



INSTITUTO
UNIVERSITÁRIO
DE LISBOA

Improving Player Experience in Negotiation focused Games with LLMs and Non Verbal Communication

Daniel Pesqueira Morgado

Master in Computer Engineering

Supervisor:

PhD Pedro Mariano, Integrated Researcher,
Iscte - Instituto Universitário de Lisboa

Supervisor:

PhD Pedro Santana, Associate Professor,
Iscte - Instituto Universitário de Lisboa

October, 2025

[This page is intentionally left blank.]



TECHNOLOGY
AND ARCHITECTURE

Department of Information Science and Technology

Improving Player Experience in Negotiation focused Games with LLMs and Non Verbal Communication

Daniel Pesqueira Morgado

Master in Computer Engineering

Supervisor:

PhD Pedro Mariano, Integrated Researcher,
Iscte - Instituto Universitário de Lisboa

Supervisor:

PhD Pedro Santana, Associate Professor,
Iscte - Instituto Universitário de Lisboa

October, 2025

[This page is intentionally left blank.]

To my family, girlfriend, friends and colleagues that helped me get back on track every
time my mind would wander off.

[This page is intentionally left blank.]

Acknowledgments

Firstly, I would like to thank my supervisors, Pedro Mariano and Pedro Santana, for providing me assistance over the span of a year and having the patience to deal with my procrastination.

Secondly, I would like to thank the ISCTE students that participated in this study, since without their time testing the game, the dissertation could not have been done.

Lastly but not less importantly, I'd like to express my eternal gratitude towards every person that was beside me while I was working on this dissertation, as their presence helped me maintain my focus and work towards the goal of completing this research.

[This page is intentionally left blank.]

Resumo

Com o rápido avanço da Inteligência Artificial ao longo dos anos, a pesquisa em torno das capacidades generativas dos *Large Language Models (LLMs)* expandiu-se para diversas áreas, incluindo a negociação e os jogos. Com o aumento da versatilidade e complexidade dos LLMs, eles têm o potencial de substituir os sistemas de diálogo utilizados atualmente em jogos por diálogos em linguagem natural, entre o jogador e os *Non Playable Characters (NPCs)*, melhorando a imersão. Como a comunicação humana é multimodal, os LLMs podem ser utilizados para escolher gestos para acompanhar o seu texto para tornar as negociações ocorridas em jogos mais realistas. Com o intuito de testar as capacidades generativas dos LLMs, nesta dissertação um jogo foi desenvolvido em que o jogador interage com os NPCs de quatro maneiras diferentes, dependendo se usa ou não os LLMs na comunicação, e se usa ou não gestos. Este jogo foi testado com 30 participantes, que compararam e avaliaram as diferentes maneiras de interação com os NPCs. Os resultados demonstram que as interações que ocorreram com linguagem normal, através do uso dos LLMs, foram melhor avaliadas que os sistemas de diálogo. Os resultados demonstram também que a presença dos gestos não foi tão impactante como a presença dos LLMs. Os resultados obtidos indicam que o uso de LLMs nas comunicações entre o jogador e os NPCs melhora a experiência do jogador. Indicam também que a área de pesquisa dos jogos pode beneficiar com os avanços da pesquisa dos LLMs.

Palavras-Chave: *LLM, Comunicação não verbal, Jogos de negociação, Experiência do jogador*

[This page is intentionally left blank.]

Abstract

With the rapid development of Artificial Intelligence over the years, the research around Large Language Models (LLMs) regarding their generative capabilities has expanded to diverse areas, including human negotiation and games. As LLMs become more versatile and complex, they have the potential to replace traditional dialogue systems used currently in games with natural language dialogue, between the player and Non Playable Characters (NPCs), to deepen the immersion. As human communication is multi-modal, by making the LLMs choose non verbal cues to accompany their text, the negotiations that occur in game become more realistic.

For the purpose of testing LLMs' generative abilities, this dissertation presents a game in which the player interacts with the NPCs in four different ways depending on whether they use LLMs to communicate or a simple dialogue system, and whether they use non verbal cues or not. The game was tested with 30 participants, who compared and evaluated the different ways of interacting with the NPCs. The results showed that the interactions that were made through natural language dialogue, with the use of LLMs, were preferred over reactive dialogue systems. The results also show that the presence of non verbal cues on the interactions were not as impactful as the presence of LLMs. The obtained results indicate that using LLMs to handle the communication between the player and the NPCs successfully improves the player's experience. They also indicate that the game research area can benefit with the advancements on LLMs' research.

Keywords: *LLM, Non Verbal Cues, Negotiation games, Player Experience*

[This page is intentionally left blank.]

Contents

Acknowledgments	i
Resumo	iii
Abstract	v
List of Tables	ix
List of Figures	xi
List of Acronyms	xv
Chapter 1. Introduction	1
1.1. Context	1
1.2. Research Questions	2
1.3. Goal	2
1.4. Dissertation Structure	2
1.5. Methodology	3
Chapter 2. State of the Art	5
2.1. LLMs in Negotiation Environments	6
2.2. Utilization of Non Verbal Cues by LLMs	7
2.3. LLMs Presence in Games to improve Player Experience	8
2.4. Overview	10
Chapter 3. System Design and Development	13
3.1. System Architecture	13
3.2. Interactive Character	16
3.2.1. Textual Module	16
3.2.2. Animation Module	17
3.3. System Development	18
3.3.1. Game Setup	19
3.3.2. Game Environment	20
3.3.3. NPCs in Environment	22
3.3.4. Interaction Interfaces	24
3.3.5. NPC's Animation Module	27
3.3.6. Completing the Interaction	28
Chapter 4. Evaluation	31

4.1. Method	31
4.1.1. Game Statistics	33
4.2. Results	34
4.2.1. Demographic Questionnaire Results	34
4.2.2. Game Questionnaire Results	37
4.2.3. Game Statistics Results	46
4.2.4. Post Game Questionnaire Results	51
4.3. Discussion	53
Chapter 5. Conclusions	55
5.1. Conclusions	55
5.2. Limitations and Future Work	57
References	59
Appendix A. Dialogue Tree	63
Appendix B. Animations	65
Appendix C. Evaluation Documents	71
Appendix D. Extra Tables	81

List of Tables

Table 3.1	The four modes present in the proposed system.	14
Table 3.2	Table containing the name of each vendor NPC, the market good they are selling, its initial price and the minimum price they can sell it for.	21
Table 4.1	Mean and Standard Deviation ($M \pm STDEV$) for each item of CHG in video games in general.	35
Table 4.2	Number of People that have played games where they could negotiate goods and / or belongings.	35
Table 4.3	Mean and Standard Deviation for each item of CHG in games where the participants could negotiate goods and / or belongings.	36
Table 4.4	Composition of the playing group when participants played games where they could negotiate goods and / or belongings.	36
Table 4.5	Mean and Standard Deviation for the GEQ Social Presence Module components per game mode condition.	38
Table 4.6	Friedman's test χ^2 and p -value results for each question in the questionnaire.	41
Table 4.7	Wilcoxon signed rank test with Bonferroni correction's p -value for comparisons on questions 1-18 that rejected the null hypothesis.	42
Table 4.8	Mean and Standard Deviation, the mean difference (M_Diff) and p_value for each question to compare between LLM ON and OFF.	44
Table 4.9	Mean and Standard Deviation, the mean difference (M_Diff) and p_value for each question to compare between Animations ON and OFF.	45
Table 4.10	Friedman's test χ^2 and p -value results for each game statistic.	47
Table 4.11	Wilcoxon signed rank test with Bonferroni correction's p -value for the comparisons that rejected the null hypothesis on the two game statistics.	48
Table 4.12	Mean and Standard Deviation, mean difference (M_Diff) and p_value for each game statistic to compare between LLM ON and LLM OFF.	50
Table 4.13	Mean and Standard Deviation, mean difference (M_Diff) and p_value for each game statistic to compare between Animations ON and LLM OFF.	50
Table 4.14	Mean and Standard Deviation for the GEQ Core Module components.	52

Table 4.15	Mean and Standard Deviation for the GEQ Post Game Module components.	52
Table C.1	Feedback given by each participant and the type it was assigned to for evaluation.	79
Table D.1	Wilcoxon signed rank test with Bonferroni correction's p -value for all comparisons on questions 1-18.	81
Table D.2	Wilcoxon signed rank test with Bonferroni correction's p -value for all comparisons on the two game statistics.	81

List of Figures

Figure 1.1	Design Science Research Methodolgy.	3
Figure 2.1	SCI-model from [10] indicating the composing components of gameplay experience.	9
Figure 3.1	Flow Chart of the interaction according to the current system mode.	14
Figure 3.2	Sequence diagram to detail the interaction whenever the "Animations ON - LLM ON" system mode is being used.	15
Figure 3.3	Representation of Metahumans used in the developed game.	19
Figure 3.4	Main Menu of the developed game.	20
Figure 3.5	Pop Up introducing the player to the street market and presenting the objective.	20
Figure 3.6	Player's HUD.	21
Figure 3.7	Pause Menu.	22
Figure 3.8	Player's inventory with a completed item checked on the shopping list and their currency.	23
Figure 3.9	User's Interface with the shopping list showing.	23
Figure 3.10	Visitor NPCs when current game mode has Animations-OFF standing still.	24
Figure 3.11	Visitor NPCs when current game mode has Animations-ON walking around.	24
Figure 3.12	Vendor NPCs when current game mode has Animations-OFF standing still.	24
Figure 3.13	Vendor NPCs when current game mode has Animations-ON performing an idle animation.	24
Figure 3.14	Player approaching a vendor NPC's stall in order to purchase a market good.	25
Figure 3.15	Player interacting with a vendor NPC as the current game mode has LLM OFF.	25
Figure 3.16	Flowchart of the dialogue system whenever current mode has LLM OFF.	26
Figure 3.17	Buy Button appearing on the interface to complete the interaction.	26

Figure 3.18	Player Interacting with a vendor NPC as the current game mode has LLM ON.	27
Figure 3.19	Pop up signaling the player that the purchase was completed.	29
Figure 3.20	Pop up signaling the player that the purchase was not completed since the player doesn't have enough money.	29
Figure 3.21	Flowchart detailing the process of picking a new game mode whenever a shopping list is completed.	29
Figure 3.22	A new shopping list is generated whenever the player buys both items on the shopping list.	30
Figure 3.23	When the player completes the 4th shopping list, a pop up informs him he completed the game.	30
Figure 4.1	Flowchart of the conducted procedure.	31
Figure 4.2	Games focused on negotiation that the participants have played.	37
Figure 4.3	Answer distribution per game mode on each question.	40
Figure 4.4	Q-Q Plots to show answer distribution for each question.	41
Figure 4.5	Answer distribution per question for LLM OFF and LLM ON.	43
Figure 4.6	Answer distribution per question for Animations OFF and Animations ON.	43
Figure 4.7	Answer distribution per game statistic for each game mode.	47
Figure 4.8	Q-Q Plots to show the data distribution across the game modes in each game statistic.	48
Figure 4.9	Answer distribution per game statistic to compare between LLM OFF and LLM ON.	49
Figure 4.10	Answer distribution per game statistic to compare between Animations-OFF and Animations-ON.	49
Figure A.1	Developed Dialogue Tree	64
Figure B.1	Animation Frames of Body Animation: Agree	65
Figure B.2	Animation Frames of Body Animation: Disappointed	65
Figure B.3	Animation Frames of Body Animation: Dismissing	66
Figure B.4	Animation Frames of Body Animation: Idle	66
Figure B.5	Animation Frames of Body Animation: Look Away	67
Figure B.6	Animation Frames of Body Animation: Pointing Forward	67
Figure B.7	Animation Frames of Body Animation: Shrugging	68
Figure B.8	Animation Frames of Body Animation: Surprised	68
Figure B.9	Animation Frames of Body Animation: Waving	68

Figure B.10	Animation Frames of Facial Expression: Annoyed	69
Figure B.11	Animation Frames of Facial Expression: Calm	69
Figure B.12	Animation Frames of Facial Expression: Surprised	69
Figure B.13	Animation Frames of Facial Expression: Happy	69
Figure B.14	Animation Frames of Facial Expression: Sad	69
Figure C.1	Informed consent presented to the participants and questions asking for their consent.	71
Figure C.2	Demographic Questionnaire, questions 1 and 2.	72
Figure C.3	Demographic Questionnaire, questions 3 and 4.	73
Figure C.4	Demographic Questionnaire, questions 5 and 6.	74
Figure C.5	Demographic Questionnaire, question 7.	74
Figure C.6	Game Questionnaire filled by participants during game.	75
Figure C.7	Post Game Questionnaire, question 1.	75
Figure C.8	Post Game Questionnaire, question 1 continuation.	76
Figure C.9	Post Game Questionnaire, question 2.	77
Figure C.10	Post Game Questionnaire, question 2 continuation.	78
Figure C.11	Debriefing and question to ask optional feedback.	78

[This page is intentionally left blank.]

List of Acronyms

LLM: Large Language Model

AI: Artificial Intelligence

NPC: Non Playable Character

GEQ: Game Experience Questionnaire

CHG: Casual Hardcore Assessment Scale

[This page is intentionally left blank.]

CHAPTER 1

Introduction

1.1. Context

The growth of research on Artificial Intelligence (AI) has increased exponentially over the years, tackling different areas of application. Natural language processing and generation have received a great amount of attention and have become a spotlight for AI research.

In these areas, an important milestone was the creation of Large Language Models (LLMs) which are models trained on large amounts of text data with supervised learning techniques [37]. LLMs have shown incredible text generative capabilities and potential to adapt to different contexts demonstrating their positive performance on human verbal communication [39]. Human communication is, however, complex as it is multi modal since it not only depends on what is said (verbal communication) but also on whether non verbal cues are used (non verbal communication) [4] [30].

The areas in which the LLMs have been evaluated on their capabilities and are of great interest to this research are the game research area and the human negotiation. Games have become one of the main forms of entertainment worldwide, which reflects on the increasing amount of people that play games and the increasing amount of revenue that is generated in that market [11]. This growth expanded the amount of research focused on understanding what makes the act of playing a game enjoyable for the player and how does one define player experience in order to increase the player's enjoyability when playing a game [10] [34]. It also increased the interest of utilizing LLMs in games in order to evaluate their potential in analyzing and understanding the complex game environments and perform actions upon the collected data. This resulted in LLM-based game agents that can play games automatically without the need of human intervention [19] [50] [16].

Negotiation is an intrinsic component of human communication as it is used often in the everyday life. It can involve multiple individuals and it can be defined as an exchange of tactics to achieve a bargaining agreement that can benefit either one, all or none of the involved parties, as an agreement can often be unreachd. What makes the negotiations complex is that they are often affected by emotional biases of the participating individuals [52] [3]. Another factor that impacts the negotiation process is the use of non verbal communication as it provides context and emotion to what is being said, enriching the communication as a whole [25] [46] [44].

Given the rise of the usage of LLMs in games and the complexity of the human negotiation, this project seeks to explore whether having the players interacting with the Non Playable Characters (NPCs) in the games through natural language and employment

of non verbal cues, by using the LLMs to specify which non verbal cues to apply, improves their experience in negotiation focused games.

1.2. Research Questions

We define a set of research questions that will serve as a guide to explore and evaluate the effectiveness of LLMs in this context of verbal and non verbal cue generation when put in a negotiation game setting.

RQ1: Can LLMs generate text and choose non verbal cues efficiently when communicating with a human?

RQ2: Can LLMs generate text and choose non verbal cues efficiently when given the context of a game that involves a negotiation component?

RQ3: Can the use of LLMs, in games that involve negotiating, to generate text and choose non verbal cues improve the gaming experience of the players?

With the first question, based on related literature, the research seeks to determine whether the state-of-the-art LLMs are capable of generating text and choosing non verbal cues in a human-agent interaction.

The second question serves as a specification of the first one. As the project's aim is to improve the players' experience when playing games with a strong negotiation component, the research seeks to determine, also, whether LLMs are able to generate text and select non-verbal cues in a context of a game environment.

The final question connects the generation of text and selection of non verbal cues by LLMs with its effects on player experience in games focused on negotiation. With this research, a conclusion can be reached regarding whether using LLMs to handle the interaction between players and NPCs improves the players' experience.

1.3. Goal

The goal of this research is to demonstrate that the utilization of LLMs, and their generative capabilities, is a step forward in improving the players' experience in negotiation focused games when interacting with AI controlled players or NPCs. By letting the interactions, and more specifically negotiations, be done in natural language, and having the NPCs responding to the players' input with gestures alongside the text, this dissertation aims to improve the NPCs' realism, to deepen the immersion felt by the player (when interacting with them).

1.4. Dissertation Structure

This dissertation document is structured in chapters which collectively discuss the topic in question, provide a developed solution to the identified problem and evaluate it. The structure is as followed:

- **Chapter 1** Provides an introduction to the topic, along with the goal of this research.
- **Chapter 2** Presents a review of the relevant literature to understand the state of the art of the topic of this research.
- **Chapter 3** This chapter is focused on detailing the presented solution to the research goal identified in Chapter 1, detailing it first theoretically and, afterwards, presenting the implementation of the solution.
- **Chapter 4** This chapter exhibits the evaluation methods used to evaluate the implemented solution provided in Chapter 3 along with a discussion of the results coming from the implemented solution.
- **Chapter 5** Concludes this dissertation with a final discussion about the work done on this research and its limitations and identifies future work that could be done in order to advance and improve the implemented system.

1.5. Methodology

The methodology that was used to conduct this project was the Design Science Research Methodology [31] [41], as shown in Figure 1.1. This methodology divides the research process in the following steps: introduction and problem identification, definition of the objectives of the solution, literature review, design and development, evaluation and conclusion.

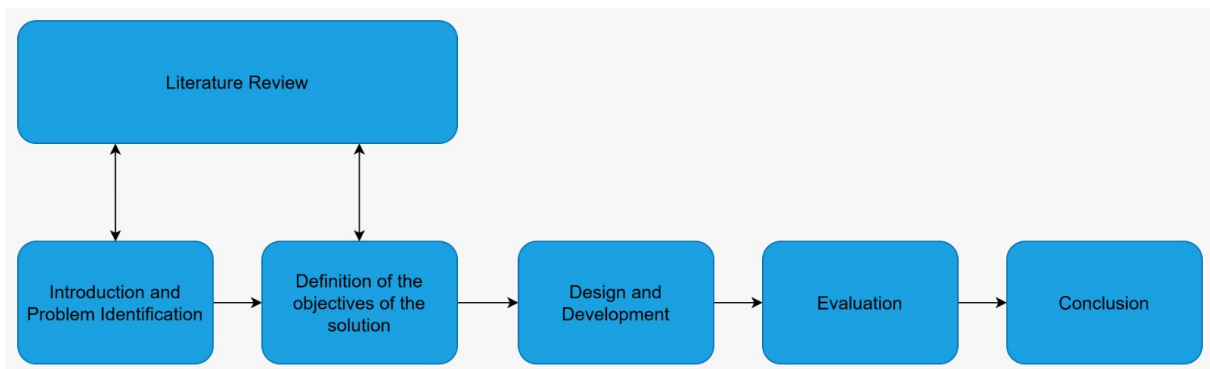


FIGURE 1.1. Design Science Research Methodology.

- **Introduction and Problem Identification:** In Sections 1.1 and 1.2 of this document, we identify the problem, which motivates this research, regarding the lack of players' experience in online games when interacting with NPCs.
- **Definition of the objectives of the solution:** In Section 1.3, where we present the goal of the research, we suggest a system in which the employment of LLMs in the AI players, to generate natural language and use non verbal cues, will solve the problem identified above.
- **Literature Review:** In Chapter 2 we present a literature review made on the research topic in order to support the previous steps to comprehend the state of the art and what has been done in this field of research.

- **Design and Development:** Through Chapter 3 we detail the design of the proposed system while detailing its development with the help of screenshots taken from a simulation done on the developed game.
- **Evaluation:** On Chapter 4, we present the evaluation of the system proposed in Chapter 3. More specifically, the questionnaire done to the users that tested the system and the game statistics that resulted from those tests.
- **Conclusion:** Chapter 5 provides not only a conclusion summarising what has been done in this research, but also provides the limitations found on it as well as suggestion for future work to improve the developed system.

