Mixed-Initiative Dialogue Management for Human-Virtual Agents Interaction in Forum Theatre Inspired Training

Samuel Otofa[1], Yacine Zerenini[2], Frederic Bechet[1], Benoit Favre[1],
Jean-Marie Pergandi[2], Magalie Ochs[1]

[1]LIS - Aix Marseille University - CNRS - France,
[2]CRVM - Aix Marseille University - CNRS - France

Correspondence: magalie.ochs@lis-lab.fr

Abstract

This work presents a virtual reality (VR) training tool designed to raise awareness of social discrimination (ethnic and gender-based) and to train individuals to respond effectively when witnessing such situations. Inspired by Augusto Boal's forum theatre, the system recreates interactive scenarios of discrimination using autonomous virtual agents. From a dialogue system perspective, the project introduces a hybrid dialogue management architecture combining state-based control with Large Language Model (LLM)-driven open dialogue. This mixed-initiative approach allows the system to manage structured training sequences while supporting flexible, context-aware interactions on sensitive topics. The demonstrator illustrates this approach through a case of ordinary sexism in a professional setting, highlighting the potential of spoken dialogue systems in VR for experiential learning and social behavior training.

1 Introduction

In the field of training, increasing attention has been directed toward *simulation-based learning* using virtual environments. Numerous studies have examined systems that simulate social interactions with *Socially Interactive Agents* (SIAs) to foster the development of social skills (Bruijnes et al., 2019), applied to the medical domain (Prange et al., 2017; Campillos-Llanos et al., 2015) or in the context of teacher training (Pautler et al., 2018). These studies have shown that SIAs can effectively enhance individuals' interpersonal abilities, for instance, in job interview preparation or intercultural communication training (Anderson et al., 2013; Hall et al., 2011). However, several application domains remain underexplored, notably the use of SIAs for *training to prevent social discrimination*, which is the focus of the present study.

In this context, the demonstration system we describe in this study is a virtual reality (VR) train-ing tool designed to raise awareness of ethnic and gender discrimination and to train users to react appropriately as witnesses. Inspired by Augusto Boal's forum theatre, the system stages interactive discrimination scenarios with autonomous virtual agents. Users observe, analyze, and reenact these situations to explore alternative responses.

Technically, the project introduces a hybrid dialogue architecture combining state-based control with Large Language Model (LLM)-driven open dialogue. This approach supports both structured training and flexible, context-sensitive interactions. The demonstrator we describe in the next sections showcases the potential of VR dialogue systems for experiential learning and social behavior training in the context of workplace sexism.

2 The Forum Theatre approach

We aim to design and deploy a *virtual reality* (VR) training tool intended to raise awareness of social discrimination (both ethnic and gender-based) and to train individuals to respond appropriately when witnessing such situations. The tool draws inspiration from the *forum theatre technique*, an interactive theatre method developed by Augusto Boal in the 1960s (Boal, 1972). Originally conceived as a form of popular education, forum theatre is now widely used to promote awareness of social issues such as discrimination and violence. It involves the dramatization of a problematic situation performed by actors, followed by active audience participation: spectators are invited to take the place of a character and explore alternative actions to change the course of events. This method not only raises awareness but also enables participants to *develop and practice concrete strategies* they can later apply in real-life contexts.

109

Figure 1: Screenshot of a VR scene depicting a discriminatory situation enacted by two autonomous virtual characters

Figure 2: Screenshot of a VR scene in which the user has the task to identify the different discriminatory behaviors observed in the previous scene through natural langage dialog with a virtual character.

3 System overview

Within this framework, our demonstrator consists of three main stages. In a first step, the user is immersed, via a VR headset, in a simulated scene depicting a discriminatory situation enacted by two autonomous virtual characters (Figure 1). This corresponds to the **Witness Stage**. In a second step, the user interacts with in natural langage with a *virtual character* to identify and analyze the different problematic behaviors observed in the scene. Through dialogue, the character helps the user identify situations of discrimination that he may have observed in the scene played out by the two virtual actors, reminding him of different moments in the scene and pointing out problematic behaviours (Figure 2). This is the **Discussion Stage**. In a final step, the user replays the same scene in VR simulated by the two virtual actors, but in the step, the user embodying the discriminated character, and attempts to respond to the discriminatory behavior of the virtual actor in order to resolve or transform the situation. This is the **Confrontation Stage**.

