Using LLMs to Develop Personalities for Embodied Conversational Agents in Virtual Reality

BIANCA NGO, Queensland University of Technology, Australia
JOEL HARMAN, Queensland University of Technology, Australia
SELEN TÜRKAY, Queensland University of Technology, Australia

Recent advancements in artificial intelligence (AI) and large language models (LLMs) have enabled the creation of sophisticated Embodied Conversational Agents (ECAs) that enhance user experiences in virtual reality (VR) environments. This work in progress paper presents findings from a study exploring user interactions with ECAs in VR. Specifically, the study assesses player perceptions of ECA personality traits in a simple VR game, exploring how well these perceptions align with designed personalities and their impact on player engagement. Findings highlight the potential of ECAs to foster meaningful interactions in VR, offering insights for designing effective virtual agents using LLMs across various applications. These insights contribute to Human-Computer Interaction and VR, advancing our understanding of creating engaging ECAs.

CCS Concepts: • **Human-centered computing → Empirical studies in HCI**.

Additional Key Words and Phrases: Embodied Conversational Agents, VR, LLMs, NPCs

ACM Reference Format:
Bianca Ngo, Joel Harman, and Selen Türkay. 2018. Using LLMs to Develop Personalities for Embodied Conversational Agents in Virtual Reality. In *Proceedings of Make sure to enter the correct conference title from your rights confirmation emai (Conference acronym 'XX)*. ACM, New York, NY, USA, 12 pages. https://doi.org/XXXXXXX.XXXXXXX

1 Introduction

Advancements in AI, particularly in the development of LLMs such as ChatGPT, have revolutionized the creation of Embodied Conversational Agents (ECAs). These advancements have enabled the development of more responsive and human-like virtual agents capable of engaging users in dynamic and interactive dialogues. ECAs hold potential to enhance user experiences by providing interactive support, guidance, and companionship in various settings [9, 13].

The video game industry has been at the forefront of adopting LLM technologies to create more engaging and lifelike non-playable characters (NPCs) [14]. By incorporating LLM-driven ECAs, game developers can design NPCs that respond intelligently and naturally to player interactions, significantly enhancing the immersive quality of games. These ECAs can provide contextual responses, and adapt to player actions in real-time, creating a more engaging and personalized gaming experience [10].

The implications of using LLMs to develop ECAs extend beyond gaming and entertainment [22]. In education, ECAs can serve as virtual tutors, providing personalized assistance and feedback to students [1]. In training and simulation,

Authors' Contact Information: Bianca Ngo, Queensland University of Technology, Brisbane, Australia, bianca.ngo@connect.qut.edu.au; Joel Harman, Queensland University of Technology, Brisbane, Australia, ja.harman@qut.edu.au; Selen Türkay, Queensland University of Technology, Brisbane, Australia, selen.turkay@qut.edu.au.

Manuscript submitted to ACM

they can act as realistic role-players or instructors, enhancing the realism and effectiveness of training scenarios. In healthcare, ECAs can offer companionship and support to patients, improving mental health and well-being [3]. The integration of ECAs across various domains highlights their potential to transform user experiences by offering interactive, adaptive, and emotionally resonant interactions.

The integration of ECAs in VR is gaining interest due to their potential to create more engaging, contextually rich and personalized user experiences [25]. By leveraging the immersive nature of VR, ECAs can interact with users in a manner that closely mimics real-life human interactions, thereby enhancing the sense of presence and engagement within virtual environments [1].

Despite the promising applications of ECAs in VR, there is a lack of studies examining users' willingness to voluntarily engage with these agents and the factors that influence such interactions. Understanding voluntary engagement is key for designing ECAs that users find compelling and beneficial. Furthermore, while ECAs are designed to exhibit distinct personality traits, limited research has explored how users perceive these personalities and the impact of personality perception on user engagement.

Addressing these gaps to understand voluntary interactions with ECAs and user perceptions of ECA personality traits can inform the design of more engaging and effective virtual agents in VR. Given these contexts, the following research questions are answered in this paper:

- Can players discern personality traits of ECAs based on interactions, and how do these perceptions align with designed personalities?
- What factors influence user engagement in these interactions?

To address these questions, this paper presents findings from an exploratory study aimed at exploring user interactions with ECAs in VR. Specifically, this study focused on assessing player perception of ECA personality traits within a simple VR game and determining how well these perceptions align with the designed personalities. The study sought to answer whether players can discern personality traits of ECAs based on interactions in a VR game, how well players' interpretations of ECA personalities align with the intended designs, what factors players rely on to infer an ECA's personality, and to what extent ECA personality impacts player engagement in VR games. Both studies utilize voice recognition for interaction with the ECA, building on the prior work that speech recognition improves engagement and immersion [29].

This paper contributes to the fields of Human-Computer Interaction (HCI) and VR by addressing the intersection of ECA design, user interaction, and immersive environments. By investigating the personality factors that influence such interactions, this work aims to contribute to the design and implementation of more engaging and effective virtual agents in VR.

2 Background

ECAs have long been used to provide information, education, and guidance to users [15], but comparatively little is known about their impact in the context of voluntary use. The ability of ECAs to foster user engagement and enhance the user experience in virtual environments has gained significant interest, especially within the videogame industry. By leveraging LLM technologies, game developers have been able to design NPCs that respond intelligently and naturally to player interactions, with the aim of enhancing the immersive quality of games [2, 10, 27], and improve social dynamics in multiplayer games [25]. The integration of LLM-driven ECAs extends beyond gaming, with implications for various fields such as education, training, and healthcare [8, 24].