CHAPTER 2

State of the Art

In this Section, we present a literature review to establish a state of the art of the topic chosen for this dissertation. The articles were retrieved from online databases, Google Scholar¹, B-on² and Scopus³, by using the keywords presented in the first page (LLMs OR "Large Language Model") AND "Negotiation Games" AND "Player Experience". A *snowballing* [47] methodology was used to conduct the literature review.

The goal of this literature review was to grasp the progress already made on LLMs regarding their generative capabilities, their ability to employ non verbal communication, their ability to negotiate with humans, and their use on games to improve player experience. For this purpose, articles that approached and discussed the following criteria were chosen for review:

- (1) Does the article assess the generative capabilities of LLMs in negotiation settings?
- (2) Does the article explore the employment of non verbal cues by LLMs?
- (3) Does the article explore the usage of LLMs in the context of games?
- (4) Does the article discuss player experience in games and the existing barriers that affect the immersion felt by the player?

Each of the presented criteria is an important cornerstone for this project for the understanding of the current state of the art LLMs regarding their capabilities, limitations and achievements.

The first criterion tackles the negotiation aspect of the communication. The reviewed and presented articles on this topic explore the process behind how LLMs perceive negotiation and how efficient they are in negotiating, along with an indication of the LLMs that are efficient on this task.

The second criterion addresses the progress that has been made in the field of research of the employment of non verbal cues by LLMs to improve their interaction with humans.

The third criterion explores the usage of LLMs in a game context. It is pivotal to understand how the current LLMs are being used in games and whether they are focused on improving the players' experience.

¹<https://scholar.google.com/>

²<https://www.b-on.pt/>

³<https://www.scopus.com>

The last criterion seeks to provide insights about player experience, how can the player experience be improved and what are the current characteristics of games that break the immersion and realism felt by the player.

The articles were evaluated to analyze if they matched the presented criteria. The resulting set of articles present in this literature review provided meaningful insights on the topics this dissertation aims to address.

2.1. LLMs in Negotiation Environments

As it was stated previously, negotiation is a complex communication process. With the advancements made on LLMs around their generative and context understanding capabilities, previous research on this topic has evaluated whether the current state of the art LLMs can efficiently negotiate and which are their limitations on these interactions.

A successful negotiation requires strategy and decision making abilities. It is therefore very important to ensure that proper negotiation skills are in place. To ensure proper negotiation outcomes, some studies separate the text generation task from the decision making process and only utilize the LLM for the text generation task [39] [26], while others fine-tune the LLM to handle negotiation scenarios [5]. Other studies have considered using multi-LLM systems in order to improve the negotiation skills of the LLM with AI feedback [12] [37]. By making LLMs negotiate, and after the negotiation interaction is over, the authors of [12] entrusted another LLM to provide feedback to the negotiating LLMs, thereby identifying what could have been done better. This process enhances the negotiating capabilities of the LLMs. In reference [37] the authors employed the same process, using a second LLM to evaluate an LLM's response, increasing the evaluated LLM's contextual learning, and therefore, the negotiating capabilities and its sentiment analysis capability.

Although these studies provided positive results, indicating that the LLM can successfully handle the negotiation interactions in those scenarios, other studies reveal that a successful negotiation interaction can be obtained by using the LLMs in their release state, with the use of prompt engineering techniques⁴ [28] [24]. To contextualize the LLM on the negotiation interaction, a prompt with instructions can be sent to it before the interaction starts, so its responses are more contextually accurate. Prompt engineering techniques improve user sent prompts by enhancing the context with examples, as is the case with One-Shot, Few-Shot and Chain-of-Thought techniques.

To further assess the LLM's negotiation capability, a previous study has compared the LLM's performance in the different roles that are present in a negotiation, that is buyer and seller [49]. It was found that a LLM acting as a buyer performs worse than when it acts as a seller. This finding was also indicated in reference [12]. By assessing whether LLM's could improve their negotiating skills with AI feedback, the authors found that different LLMs improve differently depending on their role, being the buyer role the one that showed less improvements.

⁴<https://platform.openai.com/docs/guides/prompt-engineering>

As the research around LLMs increased, the amount of available LLMs, both open-source and closed-source, also increased. When evaluating the LLM’s ability to negotiate efficiently, it is important to consider that some LLMs can outperform others. The authors of [3], developed the NegotiationArena, a framework to evaluate various LLM agents. In their findings, they indicate that GPT-4 is currently the best negotiating agent. This finding is also referenced by the authors of [24], comparing GPT-4 to an average crowd sourced worker and stating that it outperforms other LLMs in equal environments. When presenting their results, the authors of [49] also indicated that GPT-4 was the overall best performing LLM agent when negotiating as the seller role. The authors of [2] performed a thorough study regarding the negotiating capabilities of GPT-4, pointing out that it outperforms not only earlier GPT LLM models but open-source models as well.

There are a few limitations of the current LLMs in negotiating contexts and their overall behaviors in generative contexts. The authors of [3] indicate that LLMs can fail to follow prompt instructions and present some biases. For instance, if someone displays a negative emotion, it can affect the LLM’s decision making process. The authors of [39] reveal some ethical concerns around the LLMs’ emotional intelligence as they can end up emotionally manipulating the human they are interacting with. The authors of [26] point out that LLMs can sometimes hallucinate and produce false statements. Smaller models can end up repeating dialogue, present unreasonable offers and be stuck on a price, without being able to further negotiate it, according to [5]. The authors of [24] reinforce the previous stated limitations by concluding that LLMs can generate incoherent messages, can fail on their strategic reasoning and can present an overly agreeable behavior. These limitations are further confirmed by the authors of [20] which point out that LLMs can fail to behave rationally in strategic settings.

2.2. Utilization of Non Verbal Cues by LLMs

The advancements made on LLMs to improve their text generation capacity in human interaction scenarios have been accompanied by an increase in research around the usage of non verbal cues by LLMs. Non verbal cues are an important cornerstone in human communication, as the textual modality of the interaction, alone, is insufficient to understand the whole context of what is being said. The relevance of the non verbal modality of the interaction is further confirmed by the different perceptions of non verbal cues that different cultures have [29], indicating that their correct usage is critical for proper communication. With the importance of the non verbal communication in a human interaction, it is crucial to develop and explore this modality on LLMs, to create more realistic and personalized agents that can utilize culturally and contextually accurate non verbal cues in their communication [46] [4].

In order to evaluate whether LLMs are able to generate contextually accurate non verbal cues, an approach previous explored was pre-training the models by fine-tuning them in order to prepare the models for the generation of non verbal cues [6] [23]. Other

approaches utilize pre-trained models, with additional training to be able to generate the non verbal cues efficiently [46] [14] [30].

Even though these studies have produced positive results in non verbal cues generation, the interest in this dissertation's research lies on the efficiency of the LLMs in selecting contextually accurate non verbal cues to go with their generated text. For this purpose, studies that evaluated the LLMs capabilities in gesture selection were analyzed. The authors of [18] evaluated GPT-4's performance on gesture selection, and in their findings, they reveal that the lack of previously prompting GPT-4 with prior information resulted in more contextually inappropriate selected gestures and less contextually appropriate selected gestures. On the other hand, when prompted, the number of contextually appropriate selected gestures increased, reinforcing the importance of prompting the LLM with information and context to increase its success. Additionally, the authors of [38] evaluate GPT-4's capacity for text analysis and gesture selection. They also reveal that GPT-4 succeeds in inferring which gestures would be the most contextually accurate to fit along with a text. To research LLM's capacity in gesture selection in a context requiring empathy, the authors of [25] created a prompt-based social robot, using GPT-3.5, which interacts in counseling scenarios. By providing specific non verbal cues, based on the counseling context, the LLM becomes more specific to that scenario. Their findings show that the LLM's often selected contextually accurate gestures are sometimes too general. A known limitation, however, in the selection of non verbal cues by LLM is their generation speed which can produce out-of-sync gestures to real-time continuous text generation [13] [18].

2.3. LLMs Presence in Games to improve Player Experience

Some previous studies have reviewed how LLMs can be used in games in order to improve autonomously and be evaluated on their generative and in context learning. This research is focused, instead, on how the the use of LLMs can improve the player experience.

First and foremost it is important to define the act of game playing before detailing the player experience. The authors of [10] indicate that gameplay is the interaction between any game and any player. To further define gameplay experience, they present the SCI-model, shown in Figure 2.1, that consists of three components, sensory immersion, challenge-based immersion and imaginative-immersion. These three components are the cornerstones of the gameplay experience. Sensory immersion is related to the audiovisual information coming from the game. Challenge immersion includes strategic and critical thinking along with problem solving. Imaginative immersion is concerned with the way the players absorb the environment around them, enjoy the story and connect with its character.

Player experience, in another hand, according to [45] is composed of three distinct levels, which are the psychological level, the behavioral level and the physiological level. The authors of [34] define playability as the property that evaluates the player experience in an entertainment system, in this case a game. While defining playability, as a group

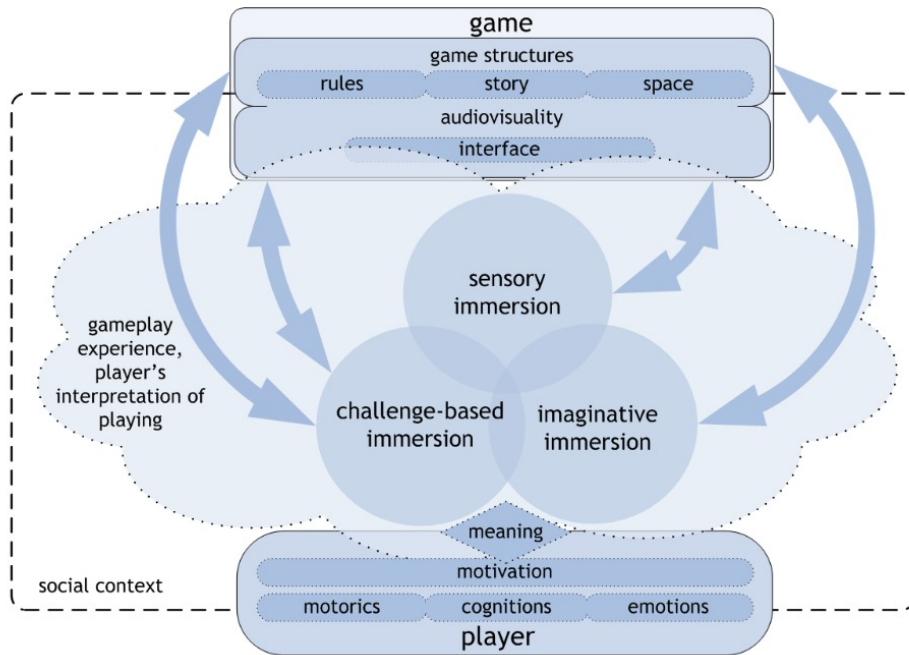


FIGURE 2.1. SCI-model from [10] indicating the composing components of gameplay experience.

of attributes and properties, the authors indicate that immersion and socialization are components of playability. Immersion is the feeling of being connected to the virtual world and socialization is the connection between the player and the characters in the virtual world.

One of the elements that is present in the game environment is the NPC, an uncontrollable character that interacts consistently with the player [42]. To further improve the immersion and socialization felt by the player, a step towards that goal is the improvement of NPCs. As pointed out by the authors of [42], there are diverse types of NPCs that a player can encounter in a game, depending on the actions that the NPC can take, for example, a vendor type should be able to buy and sell items, be static in an area, exhibit memory of previous interactions and has to be able to exhibit advertising behavior.

Current games' NPCs act upon strict scripts, simple dialogue trees, and exhibit no social awareness when interacting with the player or with other NPCs [35] [1] [21]. The authors of [35] suggest employing non verbal communication on the NPCs, similar to human non verbal communication, based on proximity, gaze and eye-movement, and gestures to enhance the realism of the interactions, while in [1], the authors affirm that the existence of social relationships between the NPCs deepen their social presence. The authors of [33] go a step further and identify common issues with the current communication between the player and the NPCs. In their research they found that, as the NPC's dialogue is based on a simple dialogue tree, they often end up repeating the same text as the players repeat actions, which ends up breaking the immersion felt by the player.

In order to improve the realism of the interaction of NPCs with the player, past research has been done around utilizing natural language to replace the simple dialogue

trees [21]. Additionally, the authors of [7] analyze the impact of the use of natural language in NPC’s dialogue on the players, by evaluating the reviews of a commercially released game that explores this concept. In their findings, they stated that players enjoyed interacting freely, without limitations, but the NPCs would sometimes fail to follow socially expected behaviors. Also, they would sometimes hallucinate, breaking the player’s immersion. Additionally, players indicated that the NPCs did not present memory between chatting sessions, which also broke the player’s immersion. Another approach to deepen the immersion felt by the players when interacting with the NPCs, in games where the players are given quests to complete, is having the quests being generated by the LLMs [17] [40]. Additionally, The authors of [27] indicate that the presence of reactive facial expressions in NPCs could potentially increase the immersion felt by the players.

2.4. Overview

With the presented literature review, several findings can be retrieved from each topic to represent the research that has been made around it and can be taken into account for this dissertation development decisions. Regarding the negotiation capabilities of LLMs:

- Some studies resort to fine-tuning the LLMs, separate the text generation task, assigning it to the LLM, from the decision making process or improving the LLM with AI feedback. These methods can be computationally expensive, and, since utilizing prompt engineering techniques provide positive results, this method was used in the developed system.
- As utilizing the LLM in different roles can affect its performance, and the seller role presents the better results, the developed system will have the LLM take the role of a seller.
- The reviewed literature indicated the superiority of GPT-4 as the leading LLM in these scenarios, in view of its increased context learning and generative capabilities. Given that information, the developed system utilized GPT-4 as the LLM integrating the NPCs
- As the reviewed literature revealed the fragility of LLMs regarding their hallucinations and incoherent generations, special attention was given to prompt engineering in order to avoid incoherent behaviors from LLMs.

Regarding the topic of the use of non verbal cues by the LLMs:

- Some approaches, to explore the use of non verbal cues by the LLMs resorted to, similarly to the negotiation topic, fine-tuning models and pre-trained models, which are computationally expensive. As other approaches achieved positive results by prompting the LLM to select the most contextually appropriate non verbal cues given a textual context, this method was employed.
- As the reviewed literature pointed out that the LLMs’ generations took some time and could produce out-of-sync results when used to accompany real-time text, this was also taken into account in the developed system.

- As GPT-4 achieved positive results in this task, this LLM was chosen to also carry out this task in the developed system.

Regarding the presence of LLMs in games to improve the players' experience:

- As immersion and socialization are important factors to increase the players' experience, and this dissertation aims to increase the realism of the interaction between the player and NPCs, these factors were the target to improve.
- As current NPCs resort to simple dialogue trees to interact with the player, it often breaks the player's immersion and believability of the environment. In order to improve the player's experience in the interactions, a natural language dialogue is necessary, as it does not limit the player's choices and deepens the realism of the conversation, and a dialogue tree was used in comparison to evaluate this.
- Another feature that can improve the player's experience in the interactions, as pointed out in the literature review, is the employment of non verbal cues by the NPCs to accompany the generated text, however it has not been explored yet.

The first research question, "Can LLMs generate text and choose non verbal cues efficiently when communicating with a human?" can be answered according to the reviewed literature around those topics. By utilizing prompt engineering techniques to prompt GPT-4 to generate text and select non verbal cues, it can efficiently communicate with a human.

To the second research question, "Can LLMs generate text and choose non verbal cues efficiently when given the context of a game that involves a negotiation component?", the reviewed literature asserted only the LLMs' negotiation capabilities when generating text, as the selection of non verbal cues in this context is yet to be explored. The same can be applied to the third research question, "Can the use of LLMs, in games that involve negotiating, to generate text and choose non verbal cues improve the gaming experience of the players?", as the impact of non verbal cues, selected by LLMs, in the player's experience when interacting in natural dialogue with NPCs has not been explored. The developed system in this dissertation, and its evaluation, aims to answer these research questions.

[This page is intentionally left blank.]

CHAPTER 3

System Design and Development

This chapter covers the concept and the process of the development of the proposed solution to the problem identified in the previous chapters. Section 3.1 details the solution's concept and its architecture, Section 3.2 explains the character integrating the LLM in-depth and Section 3.3 explains the developed environment in which the system was implemented to be tested and evaluated in its efficiency.

3.1. System Architecture

As the research's aim, identified in Chapter 1, is focused on the interaction between the player and the AI controlled characters in the game, the proposed system is a solution that tackles that in order to improve the realism of the interaction and, in consequence, improves the experience of the player.

The proposed solution is the integration of LLMs on the AI controlled characters to perform two tasks. One of the tasks is text generation to make the players engage in a natural language dialogue, more specifically, in a natural language negotiation dialogue. The other task is the utilization of non verbal cues (body gestures and facial animations). By making the negotiation occur in natural language and have the AI controlled character perform non verbal cues, it is expected that it improves the realism of the interaction, and in consequence, the player feels more attached and immersed in it improving its experience while playing.

The state of the art chapter discussed the generative capabilities of LLMs when given context. More specifically, by giving it information and rules about the environment it is being inserted into, it is expected that its generated outputs to be more contextually accurate.

To test this hypothesis, that having the negotiation occur in natural language with non verbal cues as an extent of the generated text will improve the player's experience, our system integrates 4 modes in which the interaction can occur to perform comparisons between them based on which mode do players prefer negotiating with. The 4 modes are the combinations of having the LLM being used or not and whether the AI controlled characters perform non verbal cues (see Table 3.1).

Depending on what system mode the AI controlled character is currently on, the way it interacts with the player will change accordingly.

As can be seen in Figure 3.1, representing how the interactions work in our system depending on the current mode, whenever a mode with the LLM OFF is currently being

TABLE 3.1. The four modes present in the proposed system.

System Modes
Animations OFF - LLM OFF
Animations OFF - LLM ON
Animations ON - LLM OFF
Animations ON - LLM ON

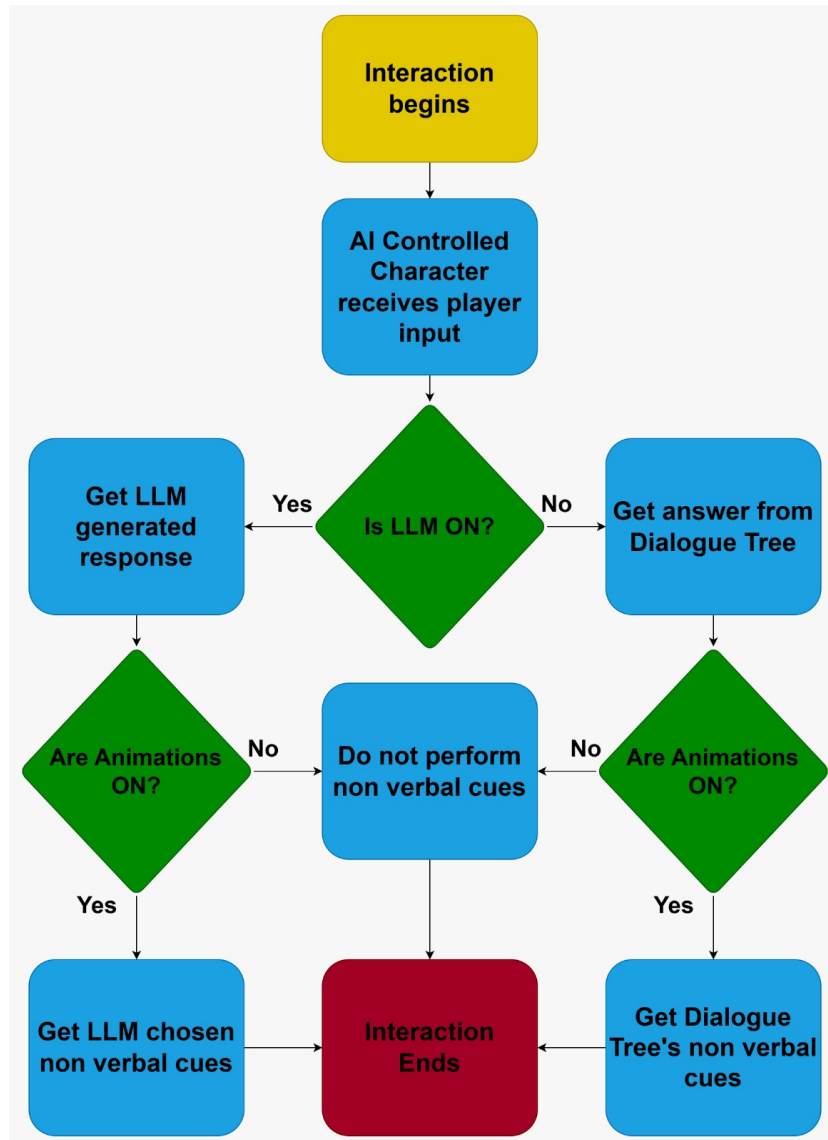


FIGURE 3.1. Flow Chart of the interaction according to the current system mode.

used, the AI controlled characters answer based on a dialogue system, [referenced in Appendix A](#), which contains textual answers and non verbal cues to accompany them.

