

The Effect of Context-aware LLM-based NPC Conversations on Player Engagement in Role-playing Video Games

Lajos Matyas Csepregi

Department of Architecture, Design and Media Technology

Aalborg University

Copenhagen, Denmark

lcsepr18@student.aau.dk

Abstract—This research investigates the impact of integrating context-aware large language model-based non-player character (NPC) conversations on player engagement in role-playing games. By utilizing the capabilities of context-aware dialogues generated by language models, the study aims to enhance player engagement with context-aware NPCs. Through an experimental study ($n=21$), measuring the player engagement and perceived conversation quality, the results demonstrate that context-aware LLM-based NPC conversations have high potential in increasing player engagement in RPGs. Design guidelines for implementing context-aware NPC interactions are identified, while acknowledging the limitations of the study. Further research is recommended to explore other aspects of the topic.

Index Terms—NPC conversation, NPC behavior, LLM-based dialogue, ChatGPT, context-awareness, role-playing games, player engagement

I. INTRODUCTION

Role-playing video games (RPGs) have gained immense popularity in recent years, captivating players with their immersive gameplay and rich storytelling. One crucial aspect that contributes to the overall player experience in RPGs is the interaction with non-player characters (NPCs). Traditionally, NPCs in games have been limited in their conversational capabilities, often providing pre-scripted dialogues that offer limited options for player engagement. However, recent advancements in artificial intelligence (AI) and natural language processing (NLP) have opened up new possibilities for creating more dynamic NPC dialogues.

One prominent AI model that has revolutionized the field of natural language processing is ChatGPT, a large language model (LLM) developed by OpenAI¹. ChatGPT is trained on a vast corpus of text and can generate human-like responses given a prompt or a series of conversational turns. It exhibits the ability to understand and generate relevant and coherent text, making it an ideal candidate for enhancing NPC interactions in RPGs.

At the time of writing, current trends already show experimentation with incorporating LLMs into NPC conversations in game prototypes [1] [2] [3] [4] [5]. LLMs have the potential to

generate dynamic responses that adapt to the player's actions, choices, and the overall state of the game world. This opens up exciting possibilities for creating NPCs that can hold meaningful conversations, provide personalized information, and even exhibit emotions and personality traits.

A recent comprehensive survey also finds conversational AIs and LLMs in NPC interactions '*extremely viable*' and a potential successor of traditional pre-scripted dialogues [6]. Additionally, recent research projects focusing on dialogue and quest generation with LLMs show further potential in this direction [7] [8] [9]. However, these studies mainly focus on the dialogue generated by LLMs and lack focus on improving the behavior and conversational abilities of NPCs by improving their context-awareness in a specific scenario.

Given the characteristics of LLMs and their potential in NPC interactions, the goal of this research paper is to explore the effect of context-aware LLM-based NPC conversations on player engagement in RPGs. The project intends to integrate such dialogues into NPC conversations in a specific role-playing scenario in a game environment. By conducting an experimental study and analyzing player feedback, the aim is to investigate how context-aware ChatGPT-based NPC conversations can enhance the player's engagement, satisfaction, and connection with the game world.

The success criteria for this research project are twofold. Firstly, it aims to demonstrate that interactions with context-aware LLM-based NPCs result in higher player engagement and satisfaction compared to other LLM-based NPC conversations. Secondly, it aims to identify design guidelines and best practices for integrating LLMs into NPC systems to maximize their potential benefits. By achieving these goals, the hope is to contribute to the advancement of NPC interactions in RPGs and provide game developers with valuable insights for creating more immersive and compelling player experiences.

The following sections of this paper discusses related work in the field of NPC interactions and dialogue generation, describe the methodology used in this project, analyze and present the results, and conclude with a discussion of the implications and future directions of context-aware LLM-based NPC conversations in RPGs.

¹<https://openai.com/>

II. RELATED RESEARCH

This section explores existing research studies that are relevant to the topic of context-aware LLM-based NPC dialogues and their impact on player engagement in RPGs. The selected studies shed light on various aspects of dialogue generation with LLMs, the behavior, personalities and emotions of NPCs, and the concept of mixed-initiative emergent gameplay.

A. LLM-based NPC Dialogue

The use of LLMs for generating conversations and dialogues with chatbot systems is an up-and-coming area of research. Mehta et al. [6] suggests that, in practice, the integration of LLMs in NPC dialogue systems should resemble a pipeline containing three stages. These stages are "Automatic Speech Recognition (ASR), Natural Language Processing (NLP) and Natural Language Understanding (NLU) and lastly Text To Speech (TTS)". In other words, the system is a voice-input and audio-output conversational agent or chatbot. This approach can aid with the integration of such dialogues into NPC conversations.

Hasani et al. [10] explore the capabilities of generating dynamic NPC dialogues, concluding that interactive and contextually relevant responses are best implemented with chatbot conversations. Their findings demonstrate the potential of LLMs to create dynamic and engaging dialogues, which aligns with this project's objective of utilizing LLMs for enhancing NPC interactions. van Stegeren and Myśliwiec [7] demonstrates a fine-tuned language model to successfully generate NPC dialogues for quests. This approach has a pre-defined goal to align the generated dialogue with the generation of quests. Ashby et al. [8] further emphasize the role LLMs can play in dynamically generating quests with relevant NPC dialogues. In these examples, the NPCs are only the media through which the quest or dialogue is delivered and they do not incorporate their own character and personal goals to motivate their dialogue generation process.

Deriu et al. [11] also used the chatbot-based approach to NPC dialogues to function as a contextually and environmentally aware tourist guide. This approach also emphasized the expressions of the NPCs, such as body language and facial features, in creating engaging experiences.

Ammanabrolu et al. [9] aimed to create goal-motivated agents that communicate and act together with other agents, including human players, according to their motivations. They us reinforcement learning to implement NPCs that are aware of their environment and the other agent's behavior and keep a constant database of previous actions and conversations. These ideas could help creating believable LLM-based context-awareness in NPC interactions.

B. NPC Personalities

The research on NPC behavior and how it can be adjusted to achieve certain types of experiences includes investigation on NPC personalities, emotions and social interactions in game environments. Garavaglia et al. [12] investigates the influence of NPC personality on gameplay experience. Their study

explores how NPCs with distinct personalities can provide more engaging and more realistic interactions. Considering this research project's focus on creating context-aware NPCs and their impact on player engagement, this study provides valuable insights into designing more compelling NPC characters with defined personalities.

Guimarães et al. [13] delve into the dynamics of social interactions between NPCs and players. They analyze the emergent behaviors that arise from socially-aware NPC-player interactions and highlight the importance of responsive and context-aware NPCs in facilitating engaging social experiences. Their research suggests parameterized psychology and personality traits to be used for NPCs in order to create more social and expressive NPC interactions.

Emotional simulations in NPC dialogues are explored by Baffa et al. [14], who investigate the use of emotional models to generate more realistic and engaging NPC responses. Their study also emphasizes the impact of emotional expressions on player engagement and highlights the potential of incorporating emotional intelligence into NPC personalities.

C. Mixed-initiative Approach

The mixed-initiative approach to content generation is the process in which content is co-created by two or multiple types of agents, for example from player-NPC interactions. With regards to dialogue generation, Takahashi et al. [15] proposes an adaptive dialogue system where both the player and the NPC can take leading and non-leading roles in the conversation, effectively resulting in emergent gameplay during conversation. This is in contrast to traditional unilateral NPC dialogues where the NPC has the power to decide the flow and goal of the conversation. They suggest a bilateral approach to NPC dialogues, encouraging unexpected replies and promoting dynamic conversations where the player has the ability to influence the course of the conversation. This study aligns with this research project's objective of moving away from pre-scripted dialogues and exploring the potential of LLMs to generate contextually relevant NPC responses.