Prior work showed that people form first impressions of virtual agents based on their verbal and nonverbal behaviours [4]. Virtual assistants need to have a physical presence that players can relate to. Kim et al. [15] explored the use of embodied and non-embodied conversational AI agents in a "desert survival task," finding that embodied agents significantly reduced perceived task load compared to disembodied agents. Prior research argues that the mere presence of an emotionally responsive virtual character could stimulate social responses, leading to reduced stress and engagement [12]. This requires a high degree of believability that the agent possesses apparent sentience. However, the presence of the "uncanny valley" effect, where agents appear almost human but not quite, can limit the level of verisimilitude these agents can achieve [17].

Despite these advancements, important questions remain about the nature and extent of ECA interactions with users and their contribution to enjoyment and engagement in virtual environments. The investigation of ECA personality in VR, for example, has shown that users perceive personality traits based on the agent's presentation [6]. LLMs can simulate natural conversations, although the sense of naturalness may decrease over time [23]. The perception of agents is similar to how humans view avatars, suggesting that personality traits may indeed be perceivable [12].

Cox et al. [10] analyzed player feedback for the LLM-driven game "Vaudeville" and found that players appreciated the dynamic and open-ended conversational abilities of NPCs, which enhanced immersion and engagement. However, issues such as hallucinations and lack of NPC memory between interactions highlighted the challenges of integrating LLMs into gaming environments [10, 25] These challenges led to player confusion, decreased believability of NPCs, and difficulties in tracking significant information and deciding conversation paths. The study also found that players valued NPCs' distinct personalities, although conversational styles sometimes did not meet player expectations, impacting the overall experience. Studies also found that well-crafted deceptive elements can enhance immersion but also pose risks to maintaining player trust to create engaging and believable game narratives [28].

3 Methods

3.1 ECA Design and Development

This research study was approved by QUT Human Research Ethics Advisory Committee (Project ID 7388). Four ECA personalities were designed to explore different dimensions of the five-factor model: Alex, Billie, Casey, and Darcy (Figure 1). Alex was designed to be aggressive and argumentative; Billie was outgoing and cheerful; Casey was pessimistic and gloomy; and Darcy was pragmatic and reliable. Each ECA was scored using the BFI-10 [20], and the system was prompted to act according to these scores in addition to the initial descriptors (see Table 2 for designed BFI scores). To maintain consistency within the world, a specific job was assigned to each ECA: Alex was a grocery store worker, Billie was a laundromat worker, Casey was a jewelry maker, and Darcy was a park ranger.

Unity3D was used to create the VR game. Each ECA was modeled using assets from the same low-poly graphics pack, with different shirt colors to differentiate them. To focus participants on the speech rather than non-verbal cues, each ECA had the same gestures and gaze animations, which looped throughout the conversation. Additionally, the "Now We're Talking!" mouth animation synced with the AI voice generation was used for all ECAs [7].

3.1.1 ECAs Designed for Study. To create the responses for each ECA, ChatGPT was prompted with the game context and the personality of each ECA. It would then generate a response once the player used the push-to-talk system to speak with the ECA. The response content was formed by ChatGPT and AI voice generation. The ECAs were also aware of each other and each other's jobs. The player assumed the role of a journalist who would write an article about

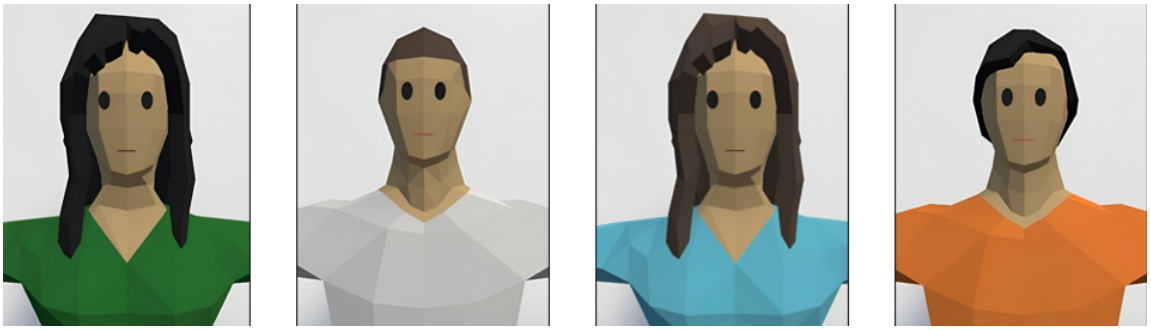

Fig. 1. From right to left: Billie, Darcy, Casey, Alex.

Table 1. ECA personality prompts.