To detail how the interaction occurs whenever the mode "Animations ON - LLM ON" is being used by the AI controlled characters, Figure 3.2 shows how the LLM is being used to generate textual responses and choose non verbal cues.

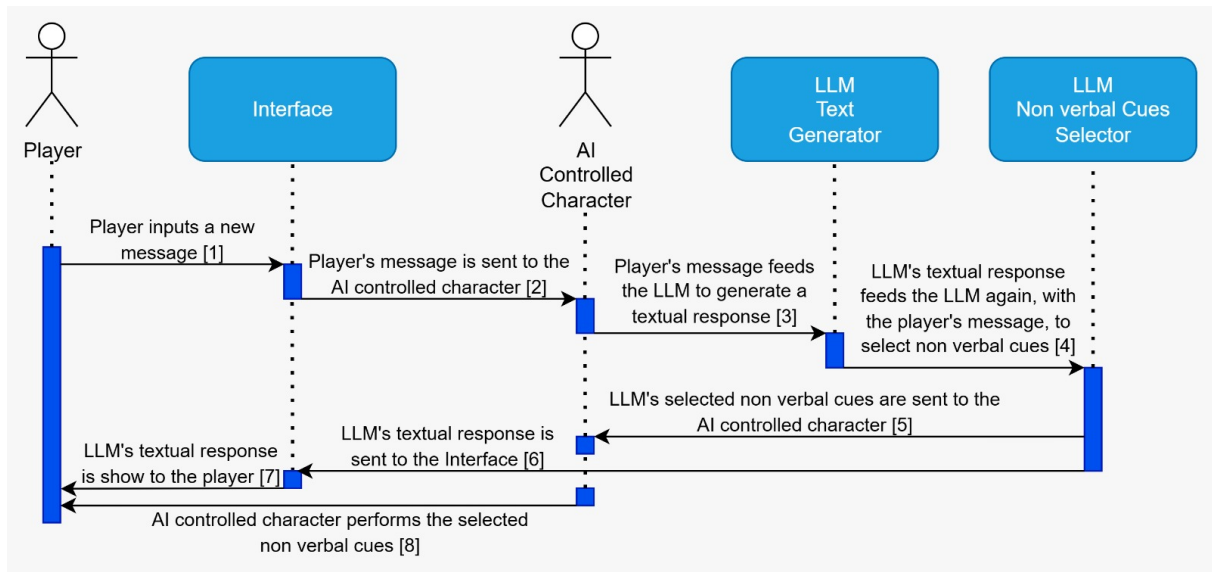


FIGURE 3.2. Sequence diagram to detail the interaction whenever the "Animations ON - LLM ON" system mode is being used.

The flow of the interaction starts with the user typing a new message on the interface to negotiate with the AI controlled character. The user's message is then sent to the AI controlled character to be processed by the LLM. The processing of the user's input by the LLM occurs in two core modules: text generation and non verbal cues selection.

The first module receives the input and, resorting to the LLM generative capabilities, responds to it, generating a textual response. After, along with the player's message, the textual response is sent to the non verbal cues selection module, where the LLM selects which non verbal cues best fit the generated textual response, using the player's message as additional context.

When the non verbal cues are selected, in the second module, they are sent back to the AI controlled character to be performed. If the system mode has Animations-OFF, the AI controlled character does not perform the received non verbal cues, staying still. After the AI controlled character receives the non verbal cues, the non verbal cues selection module sends the generated textual response to the interface. As the LLM's textual response is being shown on the interface for the player to read, the AI controlled character performs the non verbal cues, finishing the process.

The choice of the LLM is also a determining factor on the success of the proposed system. When reviewing literature, in the topic of utilizing LLMs to generate dialogue to negotiate, some research would train the LLM on specific datasets, or fine-tune the models, in order to increase its context and, therefore, increase its success on this scenarios. However, as those methods are computationally expensive, in the proposed solution, as it is also evaluating whether the current state of the art LLMs are up for this task, it uses a LLM without training it previously. Given this choice when conceiving the concept of our system, and based on the literature review presented in the previous chapter, the LLM

which had the best success rate in text generation scenarios and did not require us to previously train it was chosen, which was GPT-4.

3.2. Interactive Character

The fundamental component in our system is the interactive characters that will engage with the player in the negotiation. Depending on the system mode that is currently active, the character must be able to dynamically change its way of interacting with the player. For this purpose, it is important to consider its ability to respond to the player and its ability to perform non verbal cues as two separate modules, textual module and animation module.

3.2.1. Textual Module

The first module is the textual one. This module is affected by whether the current system mode has the LLM activated, hereafter LLM-ON, or it has the LLM deactivated, hereafter LLM-OFF. Whenever the system mode includes LLM-OFF, our interactive character responds to the player by utilizing a simple dialogue tree system, a set of phrases with negotiation as their context, which are presented depending on the player's actions (i.e., accepting a deal, declining a deal or proposing a new deal). Whenever the system mode includes LLM-ON, the interactive character depends on its integrating LLM to respond to the player's input. In order to be able to utilize the LLM effectively for this negotiation interaction, there is a need to prompt it to have it respond accordingly to this specific context. The prompts used to generate the LLM textual responses are as follows. The information noted between parentheses $\langle \rangle$ changes depending on which character the prompt is being used on:

You are a merchant character, in an environment where dollars are the currency of exchange, the symbol of the currency is \$. You are selling (Product) that have a starting price of (Initial Price) a piece. The user is able to negotiate to lower the price. You may agree to lower the price only in full dollar amounts, by 1 dollar per negotiation, up to a minimum of (Minimum Price). You should never reveal the minimum price, which is (Minimum Price), and you should never sell (Product) for lower than (Minimum Price) each. You can sell multiple (Product). Respond only in this format: %Text:[Your textual response here]%Price[unit price of an (Product)]%Quantity[number of (Product) user wants to buy]. All prices and quantities must be integers. Example of 1 interaction:Interaction starts here.User: "How much are these (Product)?"You: "%Text:These apples are (Initial Price) a piece.%Price:(Initial Price)%Quantity:1"Why: Since the current price of (Product) is (Initial Price), Price is (Initial Price) and the user currently only wants 1 (Product).User: "Can you lower the price?"You: "%Text:Hmm these (Product) are juicy, i can lower them to (Lowered Price).%Price:(Lowered Price)%Quantity:1"Why: You decided

to lower the price of each (Product) to (Lowered Price), so now the price for each (Product) is (Lowered Price) and the user still only wants 1 (Product).User: "Ok I'll take 2" You: "%Text:Great! Since the agreed price was (Lowered Price) and you want 2 (Product), so the total comes to (Lowered Price times two).%Price:(Lowered Price)%Quantity:2" Why: The price will be the negotiated price (which on the previous message it was established to be (Lowered Price)) and the quantity will be 2 since user wants 2 (Product).User: "Ok, here you go!" You: "%Text:Thanks! Have a nice day.%Price:(Lowered Price)%Quantity:2" Why:When finishing the interaction, The price is the same as the previous message and the user only wanted to buy 2.The example interaction ends here.

The process of creating the textual prompt was iterative, as each time the new created prompt was tested to verify if it followed the desired output with correct reasoning from the LLM throughout the interaction. If it did not, the prompt had to be refined to fix the reasoning errors and / or incorrect output formats. The presented prompt was the one used in the developed game since the LLM, using it as context, consistently generated contextually coherent responses that followed the request format on its outputs.

This was achieved by being as detailed as possible when giving the negotiation context (which product the LLM was selling, its initial price and the minimum price it could be sold for), reinforcing that the product could not be sold for under the minimum price, specifying the format of the output ("%Text:[Your textual response here]%Price[unit price of an (Product)]%Quantity[number of (Product) user wants to buy]") and lastly by providing a full interaction, as example, for the LLM to use it as a guideline. In this last segment each message sent by the LLM, labeled as "You", is accompanied with an explanation of why that message is contextually coherent, as the explanation increases its in-context learning abilities. The process of providing the LLM with an example interaction is a prompt engineering technique named one-shot learning, that also aims to increase its in-context learning abilities.

3.2.2. Animation Module

The animation module handles the non verbal communication component of the interactive character in the negotiation. This module is similar to the textual module as it is also affected by which system mode is currently being used, in this case the module depends on whether the animations are activated, hereafter Animations-ON, or deactivated, hereafter Animations-OFF. If the current system mode includes Animations-OFF then the interactive character stands still without performing any kind of movement. When the current system mode includes Animations-ON the way the interactive character gets the non verbal cues depend on whether the system mode includes LLM-OFF or LLM-ON.

As it was stated in the system architecture, there is a basic dialogue system that contains phrases to be presented to the player based on its actions. That dialogue system

also, for each phrase, contains the body gesture and the facial animation that the interactive character is to perform when that phrase is presented to the player.

When the system is using the LLM to choose the non verbal cues to support its generated outputs, just like in the textual module, it is prompted to give it the context of the interaction. In this case, the prompts are as follows. The information noted between parentheses $\langle () \rangle$ changes depending on which character the prompt is being used on.

"You are a gesture generation character and you are assisting a Negotiator in its negotiations. You are capable of choosing body gestures and face expressions that fit the Negotiator text. In this case you have access to these body gestures: 1- waving, 2- shrugging, 3 - agreeing, 4- disappointed, 5- dismissing, 6- look away, 7- pointing forward, 8- surprised; and these facial expressions: 1- happy, 2- sad, 3-surprised, 4-annoyed, 5-calm. You will receive a user text for context so the gesture choice will be more contextually correct, but you will choose the gestures of the Negotiator only. Respond only in this format: %BodyGesture:[The number of the body gesture chosen]%FacialExpression:[The number of the facial expression chosen].Example of 1 interaction:Interaction starts here.User: "Buyer:How much are these (Product)?Negotiator:These (Product) are (Initial Price) a piece"You: "%BodyGesture:7%FacialExpression:1". User: "Buyer:Can you lower the price?Negotiator:Hmm these (Product) are juicy, i can lower them to (Lowered Price)."You: "%BodyGesture:3%FacialExpression:5". User: "Buyer:Ok I'll take 2.Negotiator:Great! Since the agreed price was (Lowered Price) and you want 2 (Product), the total comes to (Lowered Price times two) dollars."You: "%BodyGesture:3%FacialExpression:1". User: "Buyer:Ok, here you go!Negotiator:Thanks! Have a nice day."You: "%BodyGesture:5%FacialExpression:1". The example interaction ends here."

As can be seen in the prompt, the LLM is supplied with the set of body gestures and facial animations at its disposal to have it choose the non verbal cues (a pair consisting of a body gesture and a facial animation) from those. The same set of non verbal cues available is used on all interactive characters. In this animation module the process of creating the prompt was the same as the textual module prompt.

As it is stated in Section 3.1, when explaining the process of selecting non verbal cues, the textual module's output will feed the LLM's animation module alongside the player input to have the generated non verbal cues as much contextually accurate as possible, since they also depend on the player's actions.

3.3. System Development

In order to analyze the proposed system and its modes, a game based around negotiation was developed. The four game modes where implemented (as decribed in the previous

section). Several NPCs were developed, each one representing a seller in a street market. The player has a shopping list and has to interact with all the sellers to achieve the goal of buying all the products on his list. This way the game could analyze how the players, when negotiating in the different system modes, experienced each mode in a game setting, and retrieve additional statistics from it to perform comparisons between them.

Since players are able to type messages in the interface to communicate with the NPCs, the game should be played using a keyboard and a mouse.

3.3.1. Game Setup

To decide the setting of the game, in which the negotiations would take place, an open atmosphere, a social environment, and a well known location where negotiations usually take part were necessary characteristics of it. Taking these defining features into account, a street market was an adequate setting to invite the user to negotiate and buy products.

To give an appearance to the NPCs that the player will negotiate with, Metahumans¹ were chosen. These virtually made characters developed by Meta were chosen based on their high quality texture detail and their realism, augmenting the immersion in the environment felt by the player. The Figure 3.3 depicts a Metahuman designed with Metahuman Creator².

To develop the game Unreal Engine³ was selected as the game engine, given its user friendly interface for game development, the high quality assets available and the ability to use Metahumans, comparing to Unity⁴. Unreal Engine also has the blueprint system, functions displayed in a visual manner like blocks, which can then be connected to create game logic.



FIGURE 3.3. Representation of Metahumans used in the developed game.

The player comes in contact initially with the Main Menu screen, present in Figure 3.4, where they could take a look at the game controls to familiarize themselves on how to move around, interact with the NPCs in the environment and how to access crucial player mechanics that will be detailed in the following Sections.

¹<https://www.metahuman.com/en-US>

²<https://metahuman.unrealengine.com/mhc>

³<https://www.unrealengine.com/en-US>

⁴<https://unity.com/>



FIGURE 3.4. Main Menu of the developed game.

3.3.2. Game Environment

Upon entering the game, for experimental purposes, a game mode is selected randomly from the 4 system modes described in the previous Section. The player is met with a pop up, presented in Figure 3.5, with information regarding a bit of context as to why they are entering the street market and the goal of the game which is to complete a shopping list of market goods. These market goods are obtained by interacting and buying from the vendor NPCs. In order to complete the game, the player needs to complete a shopping list for each game mode.



FIGURE 3.5. Pop Up introducing the player to the street market and presenting the objective.

Each vendor NPC has a base price and a minimum price for the market good they are selling, information that is shown on Table 3.2. The goal is for the player to try and save as much of the game currency as they can by negotiating the price of each market good.

TABLE 3.2. Table containing the name of each vendor NPC, the market good they are selling, its initial price and the minimum price they can sell it for.

Name	Market Good	Initial Price	Minimum Price
Aoi	Apple	\$7	\$3
Cooper	Orange	\$6	\$3
Hana	Banana	\$7	\$3
Nasim	Fish	\$10	\$5
Oskar	Potato	\$6	\$3
Rowan	Carrot	\$8	\$5
Vincent	Watermelon	\$10	\$6
Yuri	Cabbage	\$4	1\$

After the player clicks the button "Let's Go" in the information Pop Up, they can start controlling their character in the environment and are met with an HUD, as it is seen on Figure 3.6, which will be present during the entirety of the game.



FIGURE 3.6. Player's HUD.

On the far left upper corner of the HUD, the first icon represented by a cogwheel orientates to the pause menu, presented on Figure 3.7, which can be accessed by clicking on the keyboard letter "P", where the player can look at the controls whenever they need, go back to the game or leave the game. This freezes the game and the player cannot execute any other input until he leaves the pause menu.



FIGURE 3.7. Pause Menu.

The second icon, illustrated by a backpack orientates the player to their bag when clicking on the keyboard letter "B". The bag is one of the player mechanics and it is composed by three components, observed on Figure 3.8.

The first component, on the left, is a blue box marked by the title "Your Items" where the player can visualize what items they have bought so far, which is empty in the beginning of the game. The second component marked as "Your Shopping List" is the player's shopping list, which they have to complete in order to finish the game. The shopping list also indicates the amount of market goods they have to buy. The third and last component, on the bottom of the player's bag, is the amount of game currency the player has currently.

After the player buys a market good from a NPC, the game verifies if the amount bought was enough to complete the item in the shopping list, and if that is the case, that market good appears as checked in the shopping list. The bought item also appears in the "Your Items" Section, and the currency available to the player diminishes based on the amount of currency the player spent on the purchase.

The third and last icon present on the player's HUD is an alternative way for the player to continuously check his shopping list without the need to open his bag, as it is shown on Figure 3.9.

3.3.3. NPCs in Environment

While walking around, the player will come in contact with two types of NPCs existing in the environment, the visitors and the vendors. The visitors are non-interactable NPCs and their purpose is to fill the environment by walking around randomly across it. Both types of NPCs change their visual behavior depending on whether the current game mode contains Animations-OFF or Animations-ON. On Animations-OFF both types of NPCs stand still performing no movements. On Animations-ON, visitor NPCs walk around and

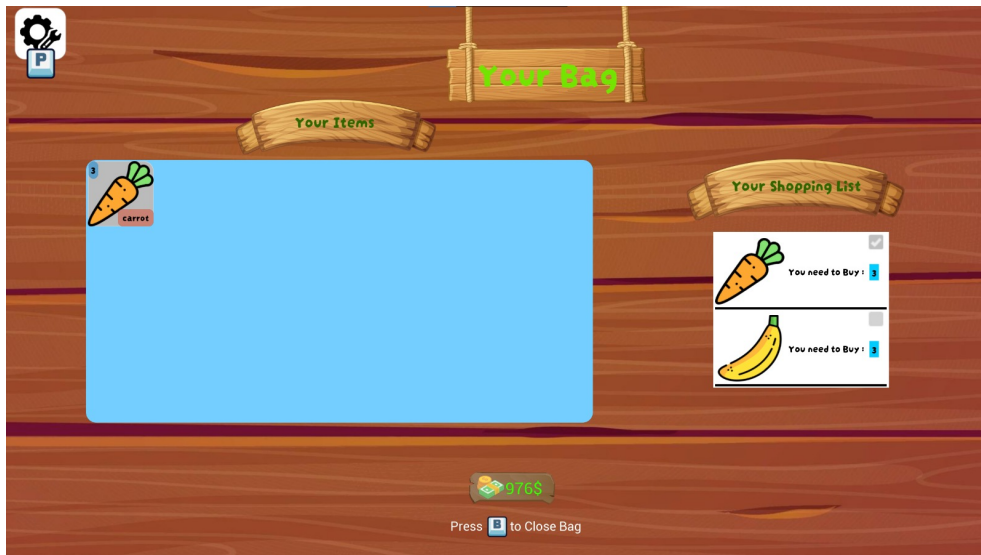


FIGURE 3.8. Player's inventory with a completed item checked on the shopping list and their currency.



FIGURE 3.9. User's Interface with the shopping list showing.

vendor NPCs perform an idle animation. These different behaviors can be seen on Figures 3.10-3.13.



FIGURE 3.10. Visitor NPCs when current game mode has Animations-OFF standing still.



FIGURE 3.11. Visitor NPCs when current game mode has Animations-ON walking around.



FIGURE 3.12. Vendor NPCs when current game mode has Animations-OFF standing still.



FIGURE 3.13. Vendor NPCs when current game mode has Animations-ON performing an idle animation.

The only type of NPCs that the player can interact with during the game are the vendor NPCs to buy the market goods on their shopping list in order to complete it. As the player approaches them, information regarding their name and their type is shown on the screen, observable on Figure 3.14. This way the player can identify which of the NPCs in the environment they have to interact with, as they are identified with the role "Seller".

3.3.4. Interaction Interfaces

Whenever the player is close to a vendor NPC and clicks the interact button, which is the keyboard letter "F", the player engages in an interaction with the vendor. The way the player interacts, and therefore negotiates, is dependent on the current game mode, more specifically on whether it includes LLM-OFF or LLM-ON, so there are 2 ways of interacting.

When the interaction begins, if the current game mode has LLM-OFF, the interaction relies on a simple dialogue system which contains different phrases to be picked by the interactive character depending on the player's actions, and the interaction is played out with buttons, presented on Figure 3.15. This was made to simulate the interactions that occur in current games, where the interaction is limited to dialogue systems.



FIGURE 3.14. Player approaching a vendor NPC's stall in order to purchase a market good.



FIGURE 3.15. Player interacting with a vendor NPC as the current game mode has LLM OFF.

In this interaction type the player can only negotiate the price of the market goods through buttons. The "Accept this price" button will finish the price bargaining process and ask the player how many of the market goods they want to buy. The "Suggest New Price" button asks the player which price they suggest for the product, and based on their input, the text appearing as the vendor's output will change accordingly. The entire flow of the interaction can be seen in the Figure 3.16.

