularly in collaborative interaction, is rapidly expanding. Within this scope, collaborative interaction with virtual agents has emerged as a notably effective and empathetic approach [3]. A critical challenge lies in developing models that accurately represent human-agent dialogue. This involves enabling agents not only to grasp the user's intentions but also social states accurately to respond appropriately based on the dialogue's context. Advances in Natural Language Processing (NLP) have been promising in deciphering dialogue intentions and managing tasks [14]. Additionally, there is growing interest in exploring how conversational strategies can be employed to foster or amplify social engagement [8]. In collaborative communication contexts, it's crucial that agents not only facilitate goal achievement but also keenly attend to the social dynamics of the interaction and emotional shifts of the person, adapting their conversational style accordingly. This proposal aims to delve into the intricacies of these interactions, prioritizing the balance between task achievement and social sensitivity in AI-facilitated dialogues. We propose drawing inspiration from Motivational Interviewing (MI) [7], a counseling approach that emphasizes empathetic and goal-oriented conversations to elicit behavior change. By incorporating MI principles, we aim to develop agents capable of engaging in collaborative dialogues that are both socially attuned and effective in achieving desired outcomes.

## Objectives

Collaborative interactions aim to achieve shared goals, which can be impeded by challenges like reluctance or ingrained habits. Collaborative interaction arises from aligned or mutually beneficial objectives, fostering progress through shared goals. Conversely, adversarial dialogue occurs when parties prioritize their own benefits, typical in debates and negotiations. As social beings, we naturally fulfill our conversations with social attributes, from casual small talk to structured interactions. Evidence suggests that enriched social interactions can bolster task-related outcomes [10], and effective task collaboration can enhance social satisfaction [6]. Consequently, when developing models for conversation, it becomes crucial to simultaneously consider both social and task attributes. This leads to our overarching research question: **How can we be more effective in modeling collaborative interactions to accurately reflect their inherent social and task characteristics?**

Integrating an agent as an assistant for collaborative communication introduces the challenge of managing imperfections, particularly in social competencies alongside task execution. An agent lacking in social skills might focus too narrowly on task progression - such as by emphasizing faults or necessary changes, without considering the conversation's social dynamics. While modern agents are adept at correcting task-related errors, their capacity for social rectification remains limited. In my Ph.D., I explored the concept of "social repair" [1] in the context of educational dialogues. I also investigated how conversational strategies, such as the interleaving of on-task and off-task conversations and hedging, can be employed to restore and maintain social factors (e.g., rapport, power, trust, etc.) to appropriate levels. Building upon this foundation, our current investigation seeks to answer two sub-questions:

**RQ1:** What constitutes *"social repair"*, and how can it be modeled and implemented in an agent?

Humans inherently engage in *self-monitoring* during interactions, a critical process of reflecting on and adjusting their behaviors based on the outcomes of those interactions and also their internal reasoning/simulation [12]. This complex and overarching mechanism encompasses both verbal and non-verbal cues, enabling individuals to develop *self-awareness* and proactively address potential missteps. In the context of virtual agents, replicating this self-monitoring capability presents a fascinating research challenge:

**RQ2:** How can we construct a self-monitoring framework for virtual agents, and what methodologies can we employ to ensure its effectiveness?

We intend to evaluate whether implementing a self-monitoring mechanism can significantly improve the agent's operational effectiveness and less "social repair" found in human-agent interaction.

## Approach/Challenges/Originality

The use of collaborative interaction with virtual agents presents a unique opportunity to promote tutoring [10], and potentially healthy lifestyles, particularly in the domain of nutrition. Grounded in evidence from recent studies [11], our approach leverages virtual agents to foster behavioral change in dietary habits. These studies highlight the importance of enjoyment, social engagement, and personalized feedback as key motivators for adopting healthier eating and physical activity habits. Drawing upon these insights, our research aims to develop models that not only grasp user intentions and social states but also encourage positive dietary changes through positive and socially engaging interactions by using strategies like "social repair" and also by empowering users to feel more competent and autonomous. Motivational Interviewing (MI) serves as a foundational methodology in our approach, where the therapists aim to modify deep-seated behaviors, a goal often shared by the interlocutor. However, unlike traditional collaborative interaction, challenges like reluctance or the difficulty of changing ingrained habits can impede progress. Therapists need to navigate not just the task of guiding interlocutors through transitions, but also the social intricacies of dialogue to prevent reluctance, this often involves "social repair" actions. We aim to effectively integrate advanced NLP techniques with social and conversational theories using Symbolic AI [5] to develop models that can dynamically adapt to the nuances of MI by combining computational modeling and empirical analysis, focusing on social repair and self-monitoring. This ensures that agents can not only follow task-oriented instructions but also engage in social interactions. Firstly, We will apply the MI corpus and theoretically and empirically analyze the existing therapeutic