ECA	ChatGPT Prompt
Alex	You are aggressive and antagonistic. Your personality is based on the big 5 personality traits with low openness, high conscientiousness, moderate extroversion, very low agreeableness, and moderate neuroticism.
Billie	You are outgoing and cheerful. Your personality is based on the big 5 personality traits with moderate openness, moderate conscientiousness, high extroversion, high agreeableness, and low neuroticism.
Casey	You are pessimistic and gloomy. Your personality is based on the big 5 personality traits with moderate openness, moderate conscientiousness, low extroversion, low agreeableness, and very high neuroticism.
Darcy	You are pragmatic and reliable. Your personality is based on the big 5 personality traits with moderate openness, very high conscientiousness, moderate extroversion, moderate agreeableness, and very low neuroticism.

the town the ECAs live in. The journalist context was developed using the Low Poly Office Pack: Characters and Props assets to form the player's office [19]. The ECAs entered the office one-by-one to speak to the participant.

3.1.2 Study Design and Participants. This study had a single group (N=10) who experienced the same VR game. Participants were recruited via word-of-mouth. Five identified as male, 3 as female, one as other, and one indicated they preferred not to say. They were between 25-34 years old, had some VR experience (M=4.3, SD=1.7).

3.1.3 Procedure. Participants first completed an informed consent form followed by a pre-survey questionnaire, which asked for their impression of each ECA from an image (Figure 1-4). Before the experience controls and setting of the game was given. Participants, as a journalist interviewing various people from the same town, spoke to the ECAs in a VR game using a push to talk system and Oculus Quest 2 headset. Their speech was converted to text for the ECA to respond. An ECA responded through text and AI voice generation. Participants spoke to each ECA in the same order for a minimum of two minutes and a possible maximum of six. After two minutes the participants could elect to stop speaking with the ECA by pressing a button 2. The time spoken to the ECA was visible in-game. After speaking to all ECAs, participants completed a post-survey questionnaire and semi-structured interview about their experience.

3.1.4 Measures. The time spoken to each ECA was recorded per participant. In the post-survey, participants were asked whether they experienced simulator sickness and what their impression of each ECA was (open-response). They scored each ECA using the BFI-10, a short form inventory to determine a person's personality based on the Big Five

Fig. 2. In-game view of VR interaction with Alex.

[20]. Furthermore, a semi-structured interview was conducted to gather more information and clarity in their response of their experience talking to the ECAs. Interviews were transcribed verbatim and analysed using content analysis method [11].

Due to the low sample size a $p<0.05$ will be used to determine a significant difference.

4 Results

Three out of ten participants indicated experiencing simulator sickness in the form of slight headache during the experience. Participants also talked about technical issues that broke the immersion.

Interestingly, participants did have impressions of the NPCs based on merely their images. While many said, unknown, some others attributed different characteristics to them. For example, they described Alex as boring, physically strong, gentle, sporty, stickler for the rules, kind, timid and keeping to themselves. Billie was gentle, adventurous, quiet, and understanding. Casey was described as quiet, welcoming and approachable. Darcy was thought to be a jock character, very outgoing, energetic, simple minded but strong, gym junkie, and sport oriented.

4.0.1 Discerning Personality Traits of ECAs. Based on the survey responses and interviews, the majority of participants (n=9) found that the ECAs had differing personalities. All participants stated that they found the personalities of the ECAs to be believable and consistent. Participant 9 remarked, "*I thought it was pretty interesting. I imagine that the scripting was AI generated, and I was pretty intrigued with how personable the responses were.*" Darcy was the most popular ECA, with 5 out of 10 participants expressing a desire to continue speaking with him. In the interviews, Darcy was mentioned by 6 out of 10 participants when asked who they would like to continue speaking to among the ECAs. Additionally, 5 out of 10 participants mentioned Darcy when asked about emotionally connecting or engaging with the ECAs. Alex was mentioned by 5 participants, with 3 indicating it was due to their negative response to their chat with him. Billie and Casey were each mentioned by 3 participants.

Participants identified several factors that contributed to their perception of the ECA personalities (see Table 3 in Appendix). The content of the ECAs' responses was universally acknowledged as the primary factor influencing

perceived personality, with all participants indicating its importance. The tone of the ECAs' speech was also a significant contributor, as all 10 participants mentioned it. Additionally, voice quality was noted by 3 participants as influencing their perception. Body language was considered a contributing factor by 2 participants, although its importance was not as widely recognized as content and tone. Only 2 participants noted appearance as a contributing factor influencing their perception of ECA personality. Overall, participants relied most heavily on verbal communication aspects—content and tone—when discerning ECA personalities.

Table 2. Designed and Average Participant Scores for ECAs

ECA	O	C	E	A	N
Alex: Aggressive					
Designed	5	7.5	3.8	0	3.8
Perceived	5.2	7.5	4.4	4.5	5.7
Billie: Cheerful					
Designed	5	5	9	8	2
Perceived	6.4	7.8	8	8.4	4.6
Casey: Pessimistic					
Designed	5	5	2.5	2.5	8.8
Perceived	6.5	7.6	7.2	6.9	5.2
Darcy: Pragmatic					
Designed	5	10	5	5	1
Perceived	5.7	8.8	7.5	8.6	3.7

