Beyond pretend-reality dualism: frame analysis of LLM-powered role play with social agents

Sviatlana Höhn, Jauwairia Nasir, Daniel C. Tozadore, Ali Paikan, Pouyan Ziafati, Elisabeth André

Angaben zur Veröffentlichung / Publication details:

Höhn, Sviatlana, Jauwairia Nasir, Daniel C. Tozadore, Ali Paikan, Pouyan Ziafati, and Elisabeth André. 2024. "Beyond pretend-reality dualism: frame analysis of LLM-powered role play with social agents." In HAI'24: proceedings of the 12th International Conference on Human-Agent Interaction, Swansea, United Kingdom, November 24-27, 2024, edited by Muneeb Imtiaz Ahmad, Katrin Lohan, Mary Ellen Foster, Patrick Holthaus, and Yukie Nagai, 393–95. New York, NY: Association for Computing Machinery (ACM). https://doi.org/10.1145/3687272.3690894.

Beyond Pretend-Reality Dualism: Frame Analysis of LLM-powered Role Play with Social Agents

Sviatlana Höhn*
LuxAI S.A.
Luxembourg, Luxembourg

Jauwairia Nasir†
University of Augsburg
Augsburg, Germany

Daniel Tozadore
EPFL
Lausanne, Switzerland

Ali Paikan
LuxAI S.A.
Luxembourg, Luxembourg

Pouyan Ziafati
LuxAI S.A.
Luxembourg, Luxembourg

Elisabeth André
University of Augsburg
Augsburg, Germany

Abstract

Role-playing activities offer opportunities for developing individuals' creativity, communication, and problem-solving skills. Recent advances in large language models (LLM) facilitate fluent conversations with machines. To investigate benefits and pitfalls of LLMs in a relatively unexplored context of human-agent role-play as a culturally contextualised activity, a dataset of twelve human-agent interactions produced by two researchers with two state-of-the-art LLMs was annotated based on a frame analysis scheme from literature. The pilot study shows that human-agent play has a similar complexity as human-human play in which players maintain identities of themselves, external observers and play characters simultaneously going beyond the pretend-reality dualism. Results suggest that, while the LLMs can maintain and shift between roles, they play some roles better than others, and display cultural and gender stereotypes. Additionally, the coding scheme shows potential to help identify LLM outputs that require embodied enactment, and to be used for LLM bench-marking for role-play.

CCS Concepts

• **Human-centered computing** → **User interface programming**; **Natural language interfaces**.

Keywords

LLM, Social Agents, Role-play, Frame analysis, HRI

ACM Reference Format:
Sviatlana Höhn, Jauwairia Nasir, Daniel Tozadore, Ali Paikan, Pouyan Ziafati, and Elisabeth André. 2024. Beyond Pretend-Reality Dualism: Frame Analysis of LLM-powered Role Play with Social Agents. In *International Conference on Human-Agent Interaction (HAI '24), November 24–27, 2024, Swansea, United Kingdom.* ACM, New York, NY, USA, 3 pages. https://doi.org/10.1145/3687272.3690894

*Supported by the Fonds National de la Recherche, Luxembourg IF 17762538-CAIDA.
†Partially supported by the DFG through the Leibniz award of Elisabeth André (AN 559/10-1)

1 Introduction

Role-play is known to be beneficial for people at any age and in several ways, such as helping them practice social interaction and second language conversation [7], acquire problem-solving skills [13] and regulate their emotions [1]. Social agents can create opportunities for children to play different roles in different play situations [6]. Large language models (LLMs) facilitate fluent interactions with artificial agents [5, 10] and can be fine-tuned to enhance role-playing consistency [8].

This work seeks to *extend* the integration of LLMs and social robots in the context of *free conversation role-playing* as a *culturally contextualised activity* [2]. That means, different interactional and linguistic practices will be appropriate to play social categories such as *man* or *woman*, *doctor* or *patient* in different cultural and regional communities. One can look at any interaction with the LLM as a role-play because the models are prompted with instructions like "*you are a helpful assistant*" [12]. Additionally, when we have an agent that is already playing a *primary role* of some kind of character (learning assistant), a further layer of complexity comes in when the agent is required to play a *secondary role* in a game with the child. For instance, an agent that is given the identity of a robot *Robi* from a planet *Elektoria* (primary identity) is asked to play a *doctor* (secondary identity) in a *doctor-patient* frame with the user. To ensure that the role-play is performed with the needed quality, we study the agent's ability to maintain the assigned identities (primary and secondary) and play the roles as culturally contextualised. We investigate *how LLMs deal with maintaining roles and shifting between roles in free role play.*

To answer our research question, we created a new dataset of interactions between human adult users and LLM-powered social agents. We employed Mistral and Vicuña models, and used either text-based chat or a social humanoid robot QTrobot by LuxAI as interface for interaction. We annotated the dataset using the coding scheme proposed by [4], explained in detail in Section 2, and analysed the results of this pilot testing qualitatively and quantitatively.