Various interactive scenarios can be simulated, encompassing different forms of social discrimination. In the proposed demonstration, we focus on *ordinary sexism* occurring in diverse social contexts (e.g., hierarchical relationships, mixed-gender interactions) and expressed through varying socio-emotional behaviors of virtual agents (e.g., conciliatory vs. aggressive attitudes). *Ordinary sexism* is defined as "stereotypes and collective representations that translate into words, gestures, behaviors or actions that exclude, marginalise or inferiorize women" (Grésy, 2009); for example, sexist remarks and jokes or devious seduction.

The virtual actors are integrated into a VR platform that simulates social interactions. As an initial step, we use a collected *corpus of forum theatre performances on discrimination* described in (Ochs et al., 2023). Motion capture was used in this corpus to recreate realistic scenes with virtual characters and to model the behavior of discriminatory agents (Figure 4). This behavioral modeling enables the virtual actors to reproduce discriminatory attitudes dynamically during interaction with the user.

4 Dialogue Management and Interaction Strategy

Our system adopts a *mixed-initiative dialogue management strategy* combining predefined sequences with open-ended interaction via a *Large Language Model* (LLM) API. Predefined sequences include replayed discrimination scenes, system explanations, and recovery mechanisms for sensitive user inputs. During open-ended dialogue, a *dialogue state representation* tracks discussion topics, scene progression, and critical events.

4.1 General Architecture

As shown in Figure 3, the system integrates an *Automatic Speech Recognition* (ASR) module, a *dialogue controller*, a *text-to-speech synthesizer*, a *VR animation manager*, and an *LLM-based chatbot*. We use the *Google Speech API* for ASR, *inworld-tts-1-max*[1] for speech synthesis, the *Unity VR engine*[2] for animation, and the *Gemini-2.5-flash* LLM for dialog turn generation.

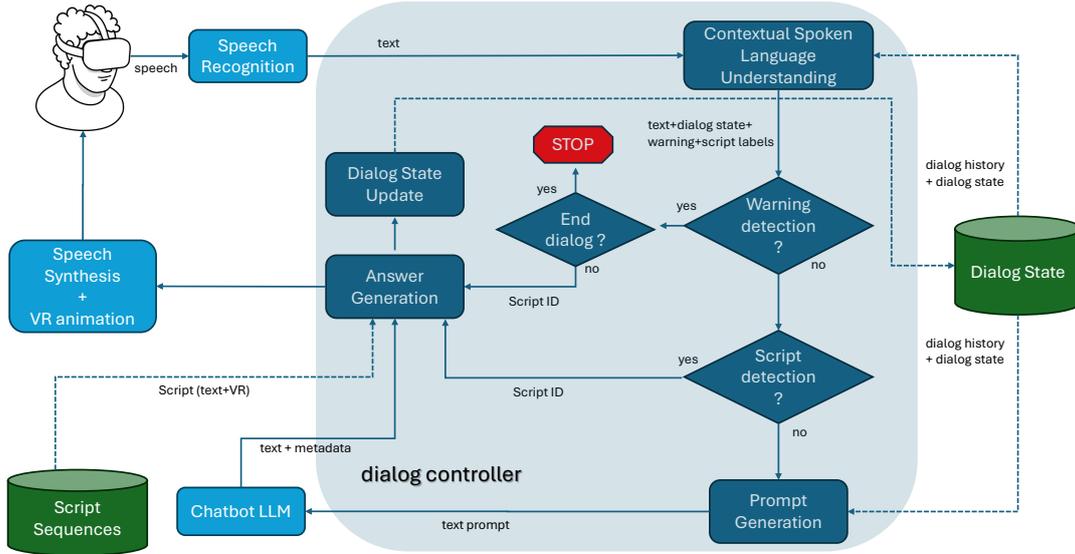The *dialogue controller* combines state-based management with open conversational capabilities,

---

[1] https://inworld.ai/

[2] https://unity.com/solutions/vr

Figure 3: Architecture of the dialogue controller.

ensuring both robustness and flexibility in handling sensitive social interactions.

4.2 Dialogue Controller

LLM-based agents face two main challenges: maintaining long-term coherence and preventing harmful or inappropriate behavior. Previous studies have addressed these issues through memory abstraction techniques (Lee et al., 2023; Seo et al., 2025) and output constraints (Rebedea et al., 2023). Our approach integrates both through a dynamic *prompt chaining mechanism* and an explicit *dialogue state representation* guiding conversation flow and goal completion.