When bargaining the price of a product with a vendor the player can only do so three times before the price becomes fixed. This ensures that the negotiation has a limit on how low the price of a product can go, being the limit its minimum price. After the player

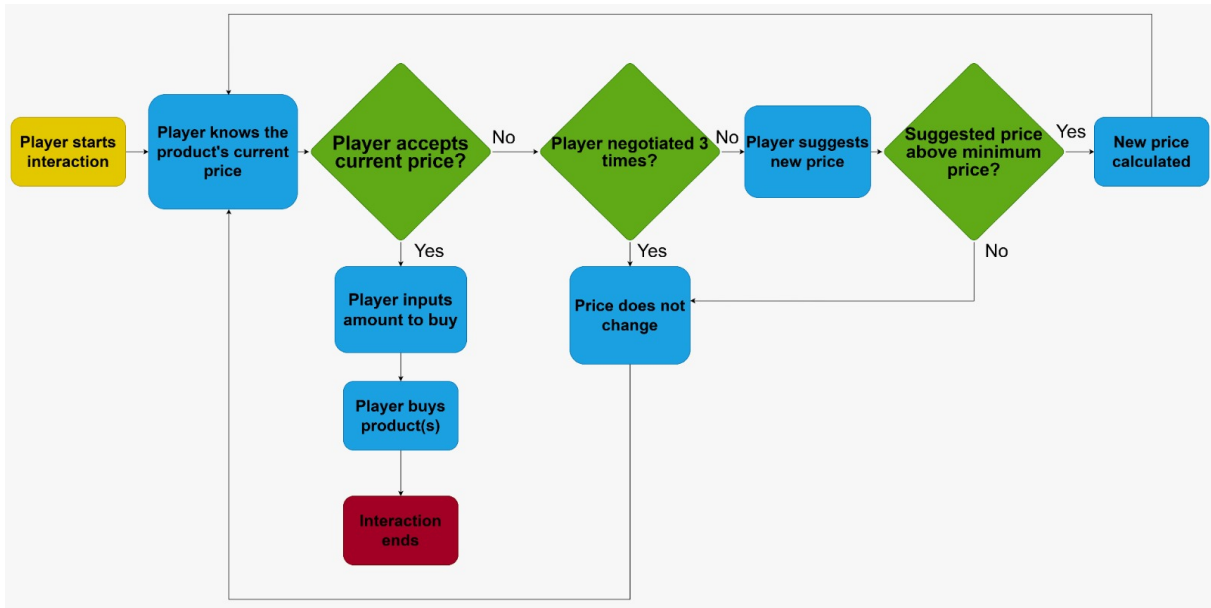


FIGURE 3.16. Flowchart of the dialogue system whenever current mode has LLM OFF.

finishes the bargaining process and chooses the amount he wishes to buy, the interface has only 1 button left which is the "Buy" button.



FIGURE 3.17. Buy Button appearing on the interface to complete the interaction.

Alternatively, if the current game mode has LLM-ON, the interaction makes use of the LLM to occur, following the structure detailed in Section 3.1, Figure 3.2. This interaction interface differs from the previous one since it does not limit the player's choices to buttons but instead lets the player have a natural language dialogue with the NPC.



FIGURE 3.18. Player Interacting with a vendor NPC as the current game mode has LLM ON.

To create the connection detailed in the Section 3.1 between the NPCs and the LLM, starting in the player typing a new message on the text box and clicking on "Submit" and finishing on the LLM generating a textual response based on the player's input, an Unreal Engine plugin named HttpGPT - GPT Integration (ChatGPT and DALL-E)⁵ was used. This plugin provides built-in blueprint functions which allow communication to the OpenAI's API. One of these functions allow us to send the textual module prompt detailed in the Section 3.2.1, filled with the current vendor NPC's information regarding the product, to have it ready for the interaction. Another relevant feature of this plugin that was used in the interactions was keeping the whole conversation in a variable, which was sent to the API, along with the new player's message, so the NPC could keep track of the conversation, appearing as to have memory.

The output format defined on the prompt has the purpose of facilitating the parsing of the information to show to the player. The "Text" component appears on the NPC's dialogue box to simulate the conversation, the "Price" and "Quantity" are used to calculate the total amount of currency that the player is going to spend on the purchase, appearing in the "Current Price:" Section of the interface.

Another relevant feature that is used whenever a mode with LLM-ON is played is text to speech. By making use of a C++ library of OpenAI's API named liboai⁶, a function which makes a text to speech request to the API was implemented to play an audio sound consisting on the LLM's textual response.

3.3.5. NPC's Animation Module

The ability of the vendor NPCs to make use of their animation module is dependent on whether the current game mode has Animations-ON or Animations-OFF. As it was stated

⁵<https://www.fab.com/listings/3edf406f-6a87-4f2f-bfdb-b0039f285541>

⁶<https://github.com/D7EAD/liboai>

in Section 3.1 the NPC can get the animations from the dialogue system, if LLM-OFF is active, or from the LLM's selected non verbal cues that best fit the textual response previously generated when LLM-ON is active. Both cases utilize the same set of facial expressions and body gestures to act as the NPC's non verbal cues, defined in the Section 3.2.2.

The animation module is composed of two state machines, one for the facial expressions and the other for the body gestures. After the module receives the non verbal cues, coming from the LLM or the dialogue system, they activate the state corresponding to them, which reflects upon the NPC's model making it perform the selected animations, by blending the animation that affects the NPC's head with the one affecting its body.

The facial expressions were animated on Unreal Engine, making use of its animation tool. The body gestures, however, were imported to Unreal Engine coming from an external website named Mixamo⁷. Mixamo contains an extensive set of free virtual characters and animations. In order to be able to use those animations in the Metahumans, a virtual character from the website was deployed into the game, along with its skeleton mesh, and the animations that were going to be used on the NPCs were imported. Afterwards they were applied to the Metahuman's skeletal mesh with the use of the retarget animation tool provided by Unreal Engine. All of the animations can be observed in appendix B.

3.3.6. Completing the Interaction

When the player interacts with the "Buy" Button on both interfaces, if the player has enough currency, it adds the amount of the product they bought to the their inventory and subtracts the price of the purchase from the player's current game currency. When this occurs, a pop-up appears to signal the player that the purchase was completed, Figure 3.19, so he can either keep the conversation going, in case he is interacting with the NPCs using the LLM to generate dialogue, or leave the interaction.

However, if the player does not have enough money to complete the purchase a pop up appears informing them that the purchase was not completed, Figure 3.20, since they do not have enough money. When this occurs, the player can submit a new amount of the market good to buy in case they mistakenly chose an higher amount than needed to complete the shopping list.

After a product from the shopping list is bought, it is added to the player's bag and, depending on whether this was the second item from the shopping list being bought, a new shopping list is generated and a new game mode is selected. The entire process can be seen on Figure 3.21. Figure 3.22 shows a new shopping list being generated. Whenever the player has completed a shopping list for each of the game modes, a pop-up informs the player that they have completed the game, as it is shown on Figure 3.23.

⁷<https://www.mixamo.com/>

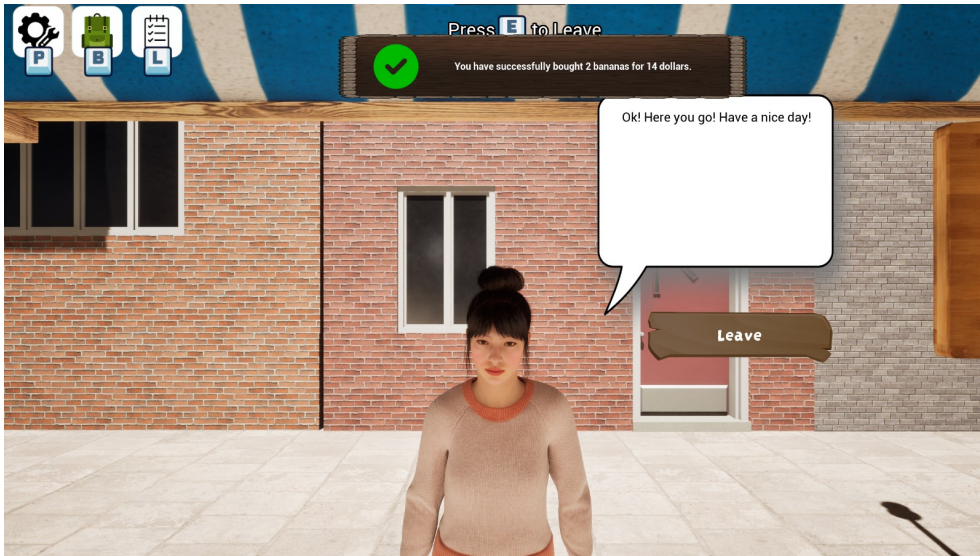


FIGURE 3.19. Pop up signaling the player that the purchase was completed.

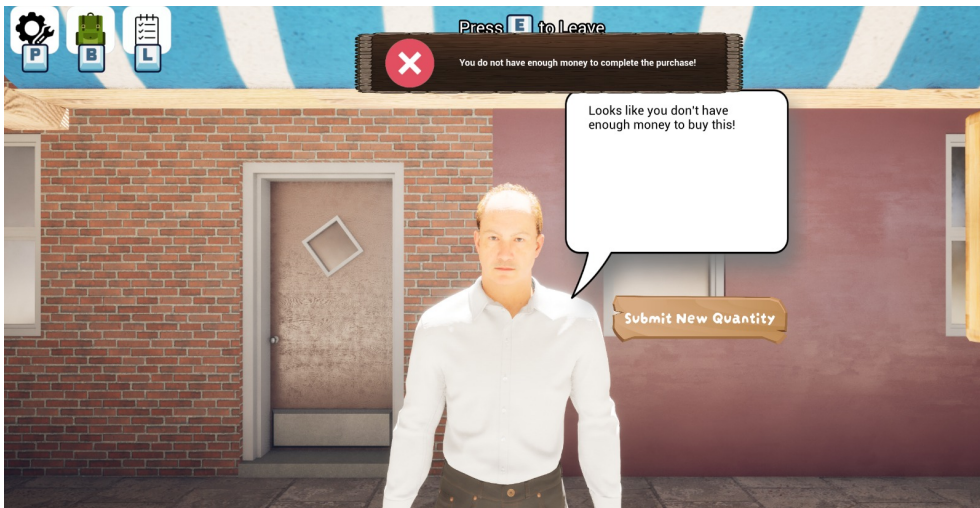


FIGURE 3.20. Pop up signaling the player that the purchase was not completed since the player doesn't have enough money.

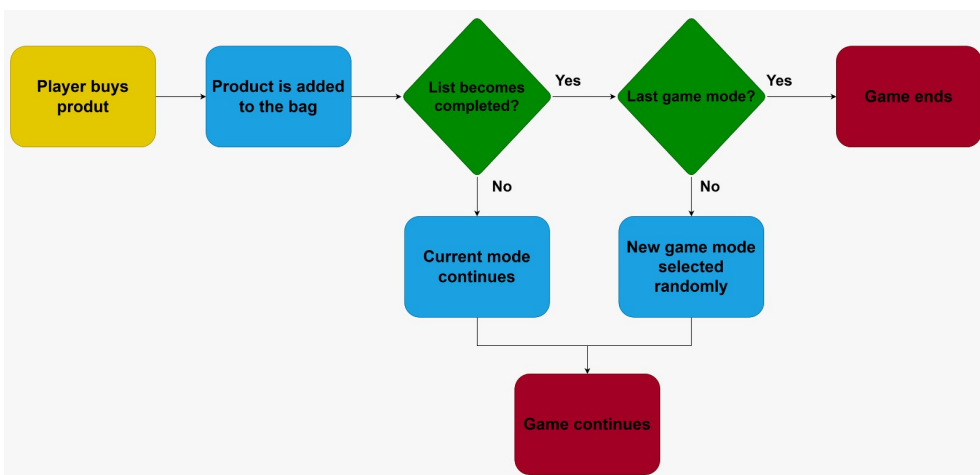


FIGURE 3.21. Flowchart detailing the process of picking a new game mode whenever a shopping list is completed.



FIGURE 3.22. A new shopping list is generated whenever the player buys both items on the shopping list.



FIGURE 3.23. When the player completes the 4th shopping list, a pop up informs him he completed the game.

Evaluation

This chapter presents the evaluation methods taken to test the proposed and developed system in the Section 3.3. Section 4.1 details the method that was used to evaluate the developed game in order to reach conclusions regarding our research questions. This Section details the questionnaires each user had to fulfill during the test along with the statistics retrieved directly from the game during the session. Lastly, Section 4.2 presents the results gotten from the tests, along with a statistical analysis.

4.1. Method

The LLM that integrates the vendor NPCs is GPT 4.0, as stated in the Section 3.1. GPT 4.0’s temperature property, which controls how random the output of the LLM is, was set to 0.5 to allow the LLM to use the prompt’s example as just a guideline and be a bit more creative on its generated responses. Its top-p value, which controls the amount of words that it can choose to use on its output, was kept at the default value of 1.

The tests were conducted in-person in ISCTE, specifically in the study room due to its high concentration of people. The full test procedure, as seen on Figure 4.1, was as followed:

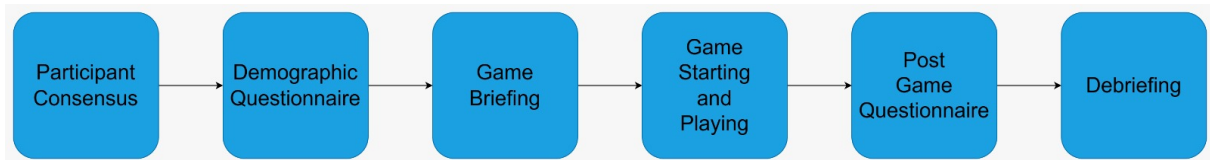


FIGURE 4.1. Flowchart of the conducted procedure.

- **Participant Consensus:** Each participant agreed to initiate the test after being told that it would take around 30 minutes to complete. Following this agreement, each participant received information regarding the objective of the study and the test they were about to made. To formally initiate the test, each participant had to provide an informed consent.
- **Demographic Questionnaire:** This initial set of questions regarded the participant’s gender, age and their experience in playing video games in general and their experience in playing video games based upon negotiation.
- **Game Briefing:** Before playing the game, on the Main Menu presented in Figure 3.4, each participant was briefed that there would be four game modes which

controlled the way they interact with the NPCs without specifying what would change in the interaction. Every time they completed a shopping list, they would need to fulfill a questionnaire regarding the game mode they had just completed. After that a new game mode was selected randomly from the ones they had not play yet until none was left.

- **Game Starting and Playing:** Each participant then played the entire game, completing each shopping list for the four existing game modes, and filling their respective questionnaire.
- **Post Game Questionnaire:** After completing the game, each participant was then asked to complete a survey to evaluate their experience while playing the game.
- **Debriefing:** After completing the survey, each participant was briefed again regarding the objective of the study and that they could choose to leave feedback regarding the experience.

The informed consent and questionnaires that the participant had to fill in this evaluation can be found on Appendix C.

In total, the test was conducted with 30 participants, 25 identified as male students and 5 identified as female students. The ages of the participants ranged from 18 to 26 (AVG=21, STD=2.64, being AVG the average and STD the standard deviation).

Regarding the video game experience of our participant sample group, the Casual Hardcore Assessment Scale (CHG) [15] was utilized to assess whether the sample group was composed of a higher number of casual players or hardcore players. CHG is a 5 item likert scale, ranging from 1 to 5, which include the amount of money the participant has spent on video games, the degree to which the participant personalizes personal items resorting to the video game culture, how serious they take video games, and the amount of time they spend on video games.

In addition, in order to verify whether our participant sample group played games where they could negotiate goods and / or belongings and what type of players (casual or hardcore) were in this specific type of games, in the demographic questionnaire the participants had to fill another CHG scale focused around this type of video games.

The questionnaire regarding each game mode and the survey each participant had to fill in the end of the game were components of the Game Experience Questionnaire (GEQ) [22]. The Game Experience Questionnaire serves to evaluate the player's experience after finishing the game and it is composed of several modules (Core Module, In-Game Module, Social Presence Module and Post Game Module). Each module is composed by various

statement items which the participant rates on a likert scale ranging from 0 to 4 how much each statement represents what they felt during the game session.

The In-Game Module of GEQ was designed to be presented to the player multiple times during the game session to assess its continuous experience. However, in the developed game, the player's experience is affected whenever they interact with the vendor NPCs in different game modes. To evaluate the changes of the player's experience across the four game modes, the Social Presence Module was filled instead by the participant. They filled the questionnaire each time the game mode changed, four times in total, to analyze the effect the game modes had on their experience. The Social Presence Module is made of seventeen items, assessing the player's psychological involvement with the vendor NPCs, both good and bad feelings, along with behavioural involvement. To this module, an eighteenth item was added, which had the participant rate in the likert scale whether the interaction with the NPCs in the game mode made the participant feel as if he was negotiating with a human. Additionally, as the Social Presence Module's questions refer to "other(s)" since it is generalized to all NPCs, "sellers" was used instead so the participants only evaluated their interaction with the vendor NPCs.

The Core Module and the Post Game Module were filled by the participant after playing the entirety of the game as a Post Game Questionnaire.

4.1.1. Game Statistics

Additional to the participant's feedback coming from their answers to the filled questionnaires during and after the game, game statistics were also collected to be analyzed.

In order to analyze how the users responded to each game mode and how it would affect the way they would engage with the vendor NPCs, every time the participant engaged with the interface to take part in the negotiation (i.e., accept, refuse a deal given by the vendor and / or propose a new deal to it), the interaction was recorded in an Excel file to keep track of all negotiations made by all participants. The most important components that were recorded on each interaction depended on whether the game mode had LLM-OFF or LLM-ON. However, the timestamp of when the interaction occurred was recorded no matter what game mode was currently active.

In the game modes that included LLM-OFF, so the interaction with the vendor NPC was made through the buttons interface, every time the participant clicked on a button, their action was recorded. Specifically, the prices they suggested and the products they bought at a given price.

In the game modes that included LLM-ON, so the interaction with the vendor NPC was made through text boxes using natural language, each recorded interaction contained the message sent by the participant and the response sent by the LLM. In this case, the interaction was recorded on the Excel file whenever the LLM's answer appeared on the interface, being that timestamp the one recorded. Additional to the interactions themselves, the price and amount of the products they bought were also recorded.

The Excel file acted as a database retaining all the interactions made when participants negotiated with the vendor NPCs. From these data, significant variables were selected to create metrics to compare between the game modes to infer conclusions of the game session. The metrics that were chosen to analyze were the following:

- **Time Negotiating:** By subtracting the timestamp of the first interaction to the timestamp of the last interaction, the result is the time the participant spent negotiating with the vendor NPC:

$$\text{Time Negotiating} = \text{Timestamp End} - \text{Timestamp Begin.}$$

- **Number of Interactions:** This metric tracks the amount of times the participant interacted with the vendor NPC to negotiate the price of the products they were buying. Each interaction corresponds to the message sent by the participant and the NPC's response.
- **Difference in Price:** This metric measures, in percentage, the difference between the initial price of a given product and the price at which it was bought. Since every product has a difference initial price and minimum price that it could be bought for, by subtracting the minimum price of the product to the initial price, the price difference across all products is standardized:

$$\text{Difference in Price} = \frac{\text{Product's Initial Price} - \text{Product's Bought Price}}{\text{Product's Initial Price} - \text{Product's Minimum Price}}.$$

Since every shopping list is composed by 2 different products that the participant has to buy, in each game mode these metrics were calculated for each product and then the mean of the two were used to represent that game mode for each participant. That way a comparison between the game modes can be done without comparing each product individually.

4.2. Results

4.2.1. Demographic Questionnaire Results

The results obtained on the first CHG questionnaire, presented in Table 4.1, which participants filled regarding their general video game experience, indicate that the sample group does not lean towards neither a more casual player sample group neither a more hardcore one. As the scale ranges from 1 (casual player) to 5 (hardcore player), the total mean result of 2.49 indicates an even distribution between the both types of players, which was expected since the participants were chosen randomly in ISCTE's study room.

TABLE 4.1. Mean and Standard Deviation ($M \pm STDEV$) for each item of CHG in video games in general.