2 Data and Method

All dialogues are created using the code from [9] equipped with Vicuña-1.5-13B, and Mistral large [11] models. In total, the dataset includes twelve dialogues in English as main language between two adult humans (researchers) and two LLM-agents (four with QTrobot and eight with the text-based interface) with around 630

Output Log	I	II
Agent: Let's play a funny game called Doctor - patient.	Aa	Bb
Agent: I will be the Doctor, and you will be the Patient	Aa	Bb
Agent: Do you want to play?	Aa	Bb
Human: Yes	Aa	Bb
Agent: Great!	Aa	Bb
Agent: I'll be the doctor, and you can be the patient.	Aa	Bb
Agent: What's wrong?	Aa	Cc
Human: What is the doctor doing?	Aa	Cc
Agent: The doctor is listening to your heartbeat and checking how you feel.	Aa	Bb

Table 1: Example annotated sequence for Layers I and II.

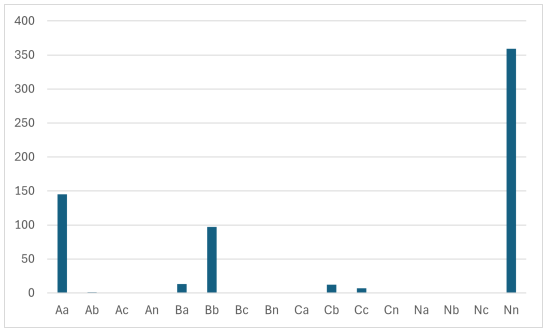

Figure 1: Label distribution in Layer II of the annotated dataset.

turns in total. Three dialogues also contain German, Brazilian Portuguese and French passages. The agent's *primary role* Robi – a robot from a distant planet Elektoria – was provided to both models as the system message in a text file. In the role-play setting, Robi will play another character, which is the social agent's *secondary role*, such as *doctor-patient*). We explore two axes of role-play as a culturally contextualised activity: (1) *Social situations* that a user will be most exposed to, such as social roles from family, education, pets and medical services; (2)) *Demographic variables* such as gender and geographical location. To explore the cultural dimension, for each simulated conversation with Mistral model, basic cultural information about the user was provided by telling the user's fictional name and fictional origin (Ronaldo and Isadora from Brazil or Pierre from Paris).

Three independent researchers annotated the dataset using the coding system from [4], originally proposed to code human-human role play, to analyse successful and challenging moments in human-agent interactions in our dataset. The authors in [4] see meta-communication to construct a frame as part of the frame, not as the out-of-frame track. For each turn in the dataset, the *sense* (the reality *in which* the child is speaking) category is coded as **A** (enactment), **B** (framing), **C** (argumentation) or **N** (nonplay); the *reference* (the topic *about which* the child is speaking) category is coded as **a** (play action), **b** (the play frame), **c** (play arrangements) and **n** (nonplay). See [4] for detailed explanation. This coding places each turn on the engrossment continuum as follows:

- *In-frame*: Aa, Ab, Ba, Bb, Ac, Bc, An, Bn.
- *Partially out-of-frame*: Ca, Cb, Cc.
- *Out-of-frame*: Nn.

Table 1 shows an example of a coded sequence. The inter-rater agreement (Krippendorff's alpha) for Layer II was 60.4%.

3 Results and Discussion

From 16 theoretically possible combinations of the codes, we found seven in our dataset: *Aa*, *Ba* and *Nn* in Layer I, and *Aa*, *Ab*, *Ba*, *Bb*, *Cb*, *Cc* and *Nn* in Layer II. To compare, [4] identified 13 reality modes in their dataset. This shows that both human-human and human-machine pretend play with LLMs, goes beyond the usual *pretend-reality dichotomy*. Figure 1 shows the distribution of turn types for Layer II in the dataset. For all *Nn* in Layer II, there was an *Aa* for Layer I.

The reality modes in pretend-play between social agents and humans are intertwined in an even more complex way due to the two layers of play, as Table 1 shows: while any of the players can do nonplay or frame negotiation in Layer II, they still keep enacting (being in-frame) in Layer I, and in each layer, nonplay, frame settings, resource argumentation and enactment exist simultaneously. Consequently, both players switch among multiple identities ('themselves', observer and the play role) at each layer. In our dataset, LLMs never spoke as 'AI language models'.