Our approach draws inspiration from these works by integrating both a dynamic *prompt chaining mechanism*, which continuously updates the LLM prompt history to guide the conversation, and an explicit *dialogue state representation* that supervises dialogue flow and ensures all goals are fulfilled before the conversation ends.

In addition to these enhanced memory methods, we implemented a *problematic situation detector* and a *contextual scripted dialogue router*, which can temporarily override LLM control when necessary. These modules are described below.

Dialogue State - This data structure maintains both the dialogue history and the set of goals achieved during the interaction. The active goal set depends on the current stage (Witness, Discussion, Confrontation) and consists of references to spe-

cific situations that must be addressed before the dialogue can be concluded. For instance, during the Discussion stage, the user is expected to comment on each discriminatory situation identified in the Witness stage. Each situation is represented by a unique label, and the dialogue state structure tracks which labels have already been discussed and which remain pending.

Contextual Spoken Language Understanding (CSLU) - This module analyzes the automatic transcription of user speech in relation to the current dialogue state. It provides four primary functions: (1) detecting discriminatory situations based on the current dialogue stage and updating the dialogue state to reflect newly discussed situations; (2) triggering predefined scripted sequences in response to the dialogue state and user input; (3) detecting the completion of a dialogue stage and returning control to the main VR menu system; and (4) identifying problematic situations originating from either the user or the LLM, and initiating corrective scripted sequences or safely terminating the interaction.

Scripted Sequences - This database contains predefined VR sequences associated with each stage. For instance, during the *Witness Stage*, the sequence is fully scripted, as will be described in the next section. In the *Discussion Stage*, scripted sequences correspond to explanatory interventions about specific discrimination situations, including legal implications or contextual clarifications.

Prompt Generation - When the CSLU module allows the LLM to generate the next system utterance, this module constructs the prompt containing instructions for the LLM API. The LLM is expected to return a JSON structure with two fields: (1) a text output corresponding to the avatar's speech; and (2) an array, possibly empty, containing dialogue state labels identified in the current conversational step. The prompt itself consists of four parts: (1) a description of the persona played by the LLM; (2) a description of the current dialogue stage, including the set of possible dialogue states, their labels, and examples; (3) general behavioral guidelines for the persona; and (4) stage-specific instructions summarizing the dialogue history and directing the next conversational goal based on the dialogue state representation.

Answer Generation - Whether the response originates from a scripted sequence or an LLM output, this module generates the corresponding VR instructions for speech synthesis and avatar movement.

Dialogue State Update - At each interaction step, the dialogue state representation is updated based on the selected scripted sequence or the processed LLM output.

5 Use Case: Ordinary Sexism in Professional Situations

The developed training tool is illustrated through a scenario depicting *ordinary sexism* in the workplace. The scene involves a conversation between a male supervisor and his female employee regarding an important assignment. Although the supervisor intends to entrust her with the task, he expresses doubts about her ability to lead a team. The scenario was designed based on descriptions of ordinary sexism in professional settings (Grésy, 2009, 2015), incorporating common behaviors such as: *condescension and paternalism* (e.g., "my sweetheart"), *denigration* (e.g., "you don't have the capacity for this job"), *indirect seduction* (encouraging stereotypical femininity), *maternity offence* (e.g., "it's not the right time for another child"), *part-time bias* (e.g., "you're lucky you're not doing anything tomorrow, Wednesday"), and *sexist remarks or jokes*.

Two professional actors from the company NextLevel performed the scene (Figure 4). Both have extensive experience in forum theatre, a tech-

Figure 4: Extract from the recorded corpus showing face-to-face interaction with discriminatory behaviors.

nique frequently used for corporate training on discrimination awareness.

Based on motion capture data from the recorded corpus, body animations were created using *iClone* and integrated into *Unity*. Facial expressions were extracted from the videos using *OpenFace* (Amos et al., 2016). In Unity, we developed two main animation libraries, *talking* and *listening*, and designed additional *idle* movements. These behaviors were derived from the corpus, enhanced with Mixamo assets, and informed by our previous work on virtual character animation. In total, we implemented 8 idle, 10 listening, 9 talking, 19 facial, and 16 head movement animations.