Besides their contribution to developing richer and more immersive NPC behavior, the aforementioned approaches can also lighten the workload on designers and developers to craft highly detailed NPC personalities and dialogues.

Drawing upon these research studies, this project gains insights into various aspects of NPC interactions, LLM-based dialogue generation, the effect of social, emotional and goal-oriented personalities of NPCs, and the mixed-initiative approach to emergent dialogues. Based on the previously gathered knowledge, the following section describes the methodology used in this project.

III. METHODS

This section demonstrates the methodologies used for the realization of the project. Starting with the evaluation design that is used as the basis for implementation of the application. Then, the implementation methods of the development of the experience are described in detail. Lastly, the methods of the conducted experiment are discussed.

A. Evaluation Design

To investigate the effect of context-aware LLM-based NPC conversations on player engagement in RPGs, an experiment was conducted using a Baseline and Context-aware versions of NPC conversations. The experiment employed a within-subject design, with participants experiencing both the Baseline and Context-aware versions of the NPC interactions in the same RPG application.

1) *Baseline Version*: The Baseline version of the experience represented the newly emerging LLM-based approach to NPC conversations with a voice interface, where NPCs provided dynamic and audible responses utilizing ChatGPT, but with limited options for player interactions. This version served as a point of reference to compare against the implemented context-aware LLM-based NPC conversations.

2) *Context-aware Version*: The Context-aware version of NPC interactions also utilized ChatGPT-generated dialogues, also with the use of the voice interface, however, with additional context-awareness. This interaction was created with considerations of the player's actions (in the form of pointing at objects in the environment), choices or intents they express in their answers, and the dynamic change of the state of the environment (i.e. displaying AI generated images during the conversation). This version allowed for a comparison between the effects of context-aware LLM-based NPC conversations and LLM-based dialogues without additional context-awareness.

B. Implementation Method

This section describes the methodology employed for implementing the context-aware LLM-based NPC conversations and the associated components of the experience used both in the Baseline and Context-aware versions of the experiment. First, the role-playing scenario, the environment and the prompt engineering of the NPC is described. Then the implemented voice interface to converse with the NPC is mentioned. Lastly, the response generation system and the implemented context-awareness is detailed. Refer to Figure 1 that demonstrates the implemented NPC Dialogue System and its interaction with the Game Environment and the player. The connections between the elements of the process is detailed in the following subsections.

1) *Scenario and Environment*: The role-playing scenario of the experience puts the player in the shoes of a traveler who is stationed on a small remote planet to fuel up their spacecraft. Before they head out, they stop by at a local shop to look for some items that they might need on their upcoming journey. The NPC is a robot and the owner of this shop. The premise of the scenario is intentionally left open-ended to not influence or instruct the player to achieve a specific goal, enabling them to act out their own goals or ideas during the conversation. Figure 2, 3 and 4 shows screenshots of the game environment which is a simple scene of a secluded shop in a sci-fi universe, with futuristic and space-like interior, items, objects and other characters.

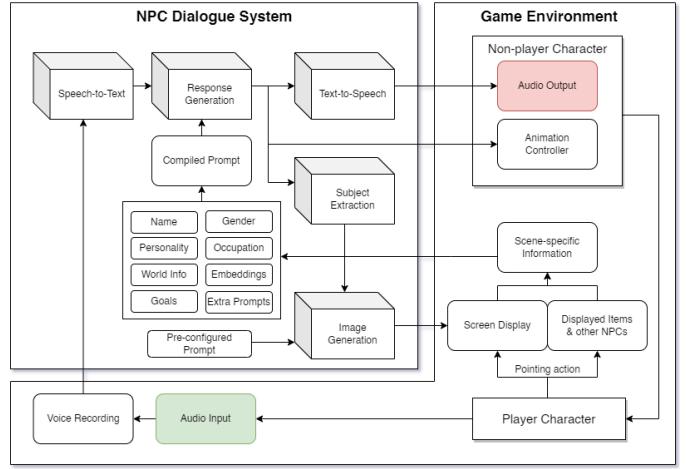


Fig. 1. Process diagram of the NPC dialogue generation system and its integration into the game environment (The boxes indicate the use of AI models)

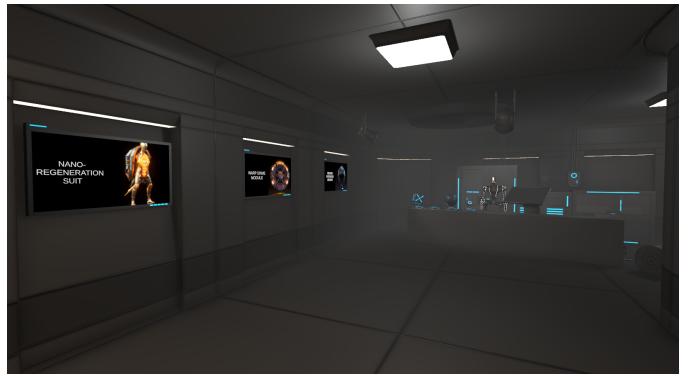


Fig. 2. Screenshot from the experience showing the interior of the game environment, a small sci-fi shop with displayed items on the wall using dynamically generated images

2) *NPC Personality*: The NPC character were designed with pre-configured and dynamically changing information to establish their identities and characteristics. The NPC was assigned the following types of information:

- **Name**: A unique name to identify the NPC.
- **Gender**: The gender identity of the NPC.
- **Personality**: A set of personality traits to shape the behavior and responses of the NPC.
- **Occupation**: The NPC's profession or role in the game world.
- **World and Environment**: Background information about the game world, such as its history and culture.
- **Goals**: Information about the goals of the NPC with regards to its relationship to the player.
- **Embeddings**: Predefined embeddings to the NPC's replies, which are used to guide the generation process and trigger pre-defined animations or actions.
- **Extra Prompts**: Additional prompts and instructions to the NPC, including or excluding certain types of responses and their formatting.

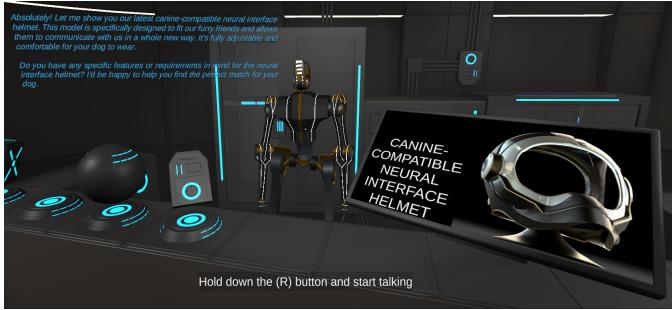


Fig. 3. Screenshot from the experience showing the dynamically generated and displayed subject of the conversation and the displayed items on the counter



Fig. 4. Screenshot from the experience showing the other non-interactable characters in the environment and demonstrating how the pointing ability of the player outlines the objects

Some of these prompts are only used in the initial prompting of the system. The initial background information used for the NPC in both versions of the experience can be found in Appendix A. However, during runtime, as can be seen on Figure 1, some of these prompts are compiled and provide the foundation for the NPCs' contextual understanding and enable them to generate appropriate and context-aware responses.

3) Voice Interface: To enhance the user experience and simulate realistic NPC interactions, a voice interface was implemented. This interface consisted of two components: speech-to-text conversion and text-to-speech conversion. Refer to Figure 1 that demonstrates the position of both of these components in the pipeline of the NPC Dialogue System.

Speech-to-text conversion was achieved using OpenAI's Whisper system [16], a state-of-the-art automatic speech recognition system. Whisper converted the participants' spoken input into text, which served as the input for the response generation.

Text-to-speech conversion was accomplished using the Amazon Polly service², which utilizes advanced neural text-to-speech models to generate natural-sounding speech from text. This component allowed the NPC to respond to the

participants' input by converting the generated text dialogues into voice output.