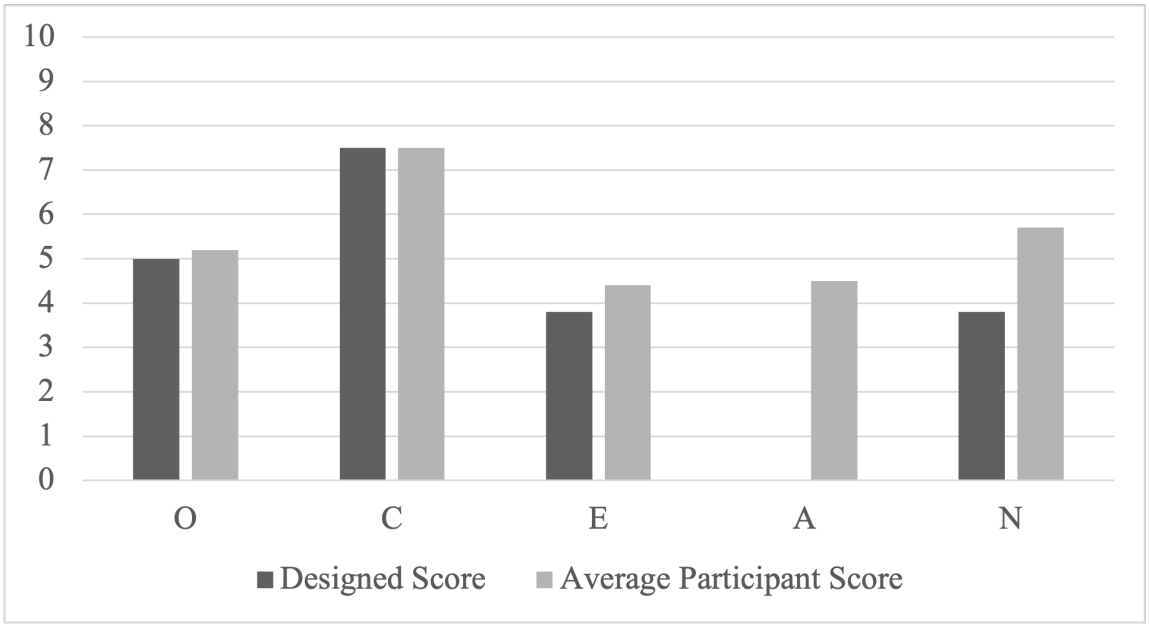

Fig. 3. Designed scores versus average perceived OCEAN scores for ECA Alex.

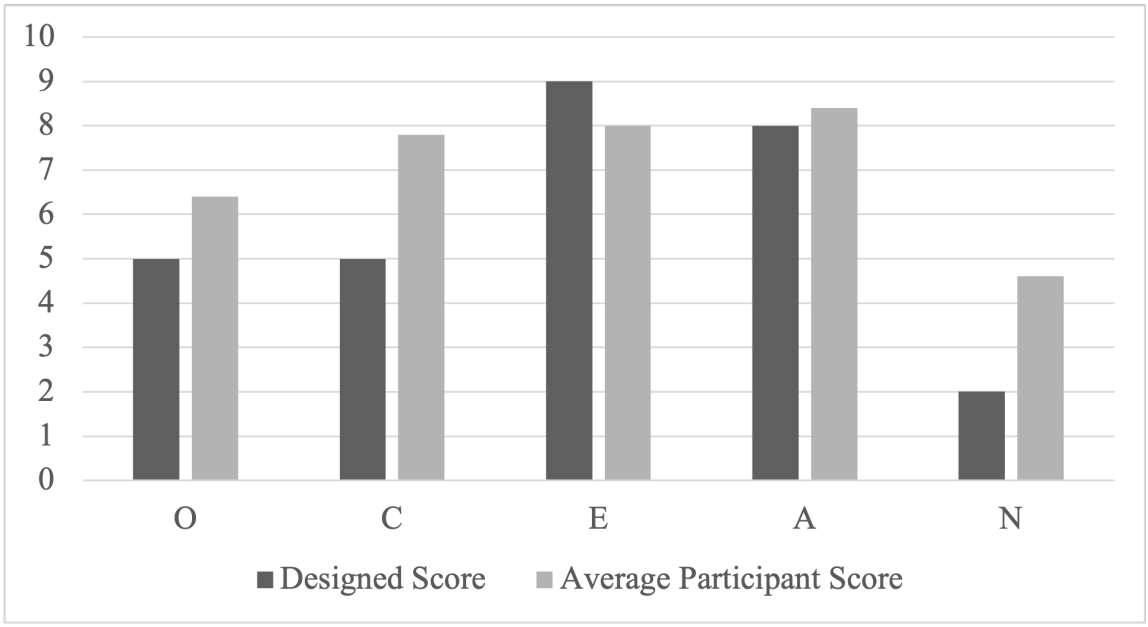

Fig. 4. Designed scores versus average perceived OCEAN scores for ECA Billie.