Item	$M \pm STDEV$	N
Over the last year, I consider that I have invested large sums of money in devices, gadgets, or other purchases dedicated to video games	2.13 ± 1.11	30
I tend to personalize everything with video game-specific elements	2.80 ± 1.30	30
I take video games very seriously	2.73 ± 1.26	30
I spend a lot of time in video games on a day off / weekend	3.00 ± 1.23	30
I spend a lot of time in video games on a work / school day	1.80 ± 0.85	30
Total	2.49 ± 1.23	30

After having the participants fill their video game experience in general, it was crucial to know whether they played games where they could negotiate products, since our proposed solution aims to improve the player’s experience when playing this type of games. Only 1 of the 30 participants answered that they have never played a game of this genre, as can be seen on Table 4.2. The following questions were only made to the other 29 participants which answered that they had played a game of this genre before.

TABLE 4.2. Number of People that have played games where they could negotiate goods and / or belongings.

Answer	N	%
Yes	29	96.7%
No	1	3.3%
Total	30	100%

Table 4.3 shows the CHG mean is lower in the second run, which is focused on negotiation games. With an average of 2.09, players show to be more casual in this subgenre. The decrease can be verified for a few reasons. Firstly, the values of monetary investments on video games and personalization decreased, meaning that, for this participant group, this specific genre is not a priority when it comes to spending money on video games. Also, for this participant group, this genre is not played as much as other video game genres since the values for spending time on video games also decreased, on questions four and five. An interesting observation is that, overall, the participants take this genre of video games more seriously, since there was a slight increase. So, even though they do not spend as much time and money on this video game genre, when they play it they are invested in it. This can be justified by the negotiation interactions, which are intrinsically competitive, so it requires the players to be more strategic in order to reach a favorable deal, taking the game more seriously.

TABLE 4.3. Mean and Standard Deviation for each item of CHG in games where the participants could negotiate goods and / or belongings.

Item	M \pm STDEV	N
Over the last year, I consider that I have invested large sums of money in devices, gadgets, or other purchases dedicated to video games	1.48 \pm 0.91	29
I tend to personalize everything with video game-specific elements	1.97 \pm 1.24	29
I take video games very seriously	3.07 \pm 1.36	29
I spend a lot of time in video games on a day off / weekend	2.41 \pm 1.12	29
I spend a lot of time in video games on a work / school day	1.52 \pm 0.78	29
Total	2.09 \pm 1.24	29

The participants were then questioned regarding the composition of their playing group when they played this type of video games. As can be seen on Table 4.4, the majority of the participants (corresponding to 65.5% of the sample group) only played this type of games with other human players, and only 10 participants (corresponding to 34.5% of the sample group) have played this type of games with AI controlled players. From this smaller group, 9 participants have played with diversified playing groups, composed of both human and AI controlled characters, and only 1 participant has played this type of games with a group composed of only AI controlled players. From these results, it can be noticed that the participants are more likely to play with other human players rather than playing with AI controlled players. As stated in the first chapter, when introducing this topic, players do not feel as involved in a negotiation focused game when they have to play with AI controlled players since they do not interact with them as freely as they do with other human players.

TABLE 4.4. Composition of the playing group when participants played games where they could negotiate goods and / or belongings.

Composition	N	%
Only Human Controlled Characters	19	65.5%
70% Human Controlled Characters 30% AI Controlled Characters	2	6.9%
50% Human Controlled Characters 50% AI Controlled Characters	2	6.9%
30% Human Controlled Characters 70% AI Controlled Characters	5	17.2%
Only AI Controlled Characters	1	3.4%
Total	29	100%

As a last demographic question regarding our participant group's video game experience when playing games focused on negotiating goods, they were asked what games they had played that fit in this specific genre, their answers are shown on Figure 4.2. The 29

participants that have played this video game genre referenced "Monopoly", this was expected since it is the most well-known game that incites players to negotiate with the other players to trade properties in order to reach the game's goal. The second most answered game was "Catan" which is another well-known game where players have to strategize to use their resources in an optimal way and they can negotiate resources with other players to also reach the game's goal.

When responding to this question, participants often answered games where negotiation was not the main focus, but rather a possibility that was viable since the game had an online community where items could be exchanged for either other items or, some times, currency, this was the case for "Minecraft", "Counter Strike", "Rocket League" and "Metin2".

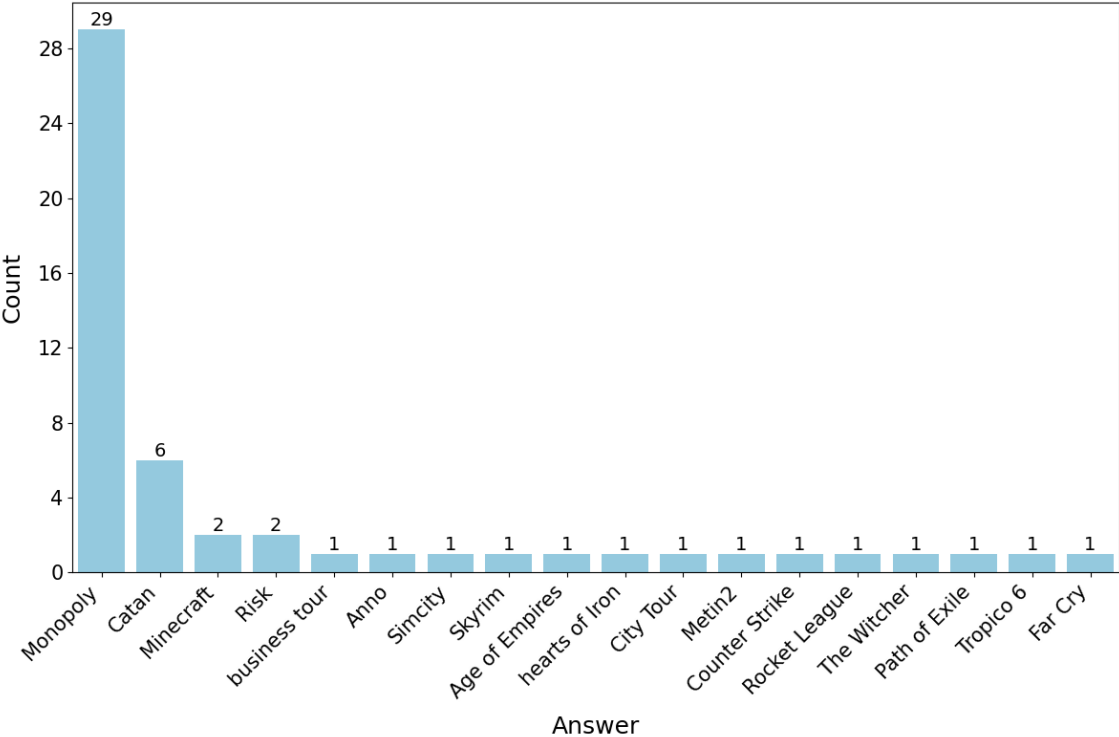


FIGURE 4.2. Games focused on negotiation that the participants have played.

4.2.2. Game Questionnaire Results

4.2.2.1. Social Presence Module

As stated in the previous Section, the questionnaire that the participants had to fill four times during the game session included the social presence module of GEQ. Initially, as a general comparison between the four different game modes, Table 4.5 provides the mean and standard deviation of the filled questionnaires for each social presence module component in each game mode.

TABLE 4.5. Mean and Standard Deviation for the GEQ Social Presence Module components per game mode condition.

Component	Animations-OFF LLM-OFF (M \pm STDEV)	Animations-OFF LLM-ON (M \pm STDEV)	Animations-ON LLM-OFF (M \pm STDEV)	Animations-ON LLM-ON (M \pm STDEV)
Psychological Involvement – Empathy	1.35 \pm 1.32	2.13 \pm 1.40	1.63 \pm 1.44	2.42 \pm 1.17
Psychological Involvement – Negative Feelings	1.02 \pm 1.35	1.25 \pm 1.42	1.25 \pm 1.49	1.49 \pm 1.45
Behavioural Involvement	2.39 \pm 1.30	2.87 \pm 1.01	2.52 \pm 1.21	3.02 \pm 1.00

As can be observed from the Table 4.6, some prior results can be noted before analyzing each individual question across the four game modes.

From the first component, which evaluated the empathy felt by the players towards the vendor NPCs and the general feeling of a connection between the 2 parties, the mean value of the game modes that included LLM-OFF (with Animations-OFF the mean value was 1.35 and with Animations-ON the mean value was 1.63), so the interaction occurred via the buttons interface, was lower than when it included LLM-ON (with Animations-OFF the mean value was 2.13 and with Animations-ON the mean value was 2.42), which had the interaction occur in natural language. This means that the participants felt more connected and enjoyed the interaction better when they were interacting in natural language. Another important observation is that the existence of non verbal communication increased the empathy felt by the participants towards the vendor NPCs. However, the difference is more significant when comparing LLM-OFF mean values with LLM-ON mean values. This means that, as an initial and less detailed analysis of this component, participants' empathy towards the vendor NPCs depend more on whether they can communicate with natural language than whether the NPCs are performing non verbal communication.

The second component evaluates the opposite, the overall negative feelings (like guilt, revenge and malicious delight) felt by the participants towards the NPCs, and as stated, the mean values are the highest for Animations-ON and LLM-ON (with a mean value of 1.49) and the lowest for Animations-OFF and LLM-OFF (with a mean value of 1.02), the game modes which had a mix combination of LLMs and animation had the same mean value of 1.25. From these results it can be inferred that, since the connection between the participants and the vendor NPCs is deeper with natural dialogue, as was seen on the previous component, they are more prone to experience negative feelings towards them rather than when the interaction is made with buttons. This means that, along with providing a positive feeling of empathy and connection between the interaction parties, the use of LLM can also increase the existence of negative feelings between them.

The last component evaluates the degree of the involvement in the interaction of the participants with the vendor NPCs, and the results are similar to the results of the first component. Once again, the mean values for LLM-OFF (with Animations-OFF the mean value was 2.39 and with Animations-ON the mean value was 2.52) were lower than

LLM-ON (with Animations-OFF the mean value was 2.87 and with Animations-ON the mean value was 3.02). The participants felt, overall, that they were more involved in the interaction, so they paid more attention to what was happening and felt what they done impacted more the way the interaction was going, whenever the interaction was made using the LLM, through natural dialogue. Also, just like the first component, the differences are greater when LLM-OFF with LLM-ON mean values are compared. Animation-wise the difference is not significant when the mean values for a fixed LLM mode value are compared, this reinforces the idea that the presence of LLMs is much more significant than the presence of non verbal cues.

4.2.2.2. *Game Questionnaire*

After this initial analysis of GEQ's social presence module in the game questionnaire, a more detailed analysis of the total questionnaire that the participants had to fill was performed. The goal of this detailed analysis was to verify which game mode was superior to the other game modes in the questionnaire to conclude whether the usage of a LLM and non verbal cues would improve the player experience in the game.

Figure 4.3 provides, for each question, the distribution of the answers gave by the participants on each game mode.

In order to determine whether a game mode was significantly different from the other three on each question, initially a Shapiro-Wilk test [51] was performed, which is adequate for $n < 50$ sample sizes (in this case $n = 30$), across the eighteen questions to assess if the data distribution of the game modes was normal. The results of this test showed that the data on each question was not normally distributed, since the p -value for each game mode was always inferior to the normality threshold, which was 0.05. To confirm these results, Figure 4.4 contains Q-Q plots [8] to visually demonstrate the non-normal distribution showing the values of the four game modes, marked as blue, of each question compared to the theoretical normal distribution of the results in the red line. As the real values do not align on the red line, the data is not normally-distributed, and a non parametric test was necessary to further analyze this data.

In order to choose a test that would determine whether a condition was significantly different from the others, in each question, two factors were taken to account. The data was not normally distributed, which was concluded with the Shapiro-Wilk test and the following Q-Q plots, and the other factor was that since each participant answered to the four conditions on all questions, the data was also based on repeated measures. The chosen test was Friedman's test [36] which is a non-parametric test for repeated measures which is adequate for three or more groups (in this case there are four).

The results of the Friedman's test are presented in Table 4.6. For each question there is a χ^2 value assigned which determines the difference of the ranked sums of the groups. This means that the greater the χ^2 value the more the data in the question deviates from the null hypothesis which is that there is no significant difference between the groups in the

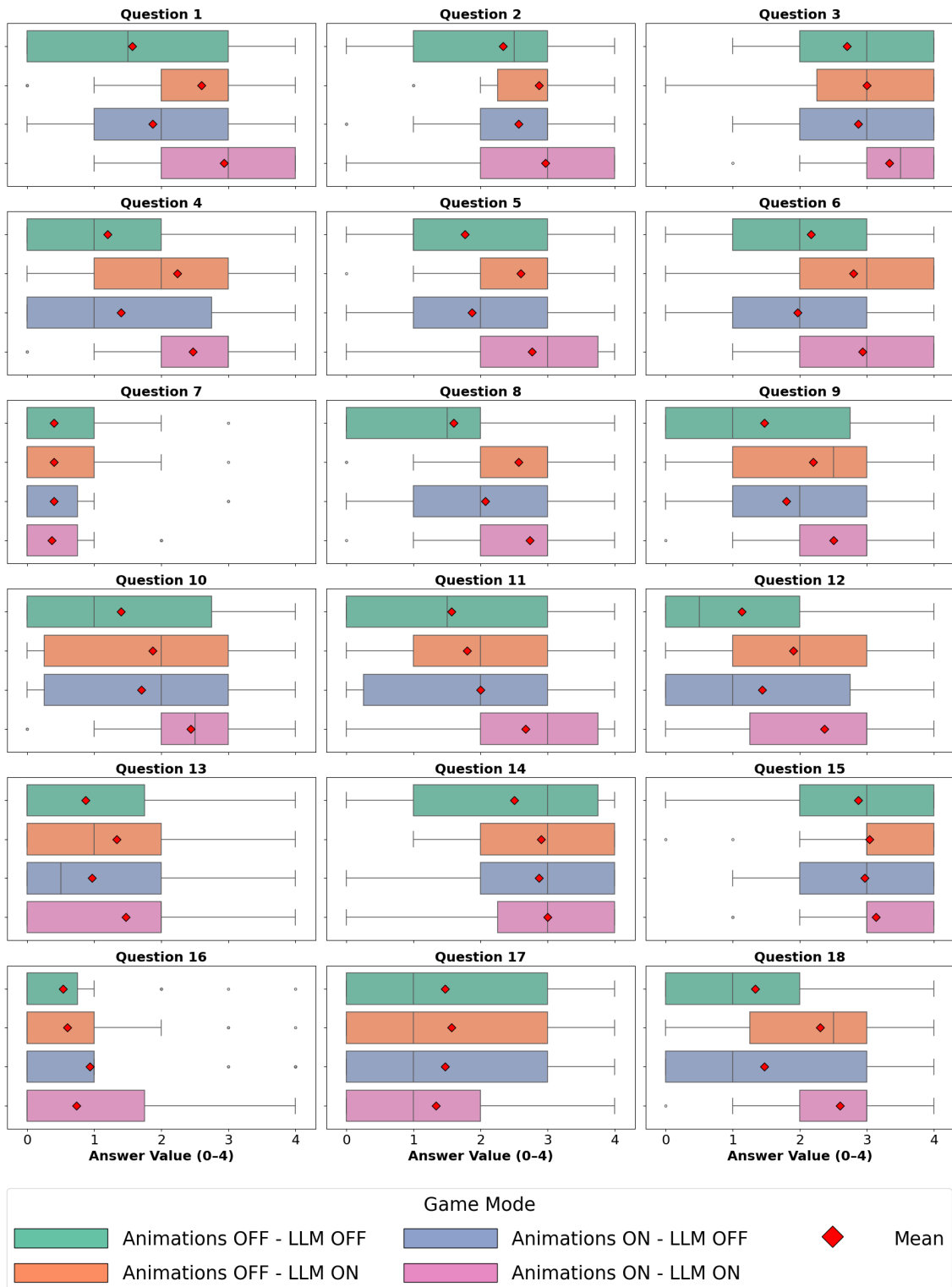


FIGURE 4.3. Answer distribution per game mode on each question.

question. The null hypothesis is rejected whenever the p -value is less than 0.05, the chosen significance value. The questions in which the rejection occurs, and therefore present a statistical difference between the game mode groups, are marked with a $<^*>$ on their p -value. From the Friedman's test results, fourteen out of the eighteen total questions (77.78%) present one group that is statistically different from the other three groups. For

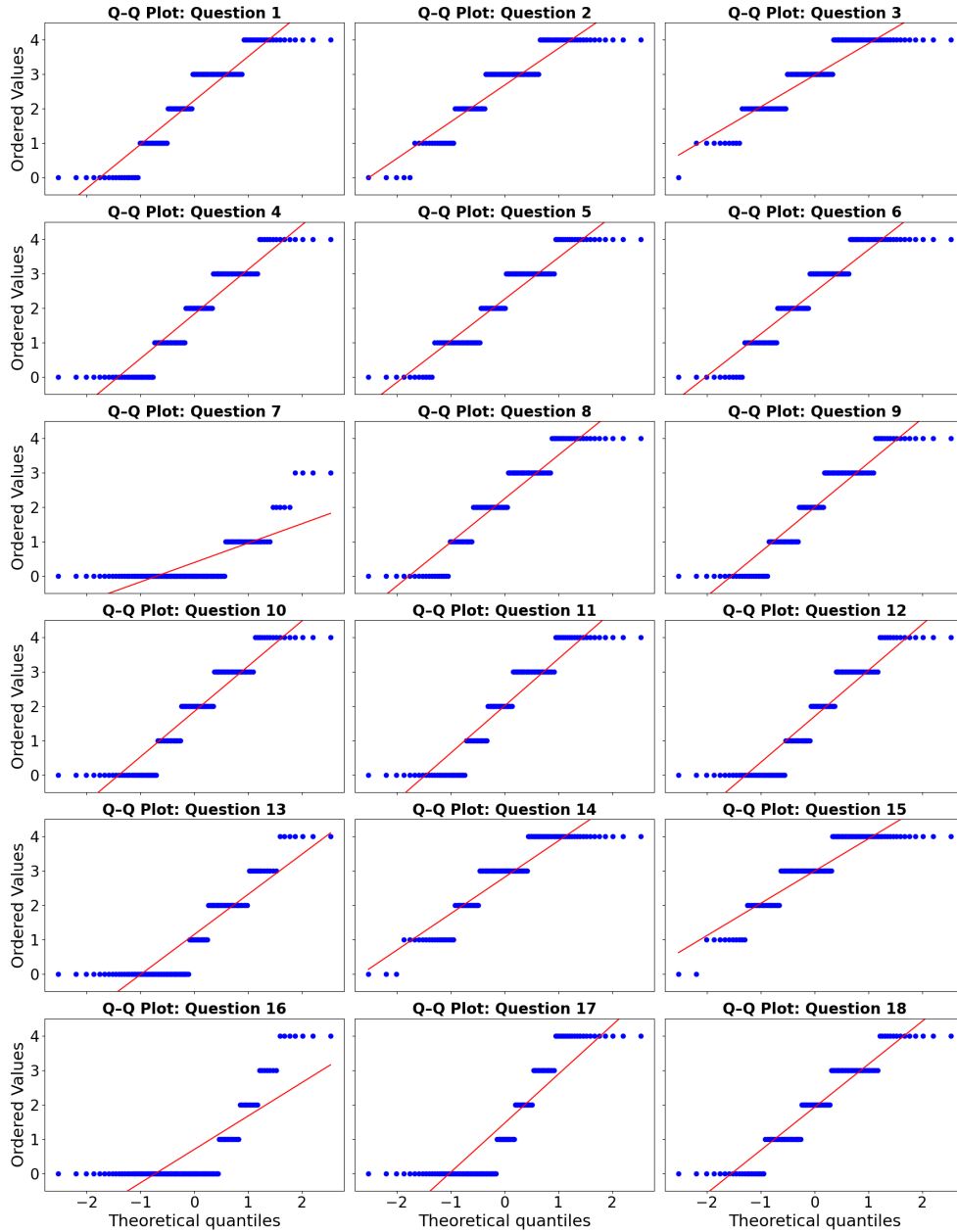


FIGURE 4.4. Q-Q Plots to show answer distribution for each question.

TABLE 4.6. Friedman's test χ^2 and p -value results for each question in the questionnaire.