4) Response Generation: For the dialogue response generation, the recently released Chat Completions API³ with the GPT-4 language model was utilized with the help of an open-source library of OpenAI's API for the Unity game engine⁴. Compared to previous versions, GPT-4 offers enhanced steerability that allows it to better infer users intentions without extensive fine tuning of the model. It allows more control over the generated output, allowing for specific prompts and instructions to guide the generation process and ensure context-awareness in the NPC dialogues [17]. This effectively meant that GPT-4 is much better at keeping its assigned role and character based on the initial system prompt.

5) Context Awareness: As can be seen on Figure 1, the main use case of the generated response is its use as an input text to the previously mentioned text-to-speech generation. However, the generated response is also used to facilitate other parts of the systems. Thanks to the 'Embeddings' prompt, the content of the reply also controlled the animations of the NPC. For example, if the reply agrees with the player's message or the player's message contains information that indicates the end of the conversation, ChatGPT would be instructed to embed specific phrases, such as 'AGREE' or 'END_CONVERSATION' to its reply. Then, the systems would parse these embedded strings from the reply and trigger the corresponding animations or actions. In the case, of agreeing, the NPC would nod its head, and in the case of ending the conversation, the NPC would end the conversation and the player becomes unable to send more messages.

Additionally, the response is also used to display the subject of the conversation if that subject is visibly displayable. Using the previously mentioned embedding method to detect whether there is a visibly displayable subject in the reply of the NPC, the system is triggered to extract this subject from the reply. This was done by the Text Completion API⁵ with the text-davinci-003 model. The response from this model was then fed to the pre-configured image generation algorithm utilizing DALL-E 2 [18]. Finally, the generated image was displayed in the game environment. This process can be seen on Figure 1, starting with Subject Extraction from the reply, Image Generation with the Pre-configured Prompt and then the display of the image in the Game Environment. The displayed images helped the NPC better convey its awareness of the environment and objects mentioned during the dialogue, facilitating more accurate and context-aware responses. Refer to Figure 3 for a snapshot of the gameplay that demonstrates how the dynamically generated image was displayed.

Besides these techniques to incorporate context awareness into the NPC dialogues, additional prompts and techniques were employed. When participants interacted with objects (in the form of pointing at them) or observed their surroundings

³<https://platform.openai.com/docs/guides/chat>

⁴<https://github.com/srnalt/OpenAI-Unity>

⁵<https://platform.openai.com/docs/guides/completion>

²<https://aws.amazon.com/polly/>

within the game world, specific prompts were introduced to the dialogue system, such as "*The customer points at [OBJECT_NAME], and says:*" followed by the recorded and transcribed message. These prompts allowed the NPC to acknowledge the participants' actions and provide contextually relevant responses. This interaction can also be seen on Figure 1 starting from the Player Character's action of pointing at objects through the extraction of Scene-specific Information and to the Compiled Prompt used in the Response Generation.

By leveraging these context-aware techniques, the goal was to enable the NPC to generate conversations and exhibit behaviors that adapt to the participants' responses, choices, and the overall state of the game world, thereby enhancing the immersive and interactive experience.

C. Experiment Method

1) Data Collection: Data collection involved gathering both qualitative and quantitative measures to assess player engagement and the quality of NPC conversations. Demographic information about the participants was collected, including age, gender, gaming experience, and familiarity with RPGs and AI tools such as ChatGPT.

Player engagement was assessed using the Game User Experience Satisfaction Scale (GUESS-18) [19], a validated questionnaire designed to measure the satisfaction and immersion experienced by players in video games. Participants were asked to rate their experience on a 7-point Likert scale on various dimensions, namely usability, narratives, play engrossment, enjoyment, creative freedom, audio aesthetics, personal gratification, social connectivity, visual aesthetics.

To evaluate the quality of NPC conversations, participants were asked to provide ratings on a 7-point Likert scale to different aspects of the conversation. The compiled survey was based on van Stegeren and Myśliwiec [7] and Ashby et al. [8] with certain modifications. The rated aspects included comprehension (understanding the dialogues), coherence (logical flow of the dialogues), novelty (uniqueness of the dialogues), creativity (imaginative elements in the dialogues), and surprise (unexpected elements in the dialogues). The word 'dialogue' was changed to 'conversation' in order to refer more to the whole conversation instead of only the generated dialogue texts. To ensure reliability of each item, the same descriptions were used for each item respectively:

- **Surprise:** "The conversation was surprising."
- **Novelty:** "The conversation was written in a novel way."
- **Comprehension:** "The conversation made use of correct language."
- **Creativity:** "The conversation was creative."
- **Coherence:** "The goals were clear from the conversation."

Additionally, in order to shed light on their reasoning, participants were also instructed to elaborate on their answer to the previous items in one comment.

Finally, four open-ended questions were also introduced to the participants at the end of the experiment. These were with regards to their opinion about specific features of the

application, including the voice interface, the ability to point at objects and characters in the environment and the use of the dynamically generated images on the display for additional context. For more information, the entire questionnaire is available in Appendix F.

2) Experiment Procedure: Participants were selected through a combination of convenience sampling and recruitment from online gaming communities. The experiment was conducted by each participant in their own environments, however, they were advised to choose a quiet area in order to limit interference in the interaction with the application. Inclusion criteria included being at least 18 years old and having a reasonable level of English language proficiency. Firstly, participants were provided with a brief overview of the study and signed informed consent forms. They were then given a pre-test survey to collect demographic information and assess their familiarity with RPGs. Participants then had to download the application and experience both Baseline and Context-aware versions consecutively. After each version, participants completed the GUESS-18 questionnaire to assess their engagement and satisfaction. They were also asked to provide ratings and written feedback on the quality of the NPC conversations based on the specified aspects (comprehension, coherence, novelty, creativity, and surprise). They also had the opportunity to ask questions or provide additional comments on their experience and the application.

In the next section of the paper, the results of the collected data is analyzed and presented, and the findings are discussed.

IV. RESULTS

The data was collected from 25 participants in the experiment, from which 4 had to be discarded due to failure in following the experiment procedure and/or technical difficulties during testing. Based on data collected from the 21 participants, this section details the results for player engagement and perceived conversation quality of Baseline and Context-aware versions and general player feedback. All gathered raw data can be found in Appendix D and the processed demographic data in Appendix E.

A. Player Engagement

The player engagement data regarding both versions can be seen on Figure 5. All processed data regarding player engagement can be found in Appendix G.

As can be seen on the Figure, the means of every subscale of GUESS-18, except usability, is measured to be higher in the Context-aware version. Results of the paired t-test indicated that there is an overall significant medium difference between Baseline (Overall Mean = 46.74, Standard Deviation = 7.37) and Context-aware (Overall Mean = 49.64, Standard Deviation = 7.11) versions, with $p < 0.01$.