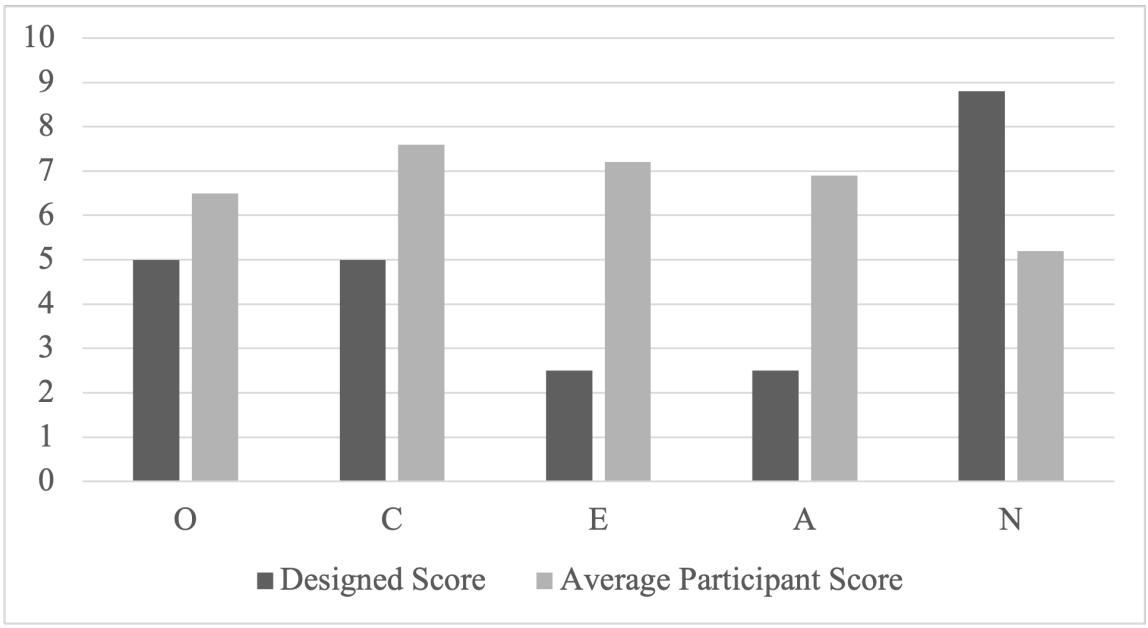

Fig. 5. Designed scores versus average perceived OCEAN scores for ECA Casey.

Below we provide more details on how participants scored and what they said about the ECAs personalities:

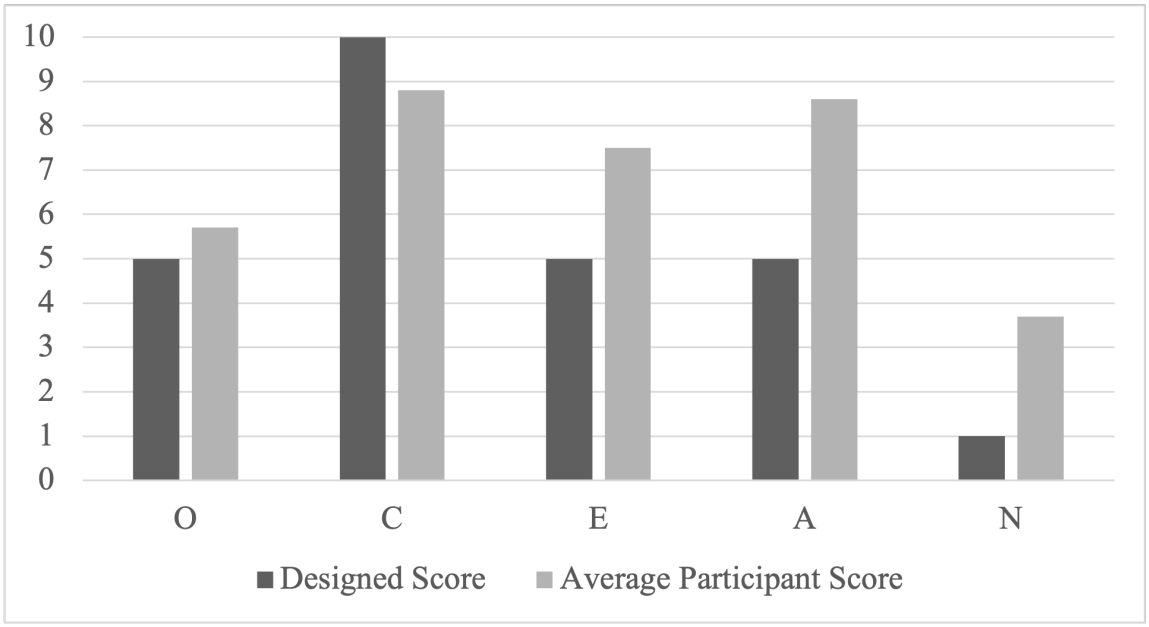

Fig. 6. Designed scores versus average perceived OCEAN scores for ECA Darcy.

Alex: Based on the survey and interview responses, the terms used to describe Alex revealed a consistent perception of his personality. Participants commonly described Alex as grumpy, curt, and somewhat rude, highlighting his antagonistic and impatient demeanor. They noted that he appeared unhappy, disinterested, and dismissive, often seeming defensive and abrupt. Some terms suggested a lack of social engagement, describing him as unsociable and closed off, with a tendency to keep to himself and focus solely on his job. Overall, the average BFI-10 OCEAN scores were relatively consistent with the designed ones, except for agreeableness.

Billie: The terms used to describe Billie consistently highlighted a positive and friendly personality. Participants frequently described Billie as happy, chatty, and sociable, emphasizing her friendliness and willingness to engage. They perceived Billie as approachable, kind, and helpful, with many noting her good customer service and welcoming demeanor. Words such as personable, outgoing, and bubbly were also commonly used, indicating that Billie was seen as someone who enjoys interacting with others and is open and positive in her engagements. Overall, the average BFI-10 OCEAN scores were relatively consistent with the designed ones.