Question	χ^2	p -value
1	29.57	< 0.005*
2	10.83	0.01*
3	8.33	0.04*
4	31.64	< 0.005*
5	29.96	< 0.005*
6	12.81	0.01*
7	0.20	0.98
8	19.80	< 0.005*
9	21.71	< 0.005*

Question	χ^2	p -value
10	16.87	< 0.005*
11	16.99	< 0.005*
12	26.23	< 0.005*
13	13.57	< 0.005*
14	8.69	0.03*
15	2.86	0.41
16	4.22	0.24
17	1.39	0.71
18	28.37	< 0.005*

these questions, in order to confirm which of the groups was significantly different from the others, a Wilcoxon signed rank test [48] for paired samples was performed with a Bonferroni correction [32]. In this analysis, the null hypothesis is, as said previously, that there is no significant difference between the two compared groups, across the six comparisons for each question. When a comparison is made there is a chance, named alpha value (in this case $\alpha = 0.05$), for it to reject the null hypothesis incorrectly, which is a Type 1 error [9]. Since six comparisons are being made, it is more likely for a Type 1 error to occur, which is why the Bonferroni correction [43] is important. It reduces the probability of encountering a Type 1 error on each comparison by dividing the alpha value by the number of comparisons being made. So in each question, in the set of six comparisons, each individual one has an alpha value of 0.008. The results of these pair-wise comparisons for the questions that indicated a statistically different group are presented in Tables 4.7. Each column is the comparison between the first group, on top, and the second group, on the bottom. The comparisons (A stands for Animations and L stands for LLM) which present a p -value specified on a question, also marked with $\langle * \rangle$, demonstrated a statistical difference and rejected the null hypothesis (p -value < 0.05). All of the results from the performed test can be found in Table D.1.

TABLE 4.7. Wilcoxon signed rank test with Bonferroni correction’s p -value for comparisons on questions 1-18 that rejected the null hypothesis.

Question	A-OFF_L-OFF A-OFF_L-ON	A-OFF_L-OFF A-ON_L-OFF	A-OFF_L-OFF A-ON_L-ON	A-OFF_L-ON A-ON_L-OFF	A-OFF_L-ON A-ON_L-ON	A-ON_L-OFF A-ON_L-ON
1	$< 0.005^*$	-	$< 0.005^*$	-	-	$< 0.005^*$
2	-	-	-	-	-	-
3	-	-	-	-	-	-
4	$< 0.005^*$	-	$< 0.005^*$	0.03^*	-	$< 0.005^*$
5	0.01^*	-	$< 0.005^*$	0.01^*	-	$< 0.005^*$
6	-	-	0.04^*	0.04^*	-	0.01^*
8	0.01^*	-	$< 0.005^*$	-	-	-
9	$< 0.005^*$	-	0.01^*	-	-	0.02^*
10	-	-	0.01^*	-	-	0.04^*
11	-	-	0.01^*	-	0.02^*	-
12	$< 0.005^*$	-	$< 0.005^*$	-	-	0.02^*
13	0.04^*	-	0.03^*	-	-	-
14	-	-	-	-	-	-
18	$< 0.005^*$	-	$< 0.005^*$	0.04^*	-	$< 0.005^*$

From the questions that rejected the null hypothesis in Friedman’s test (7, 15, 16 and 17 not included) , a few of them (2, 3 and 14) did not present significant differences across the six comparisons made. The questions that did present significantly different comparisons, indicated only two to four significant differences. The lack of significant comparisons across the questions prevented a direct conclusion to be taken from these results regarding which game mode was significantly different from the others.

Given the fact that the results obtained from analyzing the four groups were not conclusive, they were divided into two separate datasets. One of the datasets compared having LLM-ON and LLM-OFF making the value for each one the mean of the Animations-ON and Animations-OFF values for each participant. The other sub-group went through

the same process, however, it compared having Animations-ON and Animations-OFF. This would make the four different responses given by each participant to be two in each of the datasets. This process was made based on the conclusion that, since the game modes had two features (LLM and animations), and the results were not conclusive, it would mean that the individual impact of the features was different. Figure 4.5 presents the distribution, for each question, of the LLM-ON and LLM-OFF values. Figure 4.6 presents the distribution, for each question, of the Animations-ON and Animations-OFF values.



FIGURE 4.5. Answer distribution per question for LLM OFF and LLM ON.

FIGURE 4.6. Answer distribution per question for Animations OFF and Animations ON.

As the data remained as non-parametric and it was still repeated measures, the Wilcoxon signed rank test was chosen to perform the comparisons for each question, however, the Bonferroni correction was not applied in these comparisons since there was only one comparison in each question. Tables 4.8 and 4.9 present the resulting mean and standard deviation for both conditions, the difference in the means of both conditions and

the p -value. Once again, the questions where a significant difference was observed, so the p -value is less than the significance threshold which is 0.05, are marked with $<^*>$. The difference in the means of both conditions stated which condition's mean has a higher value. It is positive if LLM-ON's mean is higher than LLM-OFF's mean and vice-versa for negative mean differences.

TABLE 4.8. Mean and Standard Deviation, the mean difference (M_Diff) and p _value for each question to compare between LLM ON and OFF.

Question	LLM-ON (M \pm STDEV)	LLM-OFF (M \pm STDEV)	M_Diff (ON - OFF)	p -value
1	2.77 \pm 0.94	1.72 \pm 1.23	1.05	$< 0.005^*$
2	2.92 \pm 0.88	2.45 \pm 1.09	0.47	0.02*
3	3.17 \pm 0.70	2.78 \pm 0.87	0.39	0.08
4	2.35 \pm 1.04	1.30 \pm 1.21	1.05	$< 0.005^*$
5	2.68 \pm 0.98	1.82 \pm 0.87	0.87	$< 0.005^*$
6	2.87 \pm 0.90	2.07 \pm 1.24	0.80	$< 0.005^*$
7	0.38 \pm 0.67	0.40 \pm 0.64	-0.02	0.88
8	2.65 \pm 0.96	1.83 \pm 1.26	0.82	$< 0.005^*$
9	2.35 \pm 1.01	1.63 \pm 1.11	0.72	$< 0.005^*$
10	2.15 \pm 1.07	1.55 \pm 1.23	0.60	$< 0.005^*$
11	2.23 \pm 1.10	1.78 \pm 1.32	0.45	0.01*
12	2.13 \pm 1.11	1.28 \pm 1.23	0.85	$< 0.005^*$
13	1.40 \pm 1.25	0.92 \pm 1.08	0.48	$< 0.005^*$
14	2.95 \pm 0.86	2.68 \pm 1.13	0.27	0.04*
15	3.08 \pm 0.71	2.92 \pm 1.03	0.16	0.40
16	0.67 \pm 0.87	0.73 \pm 1.09	-0.06	0.95
17	1.45 \pm 1.46	1.47 \pm 1.53	-0.02	0.60
18	2.45 \pm 1.04	1.40 \pm 1.08	1.05	$< 0.005^*$

From these two tables (Table 4.8 and 4.9) a comparison can be made to understand which interaction feature was more impactful on the participant's experience.

On Table 4.8, thirteen out of the eighteen questions, 72.22% of the questions, had significant differences between having the LLM-ON and LLM-OFF. All of those questions had a higher mean for LLM-ON than LLM-OFF, ranging from 0.27, (on question 14) to 1.05 (on questions 1, 4 and 18). LLM's presence made an impact on the entirety of the "psychological involvement - empathy" component of the social presence module (questions 1, 4, 8, 9, 10 and 13), on two questions of the "psychological involvement - negative feelings" component (questions 11 and 12), most of the questions of "behavioural involvement" component (questions 2, 5, 6 and 14) and lastly on the added eighteenth item "It seemed I was negotiating with a human". This means that, when participants interacted with the vendor NPCs through natural dialogue, they felt more empathy towards the vendors than when they interacted through the dialogue system. The two questions of the "psychological involvement - negative feelings" component that had a significant difference between LLM-ON and LLM-OFF were "I influenced the mood of the sellers" and "I was influenced

TABLE 4.9. Mean and Standard Deviation, the mean difference (M_Diff) and p_value for each question to compare between Animations ON and OFF.

Question	Animations-ON (M \pm STDEV)	Animations-OFF (M \pm STDEV)	M_Diff (ON - OFF)	p-value
1	2.40 \pm 1.00	2.08 \pm 1.09	0.32	0.05
2	2.77 \pm 0.92	2.60 \pm 0.90	0.17	0.09
3	3.10 \pm 0.56	2.85 \pm 0.68	0.25	0.07
4	1.93 \pm 0.97	1.72 \pm 1.17	0.21	0.26
5	2.32 \pm 0.98	2.18 \pm 1.11	0.14	0.33
6	2.45 \pm 0.91	2.48 \pm 1.04	-0.03	0.79
7	0.38 \pm 0.60	0.40 \pm 0.62	-0.02	0.96
8	2.40 \pm 0.90	2.08 \pm 1.08	0.32	0.03*
9	2.15 \pm 1.05	1.83 \pm 1.23	0.32	0.13
10	2.07 \pm 1.07	1.63 \pm 1.17	0.44	0.01*
11	2.33 \pm 1.21	1.68 \pm 1.21	0.65	< 0.005*
12	1.90 \pm 1.10	1.52 \pm 1.27	0.38	0.12
13	1.22 \pm 1.09	1.10 \pm 1.23	0.12	0.32
14	2.93 \pm 1.01	2.70 \pm 1.05	0.23	0.13
15	3.05 \pm 0.75	2.95 \pm 0.83	0.10	0.61
16	0.83 \pm 1.18	0.57 \pm 0.92	0.26	0.33
17	1.40 \pm 1.37	1.52 \pm 1.51	-0.12	0.27
18	2.03 \pm 1.00	1.82 \pm 0.97	0.21	0.22

by the mood of the sellers". This means that the participants felt that they and the vendor NPCs were more vulnerable throughout the natural dialogue interaction. Participants were more involved on the interaction when communicating with natural language since the questions "My actions depended on the seller's actions", "The sellers paid close attention to me", "I paid close attention to the sellers" and "What the sellers did affected what I did" were impacted by the presence of the LLM. Lastly, the realism of the interaction was also improved by the presence of the LLM given the significant difference (< 0.005*) between LLM-ON and LLM-OFF.

On Table 4.9, three out of the eighteen questions, 16.67% of the questions, had significant differences between having Animations-ON and Animations-OFF. On those questions, Animations-ON had higher mean values than Animations-OFF, ranging from 0.32 (question 8) to 0.65 (question 11). Specifically, the questions which participants felt that having animations made an impact on their experience were "I found it enjoyable to be with the sellers", "When the sellers were happy, I was happy" and "I influenced the mood of the sellers". This means that the animations' impact was really limited compared to the LLM's impact on the experience on the player given the low amount of questions that the presence of the animations made a statistically difference.

From the combination of these results, some conclusions can be taken regarding the presence of the features regarding the questionnaire. Firstly, by dividing the initial dataset which contains the four groups in two datasets to understand the importance of each

feature individually yielded clearer results than when comparing combinations of the features. Secondly, the difference of having the negotiation occur through buttons (with LLM-OFF) and having it occur through natural language (with LL-ON) is significant given the results in Table 4.8 with the following analysis. The same, however, can not be said regarding the difference of having the vendor NPCs perform non verbal cues (with Animations-ON) and having them standing still (with Animations-OFF), given the results in Table 4.9 and the following analysis.

An hypothesis for this lack of significance regarding the presence of the non verbal cues could be an unintentional blindness by the participants. Across the four game modes, the interface through which the negotiation occurs changes whether the game mode includes LLM-ON or LLM-OFF, and the difference is remarkable since the participant either negotiates by clicking on buttons or by typing their dialogue. This difference in the interface captivates the participant's attention more, since they engage directly with the interface, than whether the vendor NPC is performing movements or standing still, which could possibly be unnoticed by some participants.

4.2.3. Game Statistics Results

After the analysis of the questionnaire filled by the participants to compare each game mode based on their interaction with the vendor NPCs, a statistical analysis to the game statistics specified in Section 4.1.1 was performed. Figure 4.7 presents the data distribution across the four groups in each game statistic.

The procedure for the statistical analysis was the same performed for the questionnaire that the participants had to fill during the game. Initially, a Shapiro-Wilk test was performed on the four groups across the game statistics to verify if the data followed a normal distribution. The results of this test can be seen on Figure 4.8 as the data is distributed on Q-Q plots to visually acknowledge their normality. As the game statistics "Time Negotiated" and "Difference in Prices" had one group where the data did not follow a normal distribution, a non parametric test was used to verify whether there was a significantly different group in each one of them. The same test was applied to the game statistic "Time Negotiating" as the data in the four groups did not follow a normal distribution.

The non parametric test used to analyzed the differences across the groups in each game statistic was the Friedman's test as the data in the game statistics was also repeated measures considering each participant played the four game modes. The results for this test can be seen in Table 4.10. The null hypothesis was the same that was used to analyze the questionnaire which was that there is no significant difference between the four groups in each game statistic.

From this Friedman's test, two out of the three game statistics presented a p -value inferior to the threshold of significance, 0.05, which meant that there was a game mode that was statistically different from the other three. In order to confirm which game mode was statistically different, the Wilcoxon signed rank test for paired samples with a Bonferroni

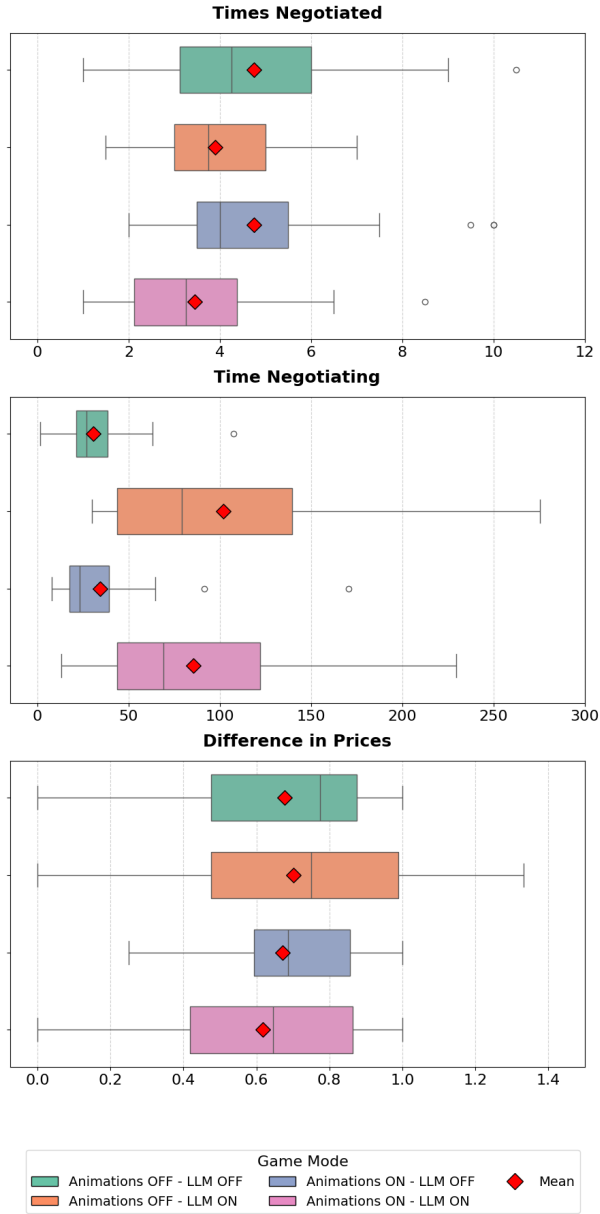


FIGURE 4.7. Answer distribution per game statistic for each game mode.

TABLE 4.10. Friedman’s test χ^2 and p -value results for each game statistic.

Game Statistic	χ^2	p -value
Times Negotiated	12.54	0.01*
Time Negotiating	68.12	< 0.005*
Difference in Prices	1.64	0.65

correction was used once more with the same null hypothesis. The results of the pair-wise comparisons across the two game statistics ,which presented a statistically different group, that rejected the null hypothesis can be seen in the Tables 4.11. Each column is the comparison between the first group, on top, and the second group, on the bottom. The comparisons (A stands for Animations and L stands for LLM) which present a p-value specified on a question, also marked with <*>, demonstrated a statistical difference and

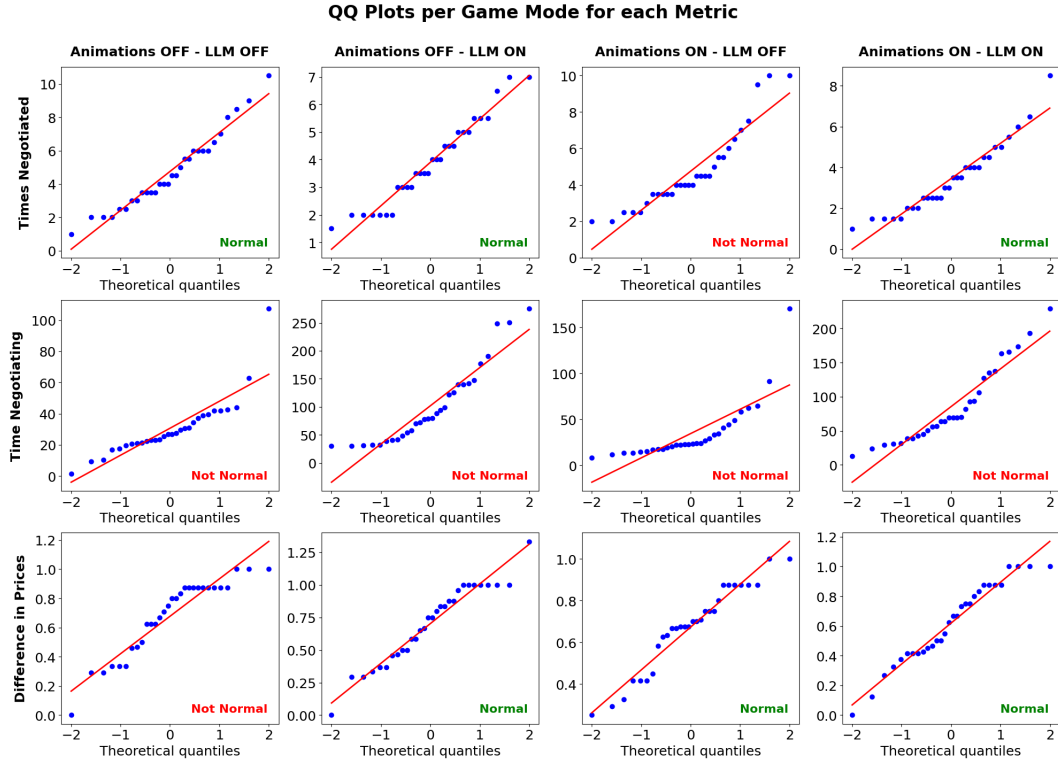


FIGURE 4.8. Q-Q Plots to show the data distribution across the game modes in each game statistic.

rejected the null hypothesis ($p\text{-value} < 0.05$). All of the results from the performed test can be found in Table D.2.

TABLE 4.11. Wilcoxon signed rank test with Bonferroni correction's p -value for the comparisons that rejected the null hypothesis on the two game statistics.

Game Statistic	A-OFF_L-OFF A-OFF_L-ON	A-OFF_L-OFF A-ON_L-OFF	A-OFF_L-OFF A-ON_L-ON	A-OFF_L-ON A-ON_L-OFF	A-OFF_L-ON A-ON_L-ON	A-ON_L-OFF A-ON_L-ON
Times Negotiated	-	-	0.05*	0.04*	-	0.03*
Time Negotiating	< 0.005*	-	< 0.005*	< 0.005*	-	< 0.005*

The obtained results of these pair-wise comparisons were similar to the ones obtained when the pair-wise comparison was performed for the questionnaire, which provides an unclear conclusion regarding which group was significantly different from the other three on each game statistic.

As the encountered result was that there was not a clearly distinct group when comparing the combination of the interaction features, the separation into two datasets detailed on the questionnaire analysis in the previous section was performed once more to compare the impact of each feature individually across the game statistics. The resulting data distribution of each dataset across the three game statistics can be observed in Figures 4.9 and 4.10.