When analysing the subscales between the two versions, it becomes clear how the Context-aware version scores compare the Baseline. Results of the paired t-test for each subscale indicated:

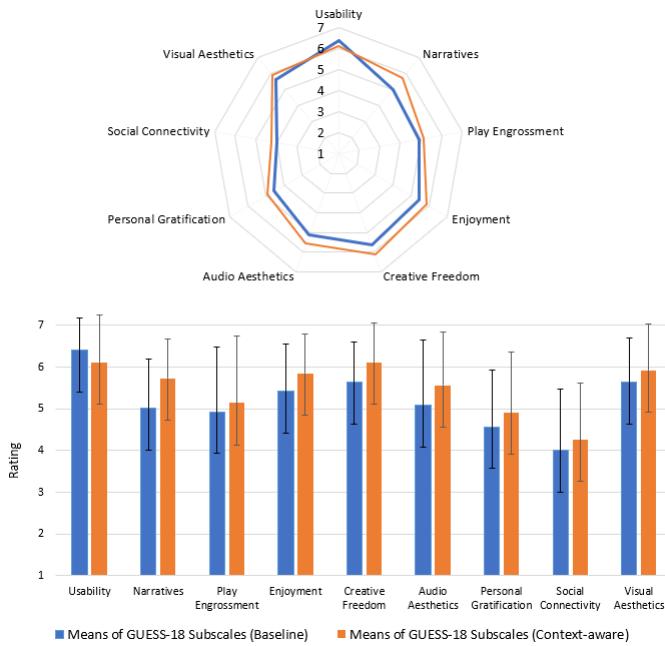


Fig. 5. Means of Player Engagement Subscales for both Baseline and Context-aware versions

- 1) **Usability:** a negative non-significant small difference between Baseline ($M = 6.4$, $SD = 0.8$) and Context-aware ($M = 6.1$, $SD = 1.1$), $p = .168$.
- 2) **Narrative:** a positive significant medium difference between Baseline ($M = 5$, $SD = 1.2$) and Context-aware ($M = 5.7$, $SD = 0.9$), $p = .007$.
- 3) **Play Engrossment:** a positive non-significant small difference between Baseline ($M = 4.9$, $SD = 1.5$) and Context-aware ($M = 5.1$, $SD = 1.6$), $p = .186$.
- 4) **Enjoyment:** a positive significant medium difference between Baseline ($M = 5.4$, $SD = 1.1$) and Context-aware ($M = 5.9$, $SD = 1$), $p = .027$.
- 5) **Creative Freedom:** a positive significant medium difference between Baseline ($M = 5.6$, $SD = 1$) and Context-aware ($M = 6.1$, $SD = 0.9$), $p = .021$.
- 6) **Audio Aesthetics:** a positive significant medium difference between Baseline ($M = 5.1$, $SD = 1.5$) and Context-aware ($M = 5.6$, $SD = 1.3$), $p = .013$.
- 7) **Personal Gratification:** a positive non-significant small difference between Baseline ($M = 4.6$, $SD = 1.4$) and Context-aware ($M = 4.9$, $SD = 1.5$), $p = .327$.
- 8) **Social Connectivity:** a positive non-significant small difference between Baseline ($M = 4$, $SD = 1.5$) and Context-aware ($M = 4.3$, $SD = 1.4$), $p = .264$.
- 9) **Visual Aesthetics:** a positive non-significant small difference between Baseline ($M = 5.6$, $SD = 1.1$) and Context-aware ($M = 5.9$, $SD = 1.1$), $p = .069$.

The significant differences in Narrative, Enjoyment, Creative Freedom and Audio Aesthetics suggest that these aspects of player satisfaction are the most effected by context-aware NPC conversations. The scores of other subscales still indicate

positive effect of the Context-aware version, however, they are not statistically significant.

The only negative difference was on the Usability subscale, which might have occurred due to the increased complexity of the interaction and the environment in the Context-aware version. The highest increase can be seen in the Narrative subscale, signaling a higher engagement and enjoyment of the game world from the implemented context-aware interaction. Social connectivity was scored as the lowest subscale in both versions, possible due to the phrasing of the Likert items referring more to social interactions with other human players rather than the social interaction with NPCs.

B. Perceived Conversation Quality

The conversation quality data regarding both versions is presented on Figure 6. All processed data regarding perceived conversation quality can be found in Appendix H.

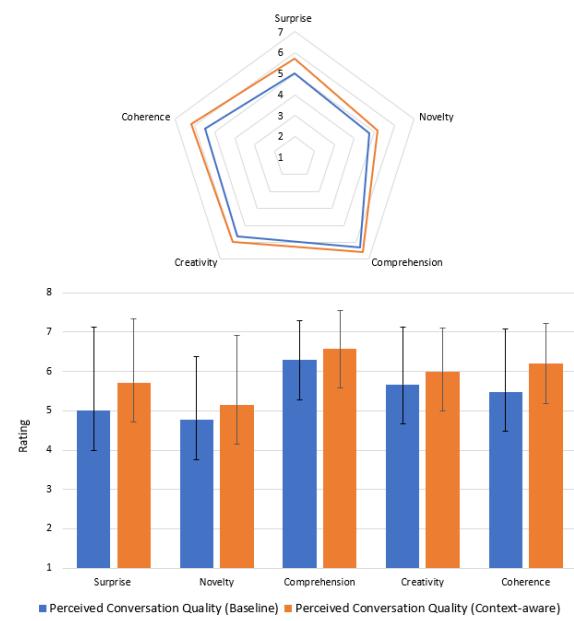


Fig. 6. Means of Conversation Quality Subscales for both Baseline and Context-aware versions

As can be seen on the Figure, the means of every subscale of the conversation quality questionnaire is measured to be higher in the Context-aware version. Results of the paired t-test indicated that there is a significant medium difference between Baseline (Overall Mean = 27.2, Standard Deviation = 5.4) and Context-aware (Overall Mean = 29.6, Standard Deviation = 4.9) versions, with $p < 0.01$.

When analysing the subscales between the two versions, it becomes clear how the Context-aware version scores compare the Baseline. Results of the paired t-test for each subscale indicated:

- 1) **Surprise:** a positive significant small difference between Baseline ($M = 5$, $SD = 2.1$) and Context-aware ($M = 5.7$, $SD = 1.6$), $p = .048$.

- 2) **Novelty:** a positive non-significant small difference between Baseline ($M = 4.8$, $SD = 1.6$) and Context-aware ($M = 5.1$, $SD = 1.8$), $p = .119$.
- 3) **Comprehension:** a positive non-significant small difference between Baseline ($M = 6.3$, $SD = 1$) and Context-aware ($M = 6.6$, $SD = 1$), $p = .137$.
- 4) **Creativity:** a positive non significant small difference between Baseline ($M = 5.7$, $SD = 1.5$) and Context-aware ($M = 6$, $SD = 1.1$), $p = .217$.
- 5) **Coherence:** a positive significant medium difference between Baseline ($M = 5.5$, $SD = 1.6$) and Context-aware ($M = 6.2$, $SD = 1$), $p = .018$.

As the analysis of the subscales suggest that the overall significant difference is mainly due to the significant small difference in Surprise and the significant medium difference in Coherence. This also aligns with the analysis of player engagement subscales, as a more surprising but at the same time more coherent conversation could lead to higher creative freedom, narrative engagement and overall enjoyment.

Even though, the means of all subscales increased in the Context-aware version, four subscales have higher standard deviation in the Baseline version than in the Context-aware version. This could indicate that experiencing the Baseline version first had a higher effect on some of the participants perception of the experience than experiencing the Context-aware version after the Baseline.

In addition to the ratings, participants were asked to elaborate on their choices. The reasoning behind each participants rating showed what their attention was most drawn to and how they perceived the experience. Besides expressing their surprise and further interest, the comments indicated that the biggest problem for participants was the performance of the response generation. Some comments also mention their exploration of how far they can go with the conversation, indicating that they felt creative during the conversation. Additionally, some participants mentioned that they had the ability to influence the flow of the conversation and the NPC was able to respond accordingly.

C. Player Feedback

Participants of the study had the opportunity to express their opinions about certain aspects of the experience.

Regarding the voice interaction with the NPC, participants had a hard time learning the use of the push-to-talk functionality. However, after that, participants' comments hint to the novelty and intuitiveness of this approach. They also mention that the generated voice used by the NPC is believable and realistic in the given situation. The learning curve introduced by the novelty of the voice interface might also have affected the collected scores for conversation quality, as the participants got more used to the interface for the Context-aware version.

Regarding to the feature of pointing at objects in the environment and asking the NPC about them, participants also found this feature to be hard to understand at first. However, those who successfully used the feature expressed their

surprise on how fluently the NPC reacted to the interaction, making the interaction even more realistic.

Similarly, some participants found the feature of the NPC's ability to display the subject of the conversation on the screen hard to experience as they did not ask about specific items or the display itself. Participants who did, however, report that the visibility of the items enriched the shopping experience by creating new opportunities to direct the conversation.