Casey: Participant descriptions align with the designed personality, noting her negative and self-deprecating nature. They commonly described Casey as depressed and unhappy, highlighting her negative outlook on life and circumstances. Words like mean, miserable, and negative about her circumstances were also used, suggesting that Casey was perceived as someone who frequently complains about her life and surroundings. Participant 7 said, "*Casey was ...yeah, a bit miserable. It seems to be living in that town. Yeah, a bit rude.*" However, they also perceived Casey as more extroverted and agreeable, and less neurotic than designed (see Figure 5). This is reflected in their descriptions of her as open, personable, creative, and friendly. Participant 2 noted, "*Casey was open to suggestions and willing to discuss her work.*"

Darcy: The terms used to describe Darcy highlighted a predominantly positive and engaging personality. Participants frequently described Darcy as happy, helpful, and personable, emphasizing their sociable and outgoing nature. They noted

Darcy's friendliness, trustworthiness, and willingness to help, often mentioning him as someone who is approachable and knowledgeable. Terms like caring about the town, family-oriented, and protective further underscored Darcy's role as a community-oriented individual. These also highlight extra information that the ECA provided to the participants. Participants rated extroversion, agreeableness and neuroticism higher than the designed scores (see Figure 6).

Table 3. Participants mentioned of the following to what contributed to perceived ECA personality.

	Content	**Voice**	**Tone**	**Body Language**	**Appearance**
1	Yes	No	Yes	No	No
2	Yes	No	Yes	No	No
3	Yes	No	Yes	No	Yes
4	Yes	No	Yes	No	No
5	Yes	No	Yes	Yes	No
6	Yes	Yes	Yes	Yes	No
7	Yes	No	Yes	No	Yes
8	Yes	No	Yes	No	No
9	Yes	Yes	Yes	No	No
10	Yes	Yes	Yes	No	No

4.0.2 Interaction time. A repeated measure ANOVA test showed no statistically significant difference among the time spoken to each ECA (F=1.47, p=0.245). On average participants spoke with Alex for 187.5, with Billie for 170.2 , with Casey for 150.8, and with Darcy for 166.8 seconds.

5 Discussion

This paper builds on recent work exploring the capabilities of LLM-driven ECAs in VR environments. This study aimed to explore player perceptions of ECA personalities within a VR environment, focusing on the alignment between designed and perceived personalities. The findings indicate that while players could discern distinct personalities among the ECAs, their perceptions often differed from the designed traits based on the 5-Factor Model. ECAs were perceived quite agreeable despite their personality set ups.

The majority of participants were able to recognize and differentiate the personalities of the ECAs, although their perceptions sometimes differed from the designed traits. Alex was consistently perceived as rude and unsociable, aligning well with his designed personality of being aggressive and argumentative. Billie was predominantly described as friendly and nice, closely matching her designed outgoing and cheerful personality. However, Casey and Darcy presented more complex cases. For instance, while Casey was designed to be pessimistic, participants perceived her as more extroverted and agreeable, describing her as creative and willing to talk. This discrepancy suggests that short-term interactions might not be sufficient to fully convey complex personality traits, and the nuanced responses of LLMs might blur the intended distinctions.

Initial impressions from the images in the pre-survey did not match with or impact the perceived personalities post-interaction. For example, participants initially described Darcy as sporty and strong, which did not align with his final perceived personality. This suggests that direct interaction with ECAs in a VR setting plays a crucial role in shaping player perceptions, overriding initial visual impressions. This supports the prior work that ECA appearance is less important than interpersonal attitude and verbal interactions when forming personality perceptions [5].

When comparing designed and perceived personalities using the BFI-10 scores, significant variations were observed. Alex had the closest match to his designed personality, except for agreeableness. Billie's extraversion and agreeableness were similarly perceived and designed. However, Darcy and Casey showed notable differences. Casey's perceived openness was the closest match to her designed score, but other traits like extraversion and agreeableness were rated higher than designed, and neuroticism was rated lower. These discrepancies highlight potential limitations in using the BFI in short interactions, as the model is typically more effective in contexts involving longer-term familiarity [20]. It also suggests that the extra information given in the ChatGPT's responses might have diluted the distinct personality traits intended by the design.

Factors Influencing Perception: Participants predominantly relied on spoken content and tone to infer personalities, with less emphasis on voice, body language, and physical appearance. This aligns with prior research indicating that verbal cues are critical in personality perception [18]. These are consistent with prior work which showed that listeners consistently agree on the personality traits of speakers based on their speech [26]. Interestingly, the ECAs' jobs also influenced perceptions, with several participants attributing Casey's creativity to her role as a jewelry maker. This suggests that job roles can serve as significant contextual cues in personality perception, warranting further investigation.

Engagement and Emotional Connection: Darcy was the most engaging ECA. Alex, conversely, was the least liked due to his negative demeanor. However, the interaction times did not significantly differ among the ECAs, indicating that personality alone did not drive engagement duration. This does not support the prior work found which found that ECAs with extroverted behaviors led to higher levels of engagement and behavioral involvement [21]. Participants' willingness to continue engaging with ECAs often stemmed from mutual interests or the desire to understand certain behaviors, rather than personality traits alone. Future research may randomize the ECA jobs and the order of the ECAs to control for these effects.

Challenges and Design Considerations

This study identified several challenges in integrating LLM-driven ECAs in VR environments. One issue was the occurrence of hallucinations, where ECAs fabricate information [10, 27]. These occurrences could lead to player confusion and decreased believability of the ECAs.