datasets (e.g., AnnoMI [15]) to identify patterns of "social repair" and design algorithms to enable the agent to predict and resolve potential conversational/social malfunctions. Secondly, we will apply social and conversational theories to design and implement a self-monitoring model to enhance the interaction interactivity. Using theories-inspired top-down modeling will also enhance the transparency and explainability of the interaction, which is surely crucial for virtual assistants, as it fosters trust, user acceptance, and long-term engagement. Studies have shown that transparent and explainable AI systems lead to increased user satisfaction, improved decision-making, and reduced cognitive load [13, 4]. Thirdly, an evaluation phase will be integral to our research, assessing the efficacy of the developed models in real-world settings. This step involves deploying the models within a controlled environment to simulate MI sessions, measuring their impact on dietary behavior change and user engagement before and after the interviewing session. The challenges of this research include the complexity of accurately modeling human-like social interactions in an agent, especially in a situation as sensitive as therapy [2]. There are also significant technical difficulties in creating a self-monitoring mechanism that focuses on creating an effective mental model, allowing the agent to assess and adjust its behavior in real-time. This requires the agent to have a deeper understanding of social cues and contexts. The originality of our work lies in its interdisciplinary approach, which combines insights from psychological sociology, linguistics, and AI. Unlike existing models that focus primarily on tasks, our research prioritizes the balance between task-orientation and social interaction.

## Positioning

Our proposal targets the confluence of AI and collaborative interaction, aiming to refine the virtual agent's role in collaborative interactions. It seeks to bridge the gap in current research by creating models that enhance agent's understanding of and participation in collaborative dialogue, focusing on social repair and self-monitoring. Leveraging advancements in NLP and HCI, this work introduces innovative methods to interpret and respond to both verbal and non-verbal interaction, enhancing agent's interaction with users. Ultimately, our research endeavors to improve collaborative outcomes and agent interaction, contributing valuable insights to both Human-agent interaction development and collaborative practices.

## Project organization

**Year 1:** In the first half, we focus on developing a computational model for "social repair" in collaborative interactions, involving analysis of aforementioned collaborative datasets and model development, aiming for a conference publication. The latter half is dedicated to creating a self-monitoring model for agents, combining insights from the earlier phase for another conference presentation.
**Year 2:** The first half of the year aims to integrate the self-monitoring model into our Greta platform [9], enhancing task and social interaction balance, with a milestone of a comprehensive system evaluation.

## Partnership

This project will engage in a strategic partnership with DR. Pelachauld and her Ph.D. student, Lucie Galland, to delve into MI and dialogue modeling, leveraging their expertise in conversational AI. Additionally, PR. Agnes Helme-Guizon and her student, Jade Broyer will contribute their expertise in social and human sciences, focusing on the impact of virtual agent behaviors on user engagement and behavior change. This integrated partnership approach aims to foster a comprehensive examination of AI-driven therapeutic interventions, combining technical development with behavioral research and social impact analysis to advance the field of HCI and health informatics.

## References


[1] A. Abulimiti et al. Towards a social planning model for rapport. In *SemDial*, 2021.

[2] T. Bickmore and J. Cassell. *Social dialongue with embodied conversational agents*. 2005.

[3] D. S. Choi et al. Improving counseling effectiveness with virtual counselors through nonverbal compassion involving eye contact, facial mimicry, and head-nodding. *Scientific Reports*, 2024.

[4] M. Eslami et al. Communicating algorithmic process in online behavioral advertising. In *CHI*, 2018.

[5] M. Garnelo and M. Shanahan. Reconciling deep learning with symbolic artificial intelligence: representing objects and relations. *COBS*, 2019.

[6] D. Harrison et al. Time, teams, and task performance. *Academy of management journal*, 2002.

[7] W. R. Miller and S. Rollnick. *Motivational interviewing: Helping people change*. Guilford press, 2012.

[8] F. Pecune et al. A framework to co-optimize task and social dialogue policies using reinforcement learning. In *ACM IVA*, 2020.

[9] C. Pelachaud. Greta: an interactive expressive embodied conversational agent. In *ICMI*, 2015.

[10] T. Sinha and J. Cassell. We click, we align, we learn: Impact of influence and convergence processes on student learning and rapport building. In *INTERPERSON*, 2015.

[11] S. Snuggs et al. A mixed-methods approach to understanding barriers and facilitators to healthy eating and exercise from five european countries: highlighting the roles of enjoyment, emotion and social engagement. *Psychology & Health*, 2023.

[12] M. Snyder. *Public appearances, Private realities: The psychology of self-monitoring*. 1987.

[13] D. Wang et al. Designing theory-driven user-centric explainable ai. In *CHI*, 2019.

[14] T.-H. Wen et al. A network-based end-to-end trainable task-oriented dialogue system. In *EACL*, 2017.

[15] Z. Wu et al. Anno-mi: A dataset of expert-annotated counselling dialogues. In *ICASSP IEEE*, 2022.