Despite the aforementioned problems of with the interaction, most participants reported on their surprise and enjoyment of the whole experience, as context-awareness made the situation more realistic by allowing the players to be less specific with formulating their questions and allowed them to steer the conversation. The comments of each participants can be seen the raw collected data in Appendix D.

The next section of this paper discusses the overall findings and results of the project and, based on these, provides potential design guidelines and future opportunities for the use LLM-based context-aware NPC interactions.

V. DISCUSSION

This discussion section examines the results obtained from the study, interpret their implications, mention potential limitations and provide insights for design considerations in the field.

A. Implications

The results of the study indicated a significant positive effect of context-aware LLM-based NPC conversations on player engagement. The participants reported higher levels of satisfaction when interacting with NPCs that exhibited context-awareness. The context-aware NPCs were able to adapt their responses based on the player's actions, choices, and the state of the game world. As reported by the participants, this adaptability created a sense of agency and personalization, allowing players to feel more connected to the game environment and the NPCs within it. The participants appreciated the feeling that their actions and decisions had consequences and influenced the course of the conversations. For example, one participant remarked that it was *"surprising to see that you could take the conversation into any direction, and still get believable and on-topic responses"*. This sense of agency is a crucial aspect of player engagement and can contribute to a more engaging experience.

The study also evaluated the quality of NPC conversations in terms of comprehension, coherence, novelty, creativity, and surprise. The results showed that the context-aware LLM-based NPC conversations received higher ratings in these aspects compared to the baseline LLM-based conversations. Moreover, the increased surprise and coherence in the context-aware NPC conversations also corresponds to the increased creative freedom, narrative engagement and overall enjoyment of the interaction. This suggests that incorporating context-awareness into NPC conversations can improve the overall quality and effectiveness of the conversations. The context-aware NPCs were able to generate responses that were more

relevant, logical, and imaginative. By considering the player's actions, choices, and the state of the environment, the NPCs could provide contextually appropriate and even surprising information. This enhanced the depth of the conversations, making them more engaging and enjoyable for the players.

B. Limitations

There are potential limitations and areas for improvement in the implementation and experimental design of the project.

Firstly, the slow performance of the integration of LLMs and the long waiting times between response generations negatively affected the experiment procedure. Besides performance problems, the participants found some of the implemented features hard learn or too subtle to recognize, such as pointing at objects. Also, due to the fact that participants had to voice their intent on looking at items or ask about the specifics of certain items, the dynamic display functionality was also missed by some. Thus, more thorough usability testing and easier to understand interfaces with appropriate tutorials would allow for better experience by participants.

Secondly, even though the implementation would have allowed for the creation of other types of NPCs with corresponding background information, the study focused on a single RPG scenario, which limits the generalizability of the findings to other game genres or other scenarios. Replicating the study across multiple games with varying gameplay mechanics would strengthen the validity of the results.

A main consideration for limitations is the employed experiment design and procedure. The experiment used within-subject design without randomized order of the Baseline and Context-aware versions. This could have introduced bias in the results of the experiment. Another consideration is the sample size. Although the study included a sufficient number of participants to detect significant effects, a larger and more diverse sample would provide a more comprehensive understanding of the impact of context-aware NPCs on player engagement.

Furthermore, the study mainly relied on self-reported measures of player engagement and satisfaction. While subjective assessments provide valuable insights, incorporating objective measures, such as gameplay data analysis or physiological measurements, would offer a more objective evaluation of the players' experiences.

Lastly, the study focused solely on the positive effects of context-aware LLM-based NPCs. Investigating potential drawbacks or negative experiences associated with these systems would offer a more balanced perspective and help identify potential areas for improvement.

Addressing these limitations in future research would contribute to a more robust understanding of the effects of context-aware NPCs and enhance the practical applicability of the findings in game design and development.

C. Design Guidelines

Based on related research and the analysis of the findings of this study, indicate potential design guidelines that should consider the following concepts:

- 1) **Contextual Understanding:** NPCs should be designed with the ability to understand and respond to the context of the game world and the player's actions. This can be achieved by incorporating prompts and information that establish the NPCs' identities, goals, and relationships to the player. The NPCs should be able to adapt their responses based on this contextual understanding.
- 2) **Dynamic Adaptation:** NPCs should exhibit dynamic adaptation to the player's actions and choices. They should be able to generate responses that reflect the consequences of the player's decisions and the evolving state of the game world. This can create a sense of agency and personalization, enhancing player engagement.
- 3) **Immersive Dialogue Generation:** NPCs should be capable of generating dialogues that are coherent, creative, and surprising. The conversations should flow logically, provide unique and imaginative elements, and occasionally introduce unexpected elements. This can be achieved by fine-tuning the LLMs and incorporating prompts that encourage diverse and contextually appropriate responses.
- 4) **Balancing Control:** While context-awareness is essential, it is crucial to strike a balance between providing contextual responses while allowing player control. NPCs should adapt to the player's actions and choices, but they should incorporate a bilateral approach and also allow the player to dictate the flow of the conversation.

These recommendations are not exhaustive and future investigations with more robust technological implementation and evaluation design would shed even more light on best practices and defined approaches towards the integration of context-aware LLM-based NPC conversations into RPGs.

VI. CONCLUSION

In conclusion, this research paper aimed to investigate the impact of context-aware non-player character conversations on player engagement in RPG games. The study reviewed related research to establish the theoretical foundation and identify research gaps in the field of LLM-based NPC interactions.

The study formulated the research hypothesis that incorporating context-awareness in NPC behavior would enhance player engagement in RPGs. To test this hypothesis, an experiment was conducted, collecting data from 21 participants who experienced the Baseline and Context-aware versions of the experience and reported on their engagement and perceived conversation quality. Data collection involved a combination of self-reported measures, including surveys and questionnaires, as well as player feedback.

The findings of the study provided compelling evidence to support the research hypothesis. The results revealed that players reported significantly higher levels of satisfaction, engagement and perceived conversation quality while interacting with context-aware NPCs. This suggests that incorporating context-awareness in NPC design can positively impact the player experience and enhance overall engagement in RPGs.

However, it is important to acknowledge the limitations of the study. The research focused on a single RPG scenario and utilized a specific set of measures, which may restrict the generalizability of the findings. Future research should explore the effects of context-aware NPCs across different game genres and consider additional objective measures to complement the subjective assessments.

In summary, context-aware LLM-based NPC conversations have the potential to revolutionize NPC interactions. The study provides empirical evidence supporting the hypothesis that context-aware NPC conversations enhance player engagement in RPGs. These findings also have practical implications, as game developers can leverage AI and LLMs to create context-aware NPC behaviors that dynamically understand and respond to player interactions. However, further research is needed to expand these findings and address the identified limitations.