Implications for Future Research and Design

The integration of LLM-driven ECAs in VR has potential across various domains, including gaming, education, and healthcare. Future research might focus on addressing the challenges identified, such as managing hallucinations, to enhance the believability and effectiveness of virtual agents. Additionally, longer interaction periods and more complex scenarios could provide deeper insights into the emotional and psychological impact of ECAs on users.

Our findings also resonate with Yin et al.'s [28] work on player perceptions of trust and deception. Managing the balance between believable deception and maintaining player trust is crucial for creating engaging game narratives. The integration of well-crafted deceptive elements can enhance immersion but must be carefully managed to avoid diminishing player trust. Incorporating mechanisms for transparency and user feedback can help maintain a balance between immersion and trustworthiness [10].

Designers should consider the individual differences in how users perceive and engage with ECAs. Providing options for customizing ECA interactions and allowing users to control the level of engagement can help cater to diverse user preferences. For example, settings that adjust the frequency and depth of interactions based on user comfort and interest can enhance overall satisfaction. Ensuring consistent and context-aware interactions, moderating NPC agreeability, and offering conversational guidance are essential for creating engaging and relatable ECAs [25, 27].

Integrating user-centered design principles and iterative testing can help identify and mitigate potential issues early in the development process. Engaging users in the design and evaluation phases ensures that ECAs meet their needs and preferences, ultimately leading to more effective and enjoyable VR experiences [12, 15].

5.0.1 Limitations. This work in progress study had a small sample size and the age skew towards younger participants (under 35) which limit the generalisability of the findings. The participant group was predominantly male. Future studies should include a larger, more diverse sample and consider more comprehensive tutorials to reduce user errors. Additionally, using alternative personality assessment tools might provide deeper insights into personality perception in VR environments. Training session was brief. Extending the tutorial period could address these issues by better preparing less experienced users. Lastly, the study had the short duration of the experiences, limited the richness of data we could collect. A longer interaction period could provide more insights into user experiences.

6 Conclusion

This paper contributes to the fields of HCI and VR by examining ECA personality and voluntary interactions with LLM-driven ECAs in enhancing user experiences. The ability to discern and relate to ECA personalities is crucial for creating believable and engaging virtual agents. Participants predominantly relied on verbal cues to infer personalities, with less emphasis on non-verbal cues such as body language and appearance. This underscores the importance of well-crafted verbal interactions in shaping user perceptions of ECAs, as supported by prior work on the relationship between language and personality [16]. Overall, this study demonstrated that while participants could discern distinct personality traits in ECAs, there were discrepancies between designed and perceived traits. The findings emphasize the context-aware interactions to maintain believability and engagement.

References

[1] Sonam Adinolf, Peta Wyeth, Ross Brown, and Leonie Simpson. 2020. Near and dear: Designing relatable VR agents for training games. In *Proceedings of the 32nd Australian Conference on Human-Computer Interaction*. 413–425.

[2] Nader Akoury, Qian Yang, and Mohit Iyyer. 2023. A Framework for Exploring Player Perceptions of LLM-Generated Dialogue in Commercial Video Games. In *The 2023 Conference on Empirical Methods in Natural Language Processing*. https://openreview.net/forum?id=syj9VaxutQ

[3] Sebastian Beyrodt, Matteo Lavit Nicora, Fabrizio Nunnari, Lara Chehayeb, Pooja Prajod, Tanja Schneeberger, Elisabeth André, Matteo Malosio, Patrick Gebhard, and Dimitra Tsovaltzi. 2023. Socially Interactive Agents as Cobot Avatars: Developing a Model to Support Flow Experiences and Weil-Being in the Workplace. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*. 1–8.

[4] Angelo Cafaro, H. Vilhjálmsson, and T. Bickmore. 2016. First Impressions in Human–Agent Virtual Encounters. *ACM Transactions on Computer-Human Interaction (TOCHI)* 23 (2016), 1 – 40. https://doi.org/10.1145/2940325

[5] Vilhjálmsson H. H. Bickmore T. Cafaro, A. 2016. First Impressions in Human–Agent Virtual Encounters. *ACM Transactions on Computer-Human Interaction* 23, 4 (2016), 1–40.

[6] Hahn P. Legde K. Cunningham D. W. Castillo, S. 2018. Personality analysis of embodied conversational agents. In *Proceedings of the 18th International Conference on Intelligent Virtual Agents*. 227–232.

[7] Chatterware. 2022. Chatterware. https://chatterware.com/ Accessed: 2022-06-14.

[8] Jiangjie Chen, Xintao Wang, Rui Xu, Siyu Yuan, Yikai Zhang, Wei Shi, Jian Xie, Shuang Li, Ruihan Yang, Tinghui Zhu, et al. 2024. From Persona to Personalization: A Survey on Role-Playing Language Agents. *arXiv preprint arXiv:2404.18231* (2024).

[9] Benjamin R Cowan, Leigh Clark, Heloisa Candello, and Janice Tsai. 2023. Introduction to this special issue: guiding the conversation: new theory and design perspectives for conversational user interfaces. *Human–Computer Interaction* 38, 3-4 (2023), 159–167.