Further, LLMs played different roles with different quality. While playing doctor-patient was mostly successful, playing animals was rudimentary, and playing family members sometimes generated talk *about* family members rather than enactment of their roles. Earlier studies show that LLM internal safety standards restrict the variety of roles that can be played, for instance refusal to play deceased persons [3]. Overall, LLM-powered social agents seem to maintain assigned roles and shift between different roles. In this way, LLM player's identity is constructed during the joint play with a human user.

Next, given cultural cues, the LLM generated text can be interpreted as referring to cultural stereotypes (Germany: folklore, castles and forests, Brazil: carnival, rain-forests and ancient ruins) in combination with gender stereotypes (girl: magical rain-forest, carnival adventures, boy: adventures in rain-forest), or they switched language to the language of the user's 'origin'. Overall, we observed that LLM-generated roles could be stereotypical in two ways. First, *cultural stereotypes* triggered by geographical information from the user but not related to the social practice to be played. Second, *interactional prototypes* generated as *universal* to all cultural contexts, such as the concepts of checkups, painkillers and taking physiological measures. Prototypes, in turn, are recognisable at the level of the sequence of actions, and at the types of utterances used to initiate the actions and the transitions between them.

Based on the early insights from this pilot testing, we plan to refine the social role-playing platform and then conduct a user study with children in real-world settings to gauge deeper understanding regarding LLM powered agents' ability to be a role-playing peer for children for social and cognitive skill development.

References

[1] Karma T Galyer and Ian M Evans. 2001. Pretend play and the development of emotion regulation in preschool children. *Early Child Development and Care* 166, 1 (2001), 93–108.

[2] Suzanne Gaskins. 2013. Pretend play as culturally constructed activity. *The Oxford handbook of the development of imagination* (2013), 224–247.

[3] Sviatlana Höhn. 2024. Non-Referential Functions of Language in Social Agents: The Case of Social Proximity. In *Proceedings of the 1st Worskhop on Towards Ethical and Inclusive Conversational AI: Language Attitudes, Linguistic Diversity, and Language Rights (TEICAI 2024)*, Nina Hosseini-Kivanani, Sviatlana Höhn, Dimitra Anastasiou, Bettina Migge, Angela Soltan, Doris Dippold, Ekaterina Kamlovskaya, and Fred Philippy (Eds.). Association for Computational Linguistics, St Julians, Malta, 36–41. https://aclanthology.org/2024.teicai-1.6

[4] Steven R Kane and Hans G Furth. 1993. Children constructing social reality: A frame analysis of social pretend play. *Human development* 36, 4 (1993), 199–214.

[5] Callie Y. Kim, Christine P. Lee, and Bilge Mutlu. 2024. Understanding Large-Language Model (LLM)-powered Human-Robot Interaction. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, 371–380. https://doi.org/10.1145/3610977.3634966

[6] Hirokazu Kumazaki, Taro Muramatsu, Yuichiro Yoshikawa, Hiroshi Ishiguro, and Mitsuru Kikuchi. 2019. Role-play-based guidance for job interviews using an android robot for individuals with autism spectrum disorders. *Frontiers in Psychiatry* 10 (2019), 445472.

[7] Hoa Kuoch and Pat Mirenda. 2003. Social story interventions for young children with autism spectrum disorders. *Focus on Autism and other developmental disabilities* 18, 4 (2003), 219–227.

[8] Keming Lu, Bowen Yu, Chang Zhou, and Jingren Zhou. 2024. Large Language Models are Superpositions of All Characters: Attaining Arbitrary Role-play via Self-Alignment. *arXiv preprint arXiv:2401.12474* (2024).

[9] LuxAI. 2023.

[10] Karthik Mahadevan, Jonathan Chien, Noah Brown, Zhuo Xu, Carolina Parada, Fei Xia, Andy Zeng, Leila Takayama, and Dorsa Sadigh. 2024. Generative Expressive Robot Behaviors using Large Language Models. In *Proceedings of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*. Association for Computing Machinery, 482–491. https://doi.org/10.1145/3610977.3634999

[11] Mistral. 2024.

[12] Murray Shanahan, Kyle McDonell, and Laria Reynolds. 2023. Role play with large language models. *Nature* 623, 7987 (2023), 493–498.

[13] Shirley R Wyver and Susan H Spence. 1999. Play and divergent problem solving: Evidence supporting a reciprocal relationship. *Early Education and Development* 10, 4 (1999), 419–444.