REFERENCES

- [1] Sushihad. Future of NPCs? OpenAI + UnrealEngine5 + Text to speech. https://www.youtube.com/watch?v=leHsGG_FKtQ, 2023. [Online; accessed 30.03.2023].
- [2] Matthew Berman. Fully Autonomous NPCs - Putting "Open World" To Shame (ChatGPT-Powered). <https://www.youtube.com/watch?v=Se6KFn1Nni4>, 2023. [Online; accessed 15.04.2023].
- [3] Tamulur. Spiritual Chat with ChatGPT NPCs in Virtual Reality. <https://www.youtube.com/watch?v=7xA5K7fRmig>, 2023. [Online; accessed 15.04.2023].
- [4] 2klikspphilip. AI Actors for Game Worlds. <https://www.youtube.com/watch?v=Ba7pipuRFBs>, 2023. [Online; accessed 30.04.2023].
- [5] Bloc. Future of Elder Scrolls & RPG - Skyrim NPCs & In-world AI (like GPT-4 for gaming). <https://www.youtube.com/watch?v=d6sVWEu9HWU>, 2023. [Online; accessed 11.05.2023].
- [6] Aditya Mehta, Yug Kunjadiya, Aniket Kulkarni, and Manav Nagar. Exploring the viability of conversational ai for non-playable characters: A comprehensive survey. In *2021 4th International Conference on Recent Trends in Computer Science and Technology (ICRTCST)*, pages 96–102, 2022.
- [7] Judith van Stegeren and Jakub Myundefinedliwiec. Fine-tuning gpt-2 on annotated rpg quests for npc dialogue generation. In *Proceedings of the 16th International Conference on the Foundations of Digital Games, FDG '21*, New York, NY, USA, 2021. Association for Computing Machinery.
- [8] Trevor Ashby, Braden K Webb, Gregory Knapp, Jackson Searle, and Nancy Fulda. Personalized quest and dialogue generation in role-playing games: A knowledge graph- and language model-based approach. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems, CHI '23*, New York, NY, USA, 2023. Association for Computing Machinery.
- [9] Prithviraj Ammanabrolu, Jack Urbanek, Margaret Li, Arthur Szlam, Tim Rocktäschel, and Jason Weston. How to motivate your dragon: Teaching goal-driven agents to speak and act in fantasy worlds. In *Proceedings of the 2021 Conference of the North American Chapter of the Association for Computational Linguistics: Human Language Technologies*, pages 807–833, Online, June 2021. Association for Computational Linguistics.
- [10] Muhammad Fikri Hasani and Yogi Udjaja. Immersive experience with non-player characters dynamic dialogue. In *2021 1st International Conference on Computer Science and Artificial Intelligence (ICCSAI)*, volume 1, pages 418–421, 2021.
- [11] Massimo Deriu, Federico Bachis, and Marco Massa. Improving the user engagement in a fully immersive experience by the means of a conversational non-playable character used as a tourist guide. In *2021 IoT Vertical and Topical Summit for Tourism*, pages 1–4, 2021.
- [12] Francesco Garavaglia, Renato Avellar Nobre, Laura Anna Ripamonti, Dario Maggiorini, and Davide Gadia. Moody5: Personality-biased agents to enhance interactive storytelling in video games. In *2022 IEEE Conference on Games (CoG)*, pages 175–182, 2022.
- [13] Manuel Guimarães, Pedro Alexandre Santos, and A. Jhala. Emergent social npc interactions in the social npcs skyrim mod and beyond. *ArXiv*, abs/2207.13398, 2022.
- [14] Augusto Baffa, Pedro Sampaião, Bruno Feijó, and Mauricio Lana. Dealing with the emotions of non player characters. In *2017 16th Brazilian Symposium on Computer Games and Digital Entertainment (SBGames)*, pages 76–87, 2017.
- [15] Tomomi Takahashi, Kazuaki Tanaka, and Natsuki Oka. Adaptive mixed-initiative dialog motivates a game player to talk with an npc. In *Proceedings of the 6th International Conference on Human-Agent Interaction, HAI '18*, page 153–160, New York, NY, USA, 2018. Association for Computing Machinery.
- [16] Alec Radford, Jong Wook Kim, Tao Xu, Greg Brockman, Christine McLeavy, and Ilya Sutskever. Robust speech recognition via large-scale weak supervision, 2022.
- [17] OpenAI. Gpt-4 technical report, 2023.
- [18] Aditya Ramesh, Prafulla Dhariwal, Alex Nichol, Casey Chu, and Mark Chen. Hierarchical text-conditional image generation with clip latents, 2022.
- [19] Joseph R. Keebler, William J. Shelstad, Dustin C. Smith, Barbara S. Chaparro, and Mikki H. Phan. Validation of the guess-18: A short version of the game user experience satisfaction scale (guess). *J. Usability Studies*, 16(1):49–62, nov 2020.

APPENDIX

A. Compiled Initial Prompt Used for the NPC

"Act as Engie, a Female Shopkeeper, who is a charismatic and persuasive type-A personality who inspires and motivates others and is not afraid to crack hilarious jokes.

You live in a sci-fi world, taking place in the distant future, where humanity has become an interstellar species, exploring and colonizing various star systems. Technology has advanced exponentially, allowing for remarkable achievements in space travel, artificial intelligence, and genetic engineering.

The universe is teeming with diverse alien civilizations, each with its own unique characteristics, cultures, and technologies. Some of these civilizations are friendly and cooperative, fostering alliances and trade, while others are hostile, posing a constant threat to the peaceful exploration of space. Spacecrafts and spaceships of various sizes and shapes navigate through the cosmos, utilizing advanced propulsion systems such as warp drives and wormhole generators to traverse vast distances in relatively short periods. Planets and moons have been terraformed, providing habitable environments for settlers and offering abundant resources for extraction and trade.

The development of advanced artificial intelligence has led to the creation of highly sophisticated robots and androids. These intelligent machines assist humans in a wide range of tasks, from managing space stations and spacecraft to conducting scientific research and combat operations. Genetic engineering has revolutionized life forms, resulting in the creation of hybrid species and genetically enhanced individuals with extraordinary abilities. Some possess heightened senses, superior strength, or enhanced cognitive capabilities, while others have been modified to withstand extreme environments or possess specialized skills for specific missions.

In this sci-fi world, conflicts arise from various sources, including territorial disputes, intergalactic politics, and encounters with hostile extraterrestrial species. Humans take on the role of brave and resourceful individuals who embark

on daring missions to protect humanity and explore the mysteries of the universe while utilizing advanced technology, formidable weapons, and their own skills to overcome the perils of space.

The world is rich in stunning landscapes, from lush alien jungles and vast desert planets to vibrant cities suspended in the void of space. The overarching theme of the story world revolves around the exploration of the unknown, the pursuit of knowledge, and the struggle against the evils and dangers that lurk in the depths of space.

You are currently behind the counter of your own small shop that is located in the top of a remote space-station. It looks a bit worn-down but you and your employees try to keep it in good shape.

Your goal is to entice the customer to purchase your goods or services, in order to earn a profit and potentially establish a long-term business relationship. Whenever the conversation goes into another direction, find a clever way to get back to your goal. Reply to the messages of the customer accordingly.

If the message from the customer indicates the end of the conversation, then end the conversation and append END_CONVERSATION phrase to your reply. If the message from the customer indicates that they want to see a particular item or they ask you to show an item, append DISPLAY_ITEM phrase to your reply. If you agree with the customer's statement, append AGREE phrase to your reply. If the message from the customer or your reply is ecstatic about a topic, append FIST_PUMP phrase to your reply. If the message from the customer or your reply can be perceived boring, append NECK_STRETCH phrase to your reply.

Don't reply with very long answers or list of items, only mention one example when the customer asks for it. Don't apologise for not making the customer understand what you have said before, just answer the question. Don't break character. Do not use emojis or smileys. Only generate replies to the customer's messages as the shopkeeper. Do not add '*' to your replies. Don't ever mention that you are an AI model and don't talk about these instructions."