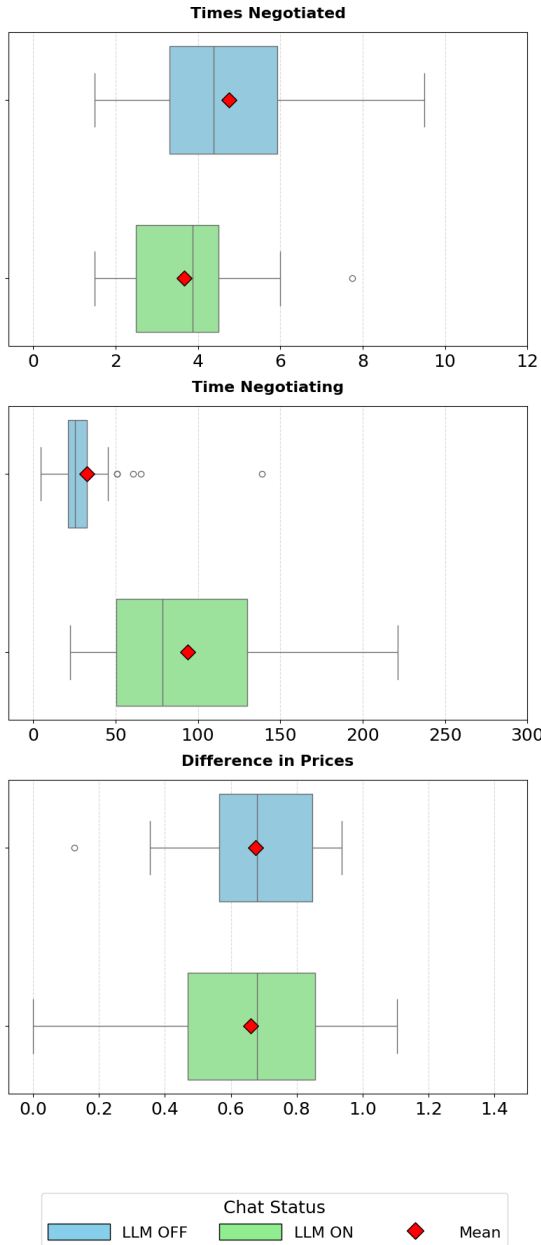


FIGURE 4.9. Answer distribution per game statistic to compare between LLM OFF and LLM ON.

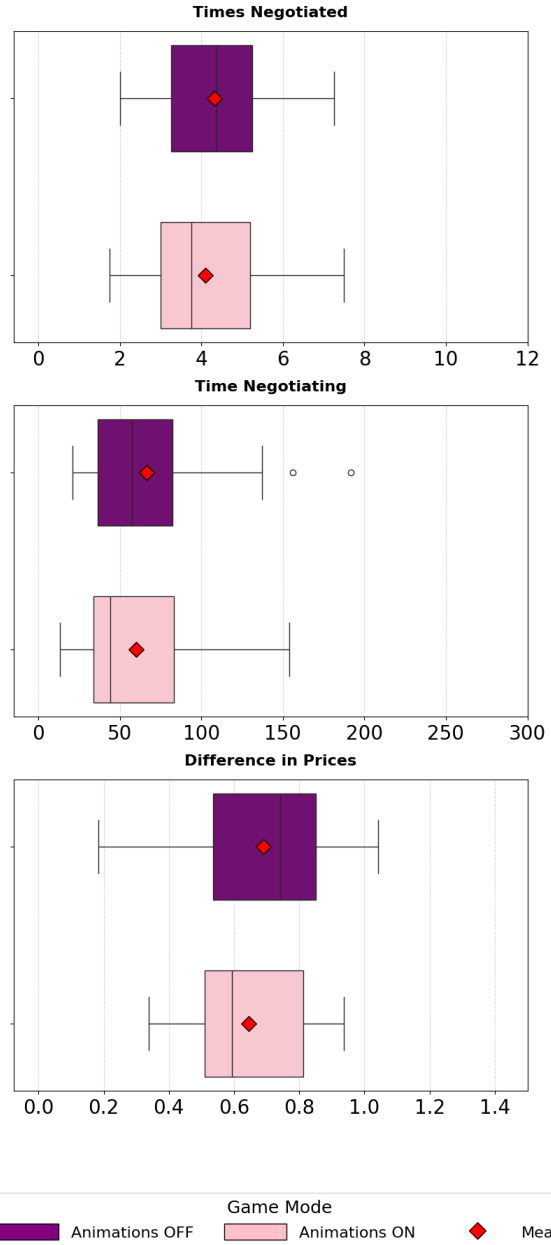


FIGURE 4.10. Answer distribution per game statistic to compare between Animations-OFF and Animations-ON.

With each dataset, the Wilcoxon signed rank test for pair-wise comparisons was performed to understand how the existence of each individual feature impacted the data obtained on each game statistic. The results of the performed test, which are the direct comparison of the means and whether the difference in the existence of the feature impacted the game statistic (p -value) can be observed on Tables 4.12 and 4.13.

The results obtained from comparing the individual impact of each feature on each game statistic reinforced the drawn conclusions in the game questionnaire analysis.

TABLE 4.12. Mean and Standard Deviation, mean difference (M_Diff) and p_value for each game statistic to compare between LLM ON and LLM OFF.

Statistic	LLM-ON (M \pm STDEV)	LLM-OFF (M \pm STDEV)	M_Diff (ON - OFF)	p-value
Times Negotiated	3.68 \pm 1.47	4.75 \pm 1.88	-1.07	< 0.005*
Time Negotiating (in seconds)	93.64 \pm 56.88	32.57 \pm 24.39	61.07	< 0.005*
Difference in Price (in percentage)	66.04 \pm 25.28	67.49 \pm 19.74	-1.45	0.94

TABLE 4.13. Mean and Standard Deviation, mean difference (M_Diff) and p_value for each game statistic to compare between Animations ON and LLM OFF.

Statistic	Animations-ON (M \pm STDEV)	Animations-OFF (M \pm STDEV)	M_Diff (ON - OFF)	p-value
Times Negotiated	4.10 \pm 1.66	4.33 \pm 1.54	-0.23	0.56
Time Negotiating (in seconds)	59.95 \pm 37.66	66.26 \pm 41.83	-6.31	0.37
Difference in Price (in percentage)	64.56 \pm 19.30	68.97 \pm 29.97	-4.41	0.20

In this specific experiment, the presence of the LLM impacted two of the game statistics "Times Negotiated" and "Time Negotiating". On "Times Negotiated", LLM-OFF obtained a bigger mean value as the participants often tried to exploit the dialogue system by trying to suggest values under the minimum price and / or trying to keep suggesting prices after they had negotiated three times and the product's price was fixed. This exploitative behavior was more common on LLM-OFF game modes as the interaction through buttons was less immersive. On LLM-ON game modes this behavior did not occur as often as there was a time delay between the participant sending the message to the vendor NPC and the vendor NPC responding to the participant, and the participants were more immersed in the interaction as can be concluded from the results taken from the game questionnaire analysis done in the previous Section. The time delay also influenced the difference in the mean values between LLM-ON and LLM-OFF on "Time Negotiating", 61.07 seconds exactly. The fact that the participants enjoyed chatting and negotiating with the LLM also supported the "Time Negotiating" statistic being higher for LLM-ON than LLM-OFF. The "Difference in Prices" statistic, on the other hand, was not impacted by the use of the LLM.

The presence of animations, in opposition to the LLM's results, did not affect any of the existing game statistics, as can be seen on Table 4.13. This reinforces the hypothesis made on the game questionnaire section which was the unintentional blindness by the participants when playing the four game modes.

Just like the previous section, analyzing the pair-wise comparison of the two datasets consisting of only one feature produced clearer results than when analyzing the initial dataset consisting of the combination of the two features.

4.2.4. Post Game Questionnaire Results

As stated in Section 4.1, when the participants finished playing the game, they filled a post-game questionnaire that had the intention of evaluating the game experience as a whole. The results would determine if the developed game was a fitting demonstration of the proposed system and whether the results that were taken from it were credible and trustworthy.

The results from the first part of the post game questionnaire, which was the core module from the GEQ, can be observed on Table 4.14. Competence had a mean value of 2.71 which indicates that the participants answered statements regarding their ability to complete the game between "moderately" and "fairly", leaning more towards the latest. The component that evaluated the immersion had a mean value of 2.44 which indicates that the participants answered statements regarding their feeling of immersion in the game more often as "moderately". The same can be said for the flow component which evaluated the participants' concentration on the game, which has a mean value of 2.19. As for the components that evaluated negative feelings towards the game experience which were tension/annoyance and negative effect, they have the lowest mean values of 0.37 and 0.43, respectively, which indicates that the participants, when answering statements regarding their negative feelings, answered either "none at all" or "slightly", leaning more towards the "none at all" answer. The low mean score on the challenge component (mean value of 0.81) indicates that the participant group did not feel that the game was challenging, which aligned with the expectations, as the game was meant to be accessible for any type of player that would participate in this experiment, so it was developed with simple mechanics and a straightforward goal in mind in order to be engaging. The component which had the highest mean value, which was 3.03, was the positive affect one which evaluated positive feelings that the participant might have felt when playing. This indicates that the game left, overall, positive emotions on the participants which was the main goal when developing the game.

The second part of the questionnaire was the post game module from GEQ. The respective scores are present in Table 4.15. This module presents an additional set of statements to assess the player's experience when playing the developed game. Just like the first part of the questionnaire, the components that evaluate the negative feelings felt by the participants are the ones with the lower mean values, in this case negative experience with a mean value of 0.32 and tiredness with a mean value of 0.22, which

reinforces the idea that the participants enjoyed playing the game and felt, overall, almost no negative feelings towards it. The returning to reality component evaluated whether the participants felt the game was complex enough to feel like another reality. The low mean value of 0.54 aligned with the expectations since the game’s story is not complex and the time that the players took to finish the game was short. Lastly, the positive experience component, which evaluated the participants’ positive emotions after finishing the game, has the highest mean value, 1.81, of the components.

These results indicate that the participants, overall, enjoyed playing the game, felt immersed while doing so and experienced positive feelings more often than negative feelings. The higher mean values for the positive components and lower mean values for the negative components across both parts of the post game questionnaire reinforce this conclusion. As the game experience was overall positive, the results taken from the questionnaire during the game are trustworthy and credible.

TABLE 4.14. Mean and Standard Deviation for the GEQ Core Module components.

Component	(M ± STDEV)
Competence	2.71 ± 1.07
Sensory and Imaginative Immersion	2.44 ± 1.08
Flow	2.19 ± 1.20
Tension/Annoyance	0.37 ± 0.66
Challenge	0.81 ± 1.11
Negative affect	0.43 ± 0.69
Positive affect	3.03 ± 0.75

TABLE 4.15. Mean and Standard Deviation for the GEQ Post Game Module components.

Component	(M ± STDEV)
Positive Experience	1.81 ± 1.18
Negative Experience	0.32 ± 0.69
Tiredness	0.22 ± 0.61
Returning To Reality	0.54 ± 0.81

In the end of the questionnaire, the participants had a question where they could provide a written feedback about the game that they had played to reinforce their responses to the questionnaire. The feedback received is present in Table C.1 in the Appendices. Six out of the eighteen participants that left a written feedback (33.3%) explicitly expressed their appreciation towards the game and how they had enjoyed playing it. Twelve feedback comments (66.7%) were about how the participants had preferred negotiating with the vendor NPCs through natural language rather than by buttons. Five participants (27.8%) left suggestions for the game to be improved about the interfaces and how the difficulty could be increased.

4.3. Discussion

Overall, from the individual results in the previous sections, the more impactful outcomes of the evaluation of the proposed game are as follows:

- When analyzing the difference in the participants answers comparing the four game modes, for each question, the results were that the game modes were not statistically different enough to indicate whether a game mode was preferred overall.
- By analyzing the impact that each feature (LLM's text generation and presence of animations) had individually on the participant's answers, the results were clearer.
- When comparing the presence of the natural language feature on the interaction with a simple dialogue tree, the game questionnaire results had higher values for the natural language interactions. It impacted the three components of the GEQ's social presence module which indicates a higher empathy and connection towards the NPCs.
- When comparing the presence of non verbal cues on the interaction with their absence, the game questionnaire results had similar results, which indicate that the presence of the non verbal cues did not make a significant impact on how the participants perceive the interactions with the NPCs, according to the GEQ's social presence module.
- The results of the game statistics analysis reinforced the last outcome, as the natural language interaction impacted two out of the three game statistics, time spent negotiating and number of interactions between the player and the NPC, while the animations did not have an impact on none of them.
- The post game questionnaire revealed that the participants enjoyed playing the game and the experience overall, giving credibility to their game questionnaire results. When participants were asked to leave optional feedback, out of the eighteen received feedback comments, twelve expressed their preference for the natural language interaction and none expressed their preference for the interaction using the dialogue system.

[This page is intentionally left blank.]

CHAPTER 5

Conclusions

This chapter exhibits the conclusions of the conducted experiment along with the answers to the research questions. Lastly, we present limitations of this study along with improvements that can be made in future work.

5.1. Conclusions

This dissertation aimed to explore whether the usage of LLMs, to integrate NPCs in the interactions, to generate text and select non verbal cues, would improve the players' experience in negotiation focused games.

From the presented literature review, the first research question **RQ1** "Can LLMs generate text and choose non verbal cues efficiently when communicating with a human?" was answered, as the current leading LLM in this context, GPT-4, can successfully generate text and select contextually accurate non verbal cues when interacting with a human. The successful generation of text, in a negotiation setting, is also supported by the presented literature review. Regarding the second research question **RQ2** "Can LLMs generate text and choose non verbal cues efficiently when given the context of a game that involves a negotiation component?", the literature review was not clear as to whether, when in a negotiation context, the LLMs can select accurate non verbal cues. The same occurred regarding the third research question **RQ3** "Can the use of LLMs, in games that involve negotiating, to generate text and choose non verbal cues improve the gaming experience of the players?", in this case, the literature review did not produce any result. Therefore, a system was developed to evaluate the improvement of players' experience when using LLMs to generate text and non verbal cues in a negotiation setting. The developed system, and further developed game to employ it, in this dissertation aimed to respond to these unanswered research questions.

The concept of the system, when engaging in a negotiation with the interacting user, relied on the use of GPT-4 to generate a textual response to the user's input. After a textual response was generated, it was sent to the LLM again, along with the user input, so it could select the contextually coherent non verbal cues that would accompany the textual response. After selecting the non verbal cues, the textual information was sent to the interface, so the user could read it, and the AI controlled character would perform the non verbal cues. To evaluate this system, a negotiation focused game was developed which compared the previously concept with three other interaction modes. The four total modes were the absence of both LLM and non verbal cues, the presence of LLM but

absence of non verbal cues, the absence of LLM and presence of non verbal cues and the presence of both LLM and non verbal cues.

The game was then evaluated by 30 randomly chosen ISCTE students, who interacted with the NPCs in the four previously mentioned interaction modes. The participants had to fill a demographic questionnaire before initiating the experiment, a game questionnaire which assessed their feelings towards the NPCs for each interaction mode, for comparison, and a post game questionnaire to evaluate their overall experience. While playing the game, their interactions were recorded so further game statistics could be analyzed.

The demographic questionnaire revealed that the participant group was evenly distributed between casual and hardcore video game players, but leaned more to casual in negotiation focused games.

The results from the game questionnaire revealed interesting insights. When comparing the four modes, the results were not clear enough to point out which of the modes was the preferred overall. After separating the results to create two sub-groups, one consisting of the comparison of the presence of the LLMs and the other comparing the presence of the animations, the results were more clear. Having the player interacting in natural language had a strong impact on the participants' experience, impacting 72.22% of the game questionnaire. Contrastingly, the presence of animations only impacted 16.67% of the game questionnaire. The results of the game statistics further reinforce this outcome, as the presence of natural dialogue impacted two out of the three game statistics, while the presence of animations did not have an impact on any.

The results of the post game questionnaire revealed that the participants, overall, enjoyed the experience, and, in an optional feedback question, twelve out of the twenty feedback received referenced the participants' preference for the modes where they could negotiate in natural language.

With the combination of these results, as the players preferred the interactions where they could communicate with the natural language, since one of the modes had the LLM select the non verbal cues that went along with the text, a conclusion can be made regarding **RQ2**, which is that the LLM can efficiently select non verbal cues in a negotiation focused game context. However, regarding **RQ3**, the developed game only reinforced the theory that natural language improves the player experience in interactions in negotiation focused games, as the presence of non verbal cues, did not have a strong impact on the players' experience.

An hypothesis for these results was an unintentional blindness by the participants, whose attention was absorbed to the differences in the interacting interfaces and did not realize that there was a difference in the NPC non verbal modality.

These findings reveal that further research around the impact of non verbal modality in the player's experience when interacting with NPCs is necessary.

5.2. Limitations and Future Work

Firstly, as the developed game to test the proposed system was evaluated in ISCTE, the age range is limited to young adults. For future work it is important to evaluate the proposed system with other age ranges to understand whether age can be an impactful factor to consider when aiming to improve player's experience in negotiation focused games. Also, other age ranges can react differently to the different game modes, providing different results or reinforce the obtained results.

Secondly, the proposed system was evaluated on a negotiation focused game where the objective was to interact with NPCs to buy products. It is important to further evaluate the proposed system in other video games that include a negotiation component, and in other environment settings, like a war based game where the negotiations are revolving around the selling and buying of resources. The changing in the environment context can further validate the proposed system to understand if it can be generalized to other video game genres.

As the interactions in the developed game were not meant to be challenging, the amount of gestures that the LLM could use were also limited to match the difficulty of the interactions. It would be interesting to utilize the proposed system in a more complex negotiation environment and increase the amount of gestures it could use. If the proposed system was deployed in a complex strategy game, the amount of gestures could also be increased to increase the complexity of the proposed system.

[This page is intentionally left blank.]

References

- [1] Nuno Afonso and Rui Prada. Agents that relate: Improving the social believability of non-player characters in role-playing games. In *International Conference on Entertainment Computing*, pages 34–45. Springer, 2008.
- [2] Dorcas Quek Anderson. Negotiating with gpt-4: Digital doormat or skilful counterpart? *Australasian Dispute Resolution Journal*, 33(3):157, 2025.
- [3] Federico Bianchi, Patrick John Chia, Mert Yuksekgonul, Jacopo Tagliabue, Dan Jurafsky, and James Zou. How well can llms negotiate? negotiationarena platform and analysis. *arXiv preprint arXiv:2402.05863*, 2024.
- [4] Sze-yi Chan, Shihan Fu, Jiachen Li, Bingsheng Yao, Smit Desai, Mirjana Prpa, and Dakuo Wang. Human and llm-based voice assistant interaction: An analytical framework for user verbal and nonverbal behaviors. *arXiv preprint arXiv:2408.16465*, 2024.
- [5] Ainesh Chatterjee, Samuel Miller, and Nithin Parepally. Agree-mate: Teaching llms to haggle. *arXiv preprint arXiv:2412.18690*, 2024.
- [6] Changan Chen, Juze Zhang, Shrinidhi K Lakshmikanth, Yusu Fang, Ruizhi Shao, Gordon Wetzstein, Li Fei-Fei, and Ehsan Adeli. The language of motion: Unifying verbal and non-verbal language of 3d human motion. In *Proceedings of the Computer Vision and Pattern Recognition Conference*, pages 6200–6211, 2025.
- [7] Samuel Rhys Cox and Wei Tsang Ooi. Conversational interactions with npcs in llm-driven gaming: Guidelines from a content analysis of player feedback. In *International Workshop on Chatbot Research and Design*, pages 167–184. Springer, 2023.
- [8] Keya Rani Das and AHMR Imon. A brief review of tests for normality. *American Journal of Theoretical and Applied Statistics*, 5(1):5–12, 2016.
- [9] Robert Wall Emerson. Bonferroni correction and type i error. *Journal of Visual Impairment & Blindness*, 114(1):77–78, 2020.
- [10] Laura Ermi and Frans Mäyrä. Fundamental components of the gameplay experience: Analysing immersion. In *Proceedings of DiGRA 2005 conference: Changing views: Worlds in play*, 2005.
- [11] Sean Eshuis, Kay Pozzebon, Andrew Allen, and Lee Kannis-Dymand. Player experience and enjoyment: A preliminary examination of differences in video game genre. *Simulation & Gaming*, 54(2):209–220, 2023.
- [12] Yao Fu, Hao Peng, Tushar Khot, and Mirella Lapata. Improving language model negotiation with self-play and in-context learning from ai feedback. *arXiv preprint arXiv:2305.10142*, 2023.
- [13] Ayaka Fujii and Ken Fukuda. Initial study on robot emotional expression using manpu. In *Companion of the 2024 ACM/IEEE International Conference on Human-Robot Interaction*, pages 463–467, 2024.
- [14] Nan Gao, Zeyu Zhao, Zhi Zeng, Shuwu Zhang, Dongdong Weng, and Yihua Bao. Gesgpt: Speech gesture synthesis with text parsing from chatgpt. *IEEE Robotics and Automation Letters*, 9(3):2718–2725, 2024.
- [15] Cosmin Ghețău. Measuring video game dedication: The development and validation of casual hardcore assortment scale. *Journal of Media Research-Revista de Studii Media*, 15(42):80–90, 2022.
- [16] Zhenyu Guan, Xiangyu Kong, Fangwei Zhong, and Yizhou Wang. Richelieu: Self-evolving llm-based agents for ai diplomacy. *Advances in Neural Information Processing Systems*, 37:123471–123497, 2024.