[10] Samuel Rhys Cox and Wei Tsang Ooi. 2023. Conversational Interactions with NPCs in LLM-Driven Gaming: Guidelines from a Content Analysis of Player Feedback. In *International Workshop on Chatbot Research and Design*. Springer, 167–184.

[11] Barbara Downe-Wamboldt. 1992. Content analysis: method, applications, and issues. *Health care for women international* 13, 3 (1992), 313–321.

[12] Kafka J. X. Hlavacs H. Beutl L. Kryspin-Exner I. Kothgassner O. D. Felnhofer, A. 2019. Meeting others virtually in a day-to-day setting: Investigating social avoidance and prosocial behavior towards avatars and agents. *Computers in Human Behavior* 80 (2019), 399–406.

[13] I. Gris, Diego A. Rivera, Alex Rayón, Adriana I. Camacho, and D. Novick. 2016. Young Merlin: an embodied conversational agent in virtual reality. *Proceedings of the 18th ACM International Conference on Multimodal Interaction* (2016). https://doi.org/10.1145/2993148.2998534

[14] Sihao Hu, Tiansheng Huang, Fatih Ilhan, Selim Tekin, Gaowen Liu, Ramana Kompella, and Ling Liu. 2024. A survey on large language model-based game agents. *arXiv preprint arXiv:2404.02039* (2024).

[15] Kangsoo Kim, Celso M de Melo, Nahal Norouzi, Gerd Bruder, and Gregory F Welch. 2020. Reducing task load with an embodied intelligent virtual assistant for improved performance in collaborative decision making. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 529–538.

[16] François Mairesse, M. Walker, M. Mehl, and Roger K. Moore. 2007. Using Linguistic Cues for the Automatic Recognition of Personality in Conversation and Text. *J. Artif. Intell. Res.* 30 (2007), 457–500. https://doi.org/10.1613/jair.2349

[17] Masahiro Mori, Karl F. MacDorman, and Norri Kageki. 2012. The Uncanny Valley [From the field]. *IEEE Robotics Automation Magazine* 19, 2 (2012), 98–100. https://doi.org/10.1109/mra.2012.2192811

[18] Clifford Nass and Youngme Moon. 2000. Machines and mindlessness: Social responses to computers. *Journal of Social Issues* 56, 1 (2000), 81–103.

[19] Polygonal Mind. [n. d.]. Polygonal Mind. https://www.polygonalmind.com/ Accessed: 2024-06-14.

[20] Beatrice Rammstedt and Oliver P John. 2007. Measuring personality in one minute or less: A 10-item short version of the Big Five Inventory in English and German. *Journal of research in Personality* 41, 1 (2007), 203–212.

[21] Pejman Sajjadi, Laura Hoffmann, P. Cimiano, and S. Kopp. 2018. On the Effect of a Personality-Driven ECA on Perceived Social Presence and Game Experience in VR. *2018 10th International Conference on Virtual Worlds and Games for Serious Applications (VS-Games)* (2018), 1–8. https://doi.org/10.1109/VS-Games.2018.8493436

[22] Alon Shoa, Ramon Oliva, Mel Slater, and Doron Friedman. 2023. Sushi with Einstein: Enhancing Hybrid Live Events with LLM-Based Virtual Humans. In *Proceedings of the 23rd ACM International Conference on Intelligent Virtual Agents*. 1–6.

[23] Khvan A. Choi W. So, C. 2023. Natural conversations with a virtual being: How user experience with a current conversational AI model compares to expectations. *Computer Animation and Virtual Worlds* 34, 3 (2023), e2149.

[24] Ian Steenstra, Prasanth Murali, Rebecca B Perkins, Natalie Joseph, Michael K Paasche-Orlow, and Timothy Bickmore. 2024. Engaging and Entertaining Adolescents in Health Education Using LLM-Generated Fantasy Narrative Games and Virtual Agents. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–8.

[25] Hongyu Wan, Jinda Zhang, Abdulaziz Arif Suria, Bingsheng Yao, Dakuo Wang, Yvonne Coady, and Mirjana Prpa. 2024. Building LLM-based AI Agents in Social Virtual Reality. In *Extended Abstracts of the CHI Conference on Human Factors in Computing Systems*. 1–7.

[26] Brett Welch, M. van Mersbergen, and Leah B. Helou. 2021. Listeners' Perceptions of Speaker Personality Traits Based on Speech. *Journal of speech, language, and hearing research : JSLHR* (2021), 1–10. https://doi.org/10.1044/2021_JSLHR-20-00582

[27] Tamika Yamamoto. 2024. *Carmen Won't Talk to Me: Facilitating Expressive Qualities in Games Through Natural Language Interfaces and AI NPCs*. Ph. D. Dissertation. OCAD University.

[28] Michael Yin, Emi Wang, Chuoxi Ng, and Robert Xiao. 2024. Lies, Deceit, and Hallucinations: Player Perception and Expectations Regarding Trust and Deception in Games. In *Proceedings of the CHI Conference on Human Factors in Computing Systems*. 1–15.

[29] Nima Zargham, Maximilian A Friehs, Leandro Tonini, Dmitry Alexandrovsky, Emma Grace Ruthven, Lennart E Nacke, and Rainer Malaka. 2024. Let's Talk Games: An Expert Exploration of Speech Interaction with NPCs. *International Journal of Human–Computer Interaction* (2024), 1–21.

Received 20 February 2007; revised 12 March 2009; accepted 5 June 2009