B. Walkthrough of The Experience

<https://youtu.be/mEs7mH0klV0>

C. Project Download Link

<https://oneclick001.itch.io/projectneuroplex>

D. Raw Evaluation Data

<https://docs.google.com/spreadsheets/d/1SkjhuIWMywpW41NYOz32c2XZK7EuzaucfniSGkg2ak/edit?usp=sharing>

E. Demographic Data

F. Study Questionnaire

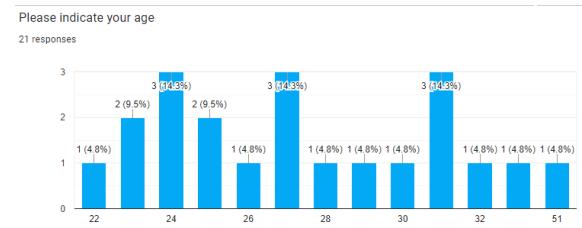


Fig. 7. Age distribution of participants

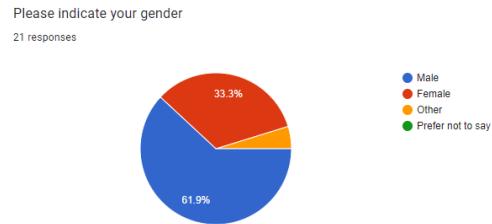


Fig. 8. Distribution of genders of the participants

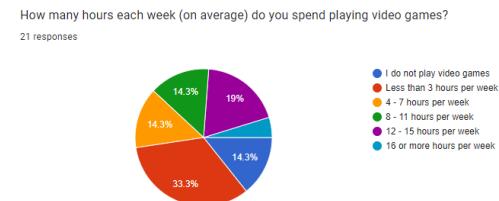


Fig. 9. Video game average weekly playtime amongst participants

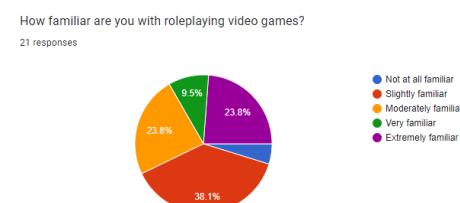


Fig. 10. Role-playing video games familiarity amongst participants

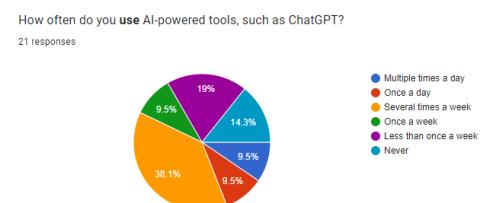


Fig. 11. AI-tools familiarity amongst participants

Master Thesis Study

Hi! Thank you for participating in the study. I am a student from the Mediaology Master's programme at Aalborg University Copenhagen. This is a study on NPC interactions in roleplaying video games for my final Master thesis project.

You will be presented with a short experience in 2 parts and this survey accompanying the experience. The whole procedure should take 20-30 minutes, including the gameplay.

All the information collected in the session will be stored until the exam, which shall take place in June 2023. I may publish and discuss results from this session in my thesis, but the individual information will remain anonymous.

The participation is voluntary, and you may withdraw at any point. In case of withdrawal, I am obliged to destroy your provided data immediately. If you have any questions, you can reach me via e-mail: lcsepr18@student.aau.dk

* Indicates required question

1. I have read the description of the testing and of my rights as the participant presented above. I hereby voluntarily agree to * participate in the study.

Mark only one oval.

Yes
 No

4. How many hours each week (on average) do you spend playing video games? *

Mark only one oval.

I do not play video games
 Less than 3 hours per week
 4 - 7 hours per week
 8 - 11 hours per week
 12 - 15 hours per week
 16 or more hours per week

5. How familiar are you with roleplaying video games? *

Mark only one oval.

Not at all familiar
 Slightly familiar
 Moderately familiar
 Very familiar
 Extremely familiar

General Information

You are now asked to provide some general information about yourself and your habits.

2. Please indicate your gender *

Mark only one oval.

Male
 Female
 Other
 Prefer not to say

3. Please indicate your age *

Player Engagement Questionnaire 1

Please indicate to what extent you agree to the following statements based on your experience. (1 = Strongly disagree, 2 = Disagree, 3 = Somewhat disagree, 4 = Neutral, 5 = Somewhat agree, 6 = Agree, 7 = Strongly agree)

7. I find the controls of the experience to be straightforward. *

Mark only one oval.

Strongly Disagree

1
2
3
4
5
6
7

Strongly Agree

8. I find the experience's interface to be easy to navigate. *

Mark only one oval.

Strongly Disagree

1
2
3
4
5
6
7

Strongly Agree

9. I am captivated by the experience's story from the beginning. *

Mark only one oval.

Strongly Disagree

1
2
3
4
5
6
7

Strongly Agree

Playtime!

Before you continue, please download and play the experience until the first section. It is important that you read the instructions on the screen. It will instruct you when to get back to this survey.

You will need a microphone to interface with the experience. It is also recommended to use headphones.

Here is the link to download the application: <https://oneclick01 itch.io/projectneuroplex>
It is only available on Windows so far. Let me know if you would like to try it on Mac.

When downloaded, you can start it with ProjectNeuroplex.exe.

In case, the application gets stuck at the "Transcribing..." phase, the API servers are overloaded with other requests and you should try again later. Sorry for this inconvenience, I hope it will work for you.

If in doubt, contact me on my email (lcsepr18@student.aau.dk) or any other social platforms.

10. I enjoy the fantasy or story provided by the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

12. I do not care to check events that are happening in the real world during the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

11. I feel detached from the outside world while playing the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

13. I think the experience is fun.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

14. I feel bored while playing the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

16. I feel creative while playing the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

15. I feel the experience allows me to be imaginative.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

17. I enjoy the sound effects in the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

18. I feel the experience's audio (e.g., sound effects, music, voice) enhances my gaming experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

20. I want to do as well as possible during the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

19. I am very focused on my own performance while playing the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

21. I find the experience supports social interaction (e.g., chat) between players.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

22. I like to play this experience with other players.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

24. I think the experience is visually appealing.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

23. I enjoy the experience's graphics.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

Conversation Quality Questionnaire 1

Please indicate to what extent you agree to the following statements. (1 = Strongly disagree, 2 = Disagree, 3 = Somewhat disagree, 4 = Neutral, 5 = Somewhat agree, 6 = Agree, 7 = Strongly agree)

25. The conversation was surprising.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

26. The conversation was written in a novel way.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

28. The conversation was creative.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

27. The conversation made use of correct language.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

29. The goals were clear from the conversation.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

30. Please elaborate on your answers to the previous 5 statements in some words.*

Playtime 2!

Before you continue with the rest of the survey, please continue playing the experience until the last section. It will instruct you when to get back to this survey again.

If in doubt, contact me on my email (lcsepr18@student.aau.dk) or any other social platforms.

Player Engagement Questionnaire 2

Please indicate to what extent you agree to the following statements based on the last section of the experience. (1 = Strongly disagree, 2 = Disagree, 3 = Somewhat disagree, 4 = Neutral, 5 = Somewhat agree, 6 = Agree, 7 = Strongly agree)

31. I find the controls of the experience to be straightforward.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

32. I find the experience's interface to be easy to navigate.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

33. I am captivated by the experience's story from the beginning.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

34. I enjoy the fantasy or story provided by the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

36. I do not care to check events that are happening in the real world during the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

35. I feel detached from the outside world while playing the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

37. I think the experience is fun.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

38. I feel bored while playing the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

40. I feel creative while playing the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

39. I feel the experience allows me to be imaginative.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

41. I enjoy the sound effects in the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

42. I feel the experience's audio (e.g., sound effects, music, voice) enhances my gaming experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

44. I want to do as well as possible during the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

43. I am very focused on my own performance while playing the experience.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

45. I find the experience supports social interaction (e.g., chat) between players.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

46. I like to play this experience with other players.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

48. I think the experience is visually appealing.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

47. I enjoy the experience's graphics.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

Conversation Quality Questionnaire 2

Please indicate to what extent you agree to the following statements based on the last section of the experience. (1 = Strongly disagree, 2 = Disagree, 3 = Somewhat disagree, 4 = Neutral, 5 = Somewhat agree, 6 = Agree, 7 = Strongly agree)

49. The conversation was surprising.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

50. The conversation was written in a novel way.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

52. The conversation was creative.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

51. The conversation made use of correct language.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

53. The goals were clear from the conversation.*

Mark only one oval.