Animations are layered (body, face, and head) to generate varied full-body behaviors. Their selection (talking, listening, or idle) depends on *animation state* variables in Unity, automatically updated by the *dialogue controller*. A lightweight algorithm ensures animation variability, while lip movements are synchronized with speech using the *SALSA lipsync* tool.

6 Conclusion

We present a VR training system that combines a Large Language Model–driven dialogue controller with realistic virtual characters to address everyday sexism in professional contexts. The system relies on a hybrid architecture that integrates scripted sequences with open-ended dialogue, balancing interaction flexibility with control in sensitive exchanges.

The proposed use case depicts a workplace scenario inspired by real instances of ordinary sexism. Future work will extend the approach to other forms of social bias and include initial evaluations of user engagement and learning outcomes based on post-interaction surveys.

References

Brandon Amos, Bartosz Ludwiczuk, Mahadev Satyanarayanan, et al. 2016. Openface: A general-purpose face recognition library with mobile applications. *CMU School of Computer Science*, 6(2):20.

Keith Anderson, Elisabeth André, Tobias Baur, Sara Bernardini, Mathieu Chollet, Evi Chryssafidou, Ionut Damian, Cathy Ennis, Arjan Egges, Patrick Gebhard, et al. 2013. The tardis framework: intelligent virtual agents for social coaching in job interviews. In *International Conference on Advances in Computer Entertainment Technology*, pages 476–491. Springer.

Augusto Boal. 1972. Catégories du théâtre populaire. *Travail théâtral*, 6:3–26.

Merijn Bruijnes, Jeroen Linssen, and Dirk Heylen. 2019. Special issue editorial: Virtual agents for social skills training.

Leonardo Campillos-Llanos, D. Bouamor, Éric Bilinski, Anne-Laure Ligozat, Pierre Zweigenbaum, and S. Rosset. 2015. Description of the patientgenesys dialogue system. In *SIGDIAL Conference*.

Brigitte Grésy. 2009. *Petit traité contre le sexisme ordinaire*. Albin Michel.

Brigitte Grésy. 2015. Le sexisme dans le monde du travail, entre déni et réalité. Technical report, Rapport du Conseil supérieur de l'égalité professionnelle entre les femmes et les hommes solidaire.

Lynne Hall, Susan J Jones, Ruth Aylett, Elisabeth Andre, Ana Paiva, Gert Jan Hofstede, Arvid Kappas, Yukiko Nakano, and Toyoaki Nishida. 2011. Fostering empathic behaviour in children and young people: interaction with intelligent characters embodying culturally specific behaviour in virtual world simulations. *INTED2011 Proceedings*, pages 2804–2814.

Gibbeum Lee, Volker Hartmann, Jongho Park, Dimitris Papailiopoulos, and Kangwook Lee. 2023. Prompted LLMs as chatbot modules for long open-domain conversation. In *Findings of the Association for Computational Linguistics: ACL 2023*, pages 4536–4554, Toronto, Canada. Association for Computational Linguistics.

Magalie Ochs, Jean-Marie Pergandi, Alain Ghio, Carine André, Patrick Sainton, Emmanuel Ayad, Auriane Boudin, and Roxane Bertrand. 2023. A forum theater corpus for discrimination awareness. *Frontiers in Computer Science*, 5:1081586.

D. Pautler, Vikram Ramanarayanan, Kirby Cofino, P. Lange, and D. Suendermann-Oeft. 2018. Leveraging multimodal dialog technology for the design of automated and interactive student agents for teacher training. In *SIGDIAL Conference*.

Alexander Prange, Margarita Chikobava, P. Poller, Michael Barz, and Daniel Sonntag. 2017. A multimodal dialogue system for medical decision support inside virtual reality. In *SIGDIAL Conference*.

Traian Rebedea, Razvan Dinu, Makesh Narsimhan Sreedhar, Christopher Parisien, and Jonathan Cohen. 2023. NeMo guardrails: A toolkit for controllable and safe LLM applications with programmable rails. In *Proceedings of the 2023 Conference on Empirical Methods in Natural Language Processing: System Demonstrations*, pages 431–445, Singapore. Association for Computational Linguistics.

Seongbum Seo, Sangbong Yoo, and Yun Jang. 2025. A prompt chaining framework for long-term recall in llm-powered intelligent assistant. In *Proceedings of the 30th International Conference on Intelligent User Interfaces*, pages 89–105.