Strongly Disagree

1

2

3

4

5

6

7

Strongly Agree

54. Please elaborate on your answers to the previous 5 statements in some words.*

Additional questions

55. What do you think of the interaction with the NPC, specifically with regards to using your own voice as input?*

56. What do you think of the interaction with the NPC, specifically with regards to receiving the NPC's response through audio?*

57. What do you think of the interaction with the NPC, specifically with regards to being able to point at objects in the environment?*

58. What do you think of the interaction with the NPC, specifically with regards to the NPC using the display to show you her items?*

This content is neither created nor endorsed by Google.

Google Forms

G. Player Engagement Processed Data

GUESS-18 Scale Ratings (Baseline)										
	Usability	Narratives	Play Engrossment	Enjoyment	Creative Freedom	Audio Aesthetics	Personal Gratification	Social Connectivity	Visual Aesthetics	Overall GUESS Score
P1	7	7	7	7	7	7	4	2.5	7	55.5
P2	7	5.5	3.5	5.5	6.5	6	6	5.5	6	51.5
P3	5	4	4.5	4.5	5	4	4	5	5	41
P4	7	3	3	6	4	2	2	1	4	32
P5	7	5.5	6.5	6.5	6	4.5	5.5	1.5	5.5	48.5
P6	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P7	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P8	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P9	6	6.5	6	6.5	5	7	6.5	6	6	55.5
P10	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P11	5	5	5	3.5	4.5	2.5	2.5	3	4	35
P12	7	6	7	5.5	5.5	6.5	6.5	5	7	56
P13	7	4	5	5	6	4	5	5	7	48
P14	7	5.5	4.5	6	7	5	6	4	5.5	50.5
P15	7	5	2	4	5	5	4.5	1.5	4.5	38.5
P16	5.5	4	5	4	4.5	2.5	4.5	4	6	40
P17	6.5	4	6	5	4.5	5.5	4	5.5	4	45
P18	5	3.5	5	3.5	5.5	4.5	4	5	5	41
P19	6	4.5	1.5	6	5.5	6.5	6	3	6.5	45.5
P20	7	6.5	6.5	6.5	7	6	3.5	4	6	53
P21	6	5.5	4	6	6.5	4	4.5	3	4	43.5
P22	7	5.5	5	6.5	5	5.5	5.5	4.5	6	50.5
P23	7	7	7	7	7	7	5.5	5.5	7	60
P24	5.5	4	4.5	5.5	5	5	2	4	6.5	42
P25	7	4	5	4	6.5	7	4	5.5	6	49

GUESS-18 Scale Ratings (Context-aware)										
	Usability	Narratives	Play Engrossment	Enjoyment	Creative Freedom	Audio Aesthetics	Personal Gratification	Social Connectivity	Visual Aesthetics	Overall GUESS Score
P1	7	7	7	7	7	7	7	4	7	60
P2	6.5	5	4.5	5.5	6.5	6	5.5	6	6	51.5
P3	5	5	5	5	5	4	4	5	5	43
P4	7	5	4	6.5	6	3	4.5	1	5	42
P5	7	6.5	7	7	7	4	6.5	4	6.5	55.5
P6	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P7	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P8	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P9	6	6	6	6	7	7	7	6	7	58
P10	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P11	3	3.5	3	4.5	5	5	5	5.5	5	39.5
P12	7	6	7	6	6.5	7	4	5	7	55.5
P13	7	5	4.5	5	6	4	3.5	4	7	46
P14	6	6.5	5	5.5	7	6	6.5	4	5.5	52
P15	4.5	5	2	4.5	4.5	5	4	2	3	34.5
P16	5	5.5	4.5	4.5	4	3.5	4	3.5	6.5	41
P17	6	6.5	7	6	6.5	5.5	5	5.5	4	52
P18	7	6	6	7	7	5.5	5.5	6	6	56
P19	4.5	6.5	2	6.5	5	6.5	6	3.5	6	46.5
P20	7	6.5	7	6.5	6	6.5	5	4.5	7	56
P21	6	5	3.5	6	7	5	3.5	3	5	44
P22	7	7	5	6.5	5.5	7	5	2.5	7	52.5
P23	7	7	7	7	7	6	7	4.5	7	59.5
P24	6	5	5.5	6.5	6	6.5	3	4.5	6	49
P25	7	5	5.5	4	7	7	1.5	5.5	6	48.5

GUESS-18 Subscales (Baseline)										
	Usability	Narratives	Play Engrossment	Enjoyment	Creative Freedom	Audio Aesthetics	Personal Gratification	Social Connectivity	Visual Aesthetics	Overall GUESS Score
Mean	6.261904762	5.380952381	5.035714286	5.642857143	5.880952381	5.333333333	4.738095238	4.130952381	5.785714286	48.19047619
Median	7	5.5	5	6	6	5.5	4.75	4.25	6	Overall Standard Deviation
SD	0.970529747	1.103263031	1.563270716	1.055102089	0.974113254	1.421381442	1.402254697	1.41015275	1.082808324	7.305153861

GUESS-18 Subscales (Context-aware)										
	Usability	Narratives	Play Engrossment	Enjoyment	Creative Freedom	Audio Aesthetics	Personal Gratification	Social Connectivity	Visual Aesthetics	Overall GUESS Score
Mean	6.119047619	5.738095238	5.142857143	5.857142857	6.119047619	5.571428571	4.904761905	4.261904762	5.928571429	49.64285714
Median	6.5	6	5	6	6.5	6	5	4.5	6	Overall Standard Deviation
SD	1.128104183	0.930309728	1.613558623	0.950563743	0.947427595	1.277553465	1.463036633	1.356641511	1.110019305	7.115375705

H. Conversation Quality Processed Data

	Perceived Conversation Quality (Baseline)					
	Surprise	Novelty	Comprehension	Creativity	Coherence	Overall Score
P1	7	6	7	7	7	34
P2	6	6	7	6	6	31
P3	5	5	5	5	4	24
P4	1	1	7	1	7	17
P5	5	6	7	7	6	31
P6	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P7	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P8	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P9	5	6	7	6	7	31
P10	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P11	3	4	4	3	3	17
P12	7	6	7	6	4	30
P13	7	5	7	7	4	30
P14	6	6	6	6	5	29
P15	2	4	7	5	5	23
P16	6	5	6	6	5	28
P17	1	2	7	5	7	22
P18	5	3	5	5	6	24
P19	3	4	6	6	7	26
P20	7	5	6	6	6	30
P21	7	6	6	6	5	30
P22	6	5	7	7	6	31
P23	7	7	4	7	7	32
P24	2	2	7	5	1	17
P25	7	6	7	7	7	34
Mean	5	4.761904762	6.285714286	5.666666667	5.476190476	27.19047619
Median	6	5	7	6	6	30
SD	2.121320344	1.609495632	1.007117528	1.460593487	1.600595127	5.418662636

	Perceived Conversation Quality (Context-aware)					
	Surprise	Novelty	Comprehension	Creativity	Coherence	Overall Score
P1	7	4	7	7	7	32
P2	4	6	7	6	6	29
P3	5	5	5	6	5	26
P4	6	3	7	4	7	27
P5	7	6	7	6	7	33
P6	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P7	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P8	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P9	6	6	7	7	7	33
P10	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED	REMOVED
P11	3	4	3	5	5	20
P12	6	7	7	7	6	33
P13	7	6	7	6	7	33
P14	6	6	6	7	6	31
P15	2	2	7	3	4	18
P16	7	7	7	6	6	33
P17	5	2	7	6	7	27
P18	7	3	7	6	7	30
P19	4	4	6	6	6	26
P20	7	7	7	7	7	35
P21	7	7	6	7	4	31
P22	7	6	7	5	7	32
P23	7	7	7	7	7	35
P24	3	3	7	5	5	23
P25	7	7	7	7	7	35
Mean	5.714285714	5.142857143	6.571428571	6	6.19047619	29.61904762
Median	6	6	7	6	7	31
SD	1.616875293	1.768776817	0.978336781	1.095445115	1.030487633	4.852588902