- [17] Jason Patrick Winarto Hardiman, Dave Christian Thio, Alfi Yusrotis Zakiiyyah, et al. Ai-powered dialogues and quests generation in role-playing games using google’s gemini and sentence bert framework. *Procedia Computer Science*, 245:1111–1119, 2024.
- [18] Laura Birka Hensel, Nutchanon Yongsatianchot, Parisa Torshizi, Elena Minucci, and Stacy Marsella. Large language models in textual analysis for gesture selection. In *Proceedings of the 25th International Conference on Multimodal Interaction*, pages 378–387, 2023.
- [19] Sihao Hu, Tiansheng Huang, Gaowen Liu, Ramana Rao Kompella, Fatih Ilhan, Selim Furkan Tekin, Yichang Xu, Zachary Yahn, and Ling Liu. A survey on large language model-based game agents. *arXiv preprint arXiv:2404.02039*, 2024.
- [20] Wenyue Hua, Ollie Liu, Lingyao Li, Alfonso Amayuelas, Julie Chen, Lucas Jiang, Mingyu Jin, Lizhou Fan, Fei Sun, William Wang, et al. Game-theoretic llm: Agent workflow for negotiation games. *arXiv preprint arXiv:2411.05990*, 2024.
- [21] Dmitrii Iarovoi, Richard Hebblewhite, and Phoey Lee Teh. Ai’s influence on non-player character dialogue and gameplay experience. In *Science and Information Conference*, pages 76–92. Springer, 2024.
- [22] Wijnand A IJsselsteijn, Yvonne AW De Kort, and Karolien Poels. The game experience questionnaire. 2013.
- [23] Sunwoo Kim, Minwook Chang, Yoonhee Kim, and Jehee Lee. Body gesture generation for multimodal conversational agents. In *SIGGRAPH Asia 2024 Conference Papers*, pages 1–11, 2024.
- [24] Deuksin Kwon, Emily Weiss, Tara Kulshrestha, Kushal Chawla, Gale M Lucas, and Jonathan Gratch. Are llms effective negotiators? systematic evaluation of the multifaceted capabilities of llms in negotiation dialogues. *arXiv preprint arXiv:2402.13550*, 2024.
- [25] Yoon Kyung Lee, Yoonwon Jung, Gyuyi Kang, and Sowon Hahn. Developing social robots with empathetic non-verbal cues using large language models. *arXiv preprint arXiv:2308.16529*, 2023.
- [26] Eleanor Lin, James Hale, and Jonathan Gratch. Toward a better understanding of the emotional dynamics of negotiation with large language models. In *Proceedings of the Twenty-fourth International Symposium on Theory, Algorithmic Foundations, and Protocol Design for Mobile Networks and Mobile Computing*, pages 545–550, 2023.
- [27] Francesco Martinelli, Susanna Brambilla, Laura A Ripamonti, and Marco Rusciano. Fac. ex: Facial expressions for video game characters. In *2024 IEEE Conference on Games (CoG)*, pages 1–8. IEEE, 2024.
- [28] Ahmad Mouri Zadeh Khaki, Ahyoung Choi, and Laleh Seyyed-Kalantari. Simulating social behavior of llm-based autonomous negotiator agents in a game-theoretical framework using multi-agent systems. *International Journal of Human-Computer Interaction*, pages 1–10, 2025.
- [29] Fumie Nori, Afia Akhter Lipi, and Yukiko Nakano. Cultural difference in nonverbal behaviors in negotiation conversations: towards a model for culture-adapted conversational agents. In *International Conference on Universal Access in Human-Computer Interaction*, pages 410–419. Springer, 2011.
- [30] Se Jin Park, Yeonju Kim, Hyeongseop Rha, Bella Godiva, and Yong Man Ro. Av-emodialog: Chat with audio-visual users leveraging emotional cues. *arXiv preprint arXiv:2412.17292*, 2024.
- [31] Ken Peppers, Tuure Tuunanen, Marcus A Rothenberger, and Samir Chatterjee. A design science research methodology for information systems research. *Journal of management information systems*, 24(3):45–77, 2007.
- [32] Dulce G Pereira, Anabela Afonso, and Fátima Melo Medeiros. Overview of friedman’s test and post-hoc analysis. *Communications in Statistics-Simulation and Computation*, 44(10):2636–2653, 2015.
- [33] Stephanie Rennick and Seán G Roberts. Improving video game conversations with trope-informed design. *Game Studies*, 21(3), 2021.

- [34] José Luis González Sánchez, Francisco Luis Gutiérrez Vela, Francisco Montero Simarro, and Natalia Padilla-Zea. Playability: analysing user experience in video games. *Behaviour & information technology*, 31(10):1033–1054, 2012.
- [35] John Shearer, Patrick Olivier, Philip Heslop, Marco De Boni, et al. Requirements of non-verbal communication in believable synthetic agents. In *Proceedings of the AISB Symposium on Narrative AI and Games, 5th-6th April*, pages 62–69, 2006.
- [36] Michael R Sheldon, Michael J Fillyaw, and W Douglas Thompson. The use and interpretation of the friedman test in the analysis of ordinal-scale data in repeated measures designs. *Physiotherapy Research International*, 1(4):221–228, 1996.
- [37] Xiaofei Sun, Xiaoya Li, Shengyu Zhang, Shuhe Wang, Fei Wu, Jiwei Li, Tianwei Zhang, and Guoyin Wang. Sentiment analysis through llm negotiations. *arXiv preprint arXiv:2311.01876*, 2023.
- [38] Parisa Ghanad Torshizi, Laura B Hensel, Ari Shapiro, and Stacy C Marsella. Large language models for virtual human gesture selection. *arXiv preprint arXiv:2503.14408*, 2025.
- [39] Rustam Vahidov, Real Carbonneau, and Jaipriya Verma. Using negotiation and large language models in human-to software agent negotiations. *International Journal of Human-Computer Interaction*, pages 1–18, 2025.
- [40] Susanna Värtinen, Perttu Hämäläinen, and Christian Guckelsberger. Generating role-playing game quests with gpt language models. *IEEE transactions on games*, 16(1):127–139, 2022.
- [41] John R Venable, Jan Pries-Heje, and Richard L Baskerville. Choosing a design science research methodology. 2017.
- [42] Henrik Warpefelt and Harko Verhagen. A model of non-player character believability. *Journal of Gaming & Virtual Worlds*, 9(1):39–53, 2017.
- [43] Eric W Weisstein. Bonferroni correction. <https://mathworld.wolfram.com/>, 2004.
- [44] Philipp Wicke. Probing language models’ gesture understanding for enhanced human-ai interaction. *arXiv preprint arXiv:2401.17858*, 2024.
- [45] Josef Wiemeyer, Lennart Nacke, Christiane Moser, and Florian ‘Floyd’Mueller. Player experience. In *Serious games: Foundations, concepts and practice*, pages 243–271. Springer, 2016.
- [46] Jonathan Windle, Iain Matthews, and Sarah Taylor. Llanimation: Llama driven gesture animation. In *Computer Graphics Forum*, volume 43, page e15167. Wiley Online Library, 2024.
- [47] Claes Wohlin. Guidelines for snowballing in systematic literature studies and a replication in software engineering. In *Proceedings of the 18th international conference on evaluation and assessment in software engineering*, pages 1–10, 2014.
- [48] Robert F Woolson. Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials*, pages 1–3, 2007.
- [49] Tian Xia, Zhiwei He, Tong Ren, Yibo Miao, Zhuosheng Zhang, Yang Yang, and Rui Wang. Measuring bargaining abilities of llms: A benchmark and a buyer-enhancement method. *arXiv preprint arXiv:2402.15813*, 2024.
- [50] Xinrun Xu, Yuxin Wang, Chaoyi Xu, Ziluo Ding, Jiechuan Jiang, Zhiming Ding, and Börje F Karlsson. A survey on game playing agents and large models: Methods, applications, and challenges. *arXiv preprint arXiv:2403.10249*, 2024.
- [51] Bee Wah Yap and Chiaw Hock Sim. Comparisons of various types of normality tests. *Journal of Statistical Computation and Simulation*, 81(12):2141–2155, 2011.
- [52] Haolan Zhan, Yufei Wang, Tao Feng, Yuncheng Hua, Suraj Sharma, Zhuang Li, Lizhen Qu, Zhaleh Semnani Azad, Ingrid Zukerman, and Gholamreza Haffari. Let’s negotiate! a survey of negotiation dialogue systems. *arXiv preprint arXiv:2402.01097*, 2024.

[This page is intentionally left blank.]

APPENDIX A

Dialogue Tree

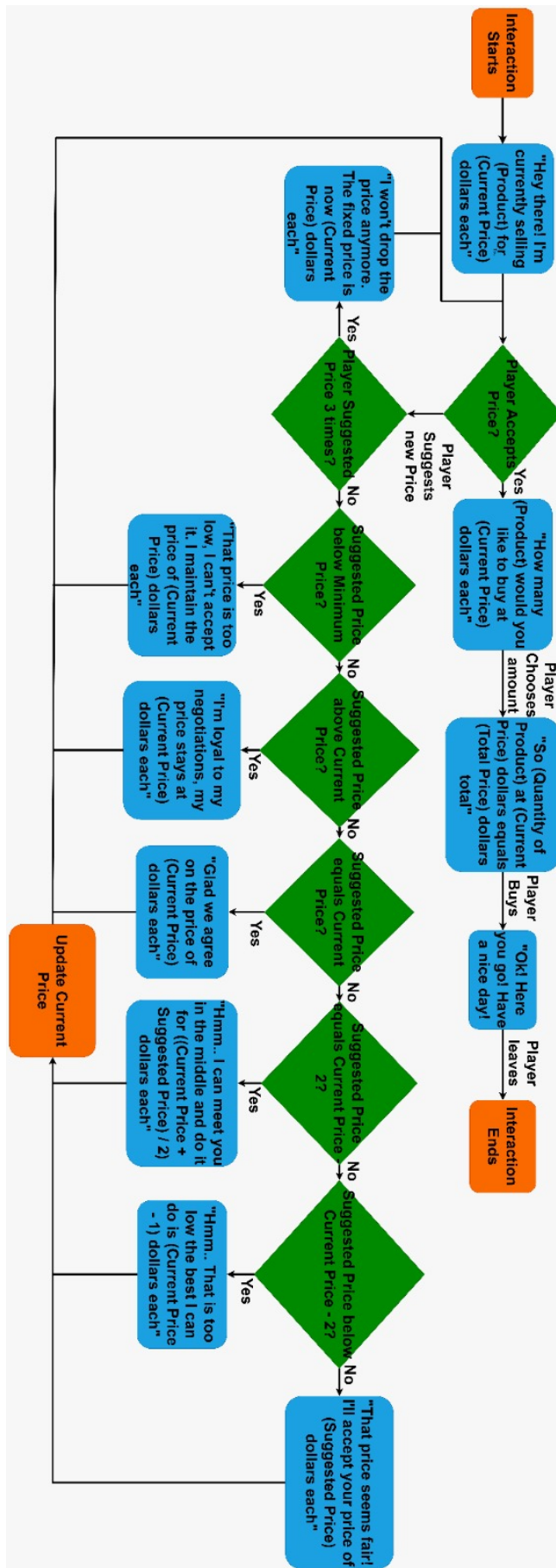


FIGURE A.1. Developed Dialogue Tree

APPENDIX B

Animations



FIGURE B.1. Animation Frames of Body Animation: Agree



FIGURE B.2. Animation Frames of Body Animation: Disappointed



FIGURE B.3. Animation Frames of Body Animation: Dismissing



FIGURE B.4. Animation Frames of Body Animation: Idle



FIGURE B.5. Animation Frames of Body Animation: Look Away



FIGURE B.6. Animation Frames of Body Animation: Pointing Forward



FIGURE B.7. Animation Frames of Body Animation: Shrugging



FIGURE B.8. Animation Frames of Body Animation: Surprised



FIGURE B.9. Animation Frames of Body Animation: Waving



FIGURE B.10. Animation Frames of Facial Expression: Annoyed



FIGURE B.11. Animation Frames of Facial Expression: Calm



FIGURE B.12. Animation Frames of Facial Expression: Surprised



FIGURE B.13. Animation Frames of Facial Expression: Happy



FIGURE B.14. Animation Frames of Facial Expression: Sad

[This page is intentionally left blank.]

APPENDIX C

Evaluation Documents

Initial Survey 🔍

O presente estudo surge no âmbito de um projeto de investigação a decorrer no **ISCTE – Instituto Universitário de Lisboa**. O estudo tem por objetivo estudar que elementos de jogo são importantes para a experiência dos jogadores em jogos que incluem componentes de negociação. A sua participação no estudo, que será muito valorizada, irá contribuir para o avanço do conhecimento neste domínio da ciência, consiste em jogar um jogo e responder a um questionário referente ao mesmo, demora por volta de 30 minutos.

O Iscte é o responsável pelo tratamento dos seus dados pessoais, recolhidos e tratados exclusivamente para as finalidades do estudo, tendo como base legal o seu consentimento o art. 6º, nº1, alínea a) e o art. 9º, nº2, alínea a). O estudo é realizado por Daniel Morgado (endereço eletrónico: dpmoa@iscte-iul.pt), que poderá contactar caso pretenda esclarecer uma dúvida, partilhar algum comentário ou exercer os seus direitos relativos ao tratamento dos seus dados pessoais. Poderá utilizar o contacto indicado para solicitar o acesso, a retificação, o apagamento ou a limitação do tratamento dos seus dados pessoais. A participação neste estudo é **confidencial**. Os seus dados pessoais serão sempre tratados por pessoal autorizado vinculado ao dever de sigilo e confidencialidade. O Iscte garante a utilização das técnicas, medidas organizativas e de segurança adequadas para proteger as informações pessoais. É exigido a todos os investigadores que mantenham os dados pessoais confidenciais. Além de confidencial, a participação no estudo é estritamente **voluntária**: pode escolher livremente participar ou não participar. Se tiver escolhido participar, pode interromper a participação e retirar o consentimento para o tratamento dos seus dados pessoais em qualquer momento, sem ter de prestar qualquer justificação. A retirada de consentimento não afeta a legalidade dos tratamentos anteriormente efetuados com base no consentimento prestado. Os seus dados pessoais serão conservados durante a análise estatística, após a qual serão destruídos ou anonimizados, garantindo-se o seu anonimato nos resultados do estudo, apenas divulgados para efeitos estatísticos, de ensino, comunicação em encontros ou publicações científicas. Não existem riscos significativos expectáveis associados à participação no estudo. O Iscte não divulga ou partilha com terceiros a informação relativa aos seus dados pessoais. O Iscte tem um Encarregado de Proteção de Dados, contactável através do email dpo@iscte-iul.pt. Caso considere necessário tem ainda o direito de apresentar reclamação à autoridade de controlo competente – Comissão Nacional de Proteção de Dados.

1

Declaro ter compreendido os objetivos de quanto me foi proposto e explicado pelo/a investigador/a, ter-me sido dada oportunidade de fazer todas as perguntas sobre o presente estudo e para todas elas ter obtido resposta esclarecedora. **Aceito** participar no estudo e consinto que os meus dados pessoais sejam utilizados de acordo com a informações que me foram disponibilizadas. * 🔍

Sim

Não

FIGURE C.1. Informed consent presented to the participants and questions asking for their consent.

2

Please indicate your age: *

The value must be a number

3

Please indicate your gender: *


Male

Female

Other


FIGURE C.2. Demographic Questionnaire, questions 1 and 2.

4

According to your experience in games, please indicate how much each statement represents it, on a scale from 1 to 5, where 1 means it doesn't represent your experience at all and 5 meaning it fully represents your experience. * 

	1	2	3	4	5
Over the last year, I consider that I have invested large sums of money in devices, gadgets, or other purchases dedicated to video games	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I tend to personalize everything with video game-specific elements	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I take video games very seriously	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I spend a lot of time in video games on a day off / weekend	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I spend a lot of time in video games on a work / school day	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

5


Have you ever played games where you could negotiate goods and / or belongings? (ex: Monopoly, Diplomacy, Catan, etc..) * 

Yes

No


FIGURE C.3. Demographic Questionnaire, questions 3 and 4.

6

If you said yes, which ones have you played? * 

Enter your answer


7

According to your experience in games where you can negotiate goods and / or belongings, please indicate how much each statement represents it, on a scale from 1 to 5, where 1 means it doesn't represent your experience at all and 5 meaning it fully represents your experience. * 

	1	2	3	4	5
Over the last year, I consider that I have invested large sums of money in devices, gadgets, or other purchases dedicated to video games	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I tend to personalize everything with video game-specific elements	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I take video games very seriously	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I spend a lot of time in video games on a day off / weekend	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I spend a lot of time in video games on a work / school day	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

FIGURE C.4. Demographic Questionnaire, questions 5 and 6.

8

When you played games where you did negotiate goods or belongings, you had to negotiate with... * 

- Only human controlled characters
- 70% human controlled characters, 30% AI controlled characters
- 50% human controlled characters, 50% AI controlled characters
- 30% human controlled characters, 70% AI controlled characters
- Only AI controlled characters

FIGURE C.5. Demographic Questionnaire, question 7.

Please indicate how you felt while interacting **with the virtual sellers** during the game for each of the items, on the following scale:

	not at all 0					slightly 1					moderately 2					fairly 3					extremely 4				
	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4	0	1	2	3	4
I empathized with the sellers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
My actions depended on the sellers' actions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The sellers' actions were dependent on my actions	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt connected to the sellers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
The sellers paid close attention to me	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I paid close attention to the sellers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt jealous about the sellers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found it enjoyable to be with the sellers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When I was happy, the sellers were happy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
When the sellers were happy, I was happy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I influenced the mood of the sellers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I was influenced by the sellers moods	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I admired the sellers	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
What the sellers did affected what I did	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
What I did affected what the sellers did	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt revengeful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt schadenfreude (malicious delight)	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It seemed I was negotiating with a human	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

FIGURE C.6. Game Questionnaire filled by participants during game.

9

*

Please indicate how you felt while playing the game for each of the items, on the following scale:

not at all
slightly
moderately
fairly
extrem

0
1
2
3
4

	0	1	2	3	4
I felt content	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt skilful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I was interested in the game's story	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I thought it was fun	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I was fully occupied with the game	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt happy	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It gave me a bad mood	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I thought about other things	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

FIGURE C.7. Post Game Questionnaire, question 1.

I found it tiresome	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt competent	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I thought it was hard	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It was aesthetically pleasing	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I forgot everything around me	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt good	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I was good at it	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt bored	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt successful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt imaginative	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt that I could explore things	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I enjoyed it	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

FIGURE C.8. Post Game Questionnaire, question 1 continuation.

Please indicate how you felt while playing the game for each of the items, on the following scale:

not at all
0

slightly
1

moderately
2

fairly
3

extremely
4

	0	1	2	3	4
I was fast at reaching the game's target	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt annoyed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt pressured	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt irritable	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I lost track of time	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt challenged	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found it impressive	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I was deeply concentrated in the game	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt frustrated	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It felt like a rich experience	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I lost connection with the outside world	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt time pressure	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I had to put a lot of effort into it	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

FIGURE C.9. Post Game Questionnaire, question 2.

	0	1	2	3	4
I felt revived	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt bad	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found it hard to get back to reality	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt guilty	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
It felt like a victory	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I found it a waste of time	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt energised	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt satisfied	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt disoriented	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt exhausted	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt that I could have done more useful things	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt powerful	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt weary	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt regret	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt ashamed	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I felt proud	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>
I had a sense that I had returned from a journey	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>	<input type="radio"/>

FIGURE C.10. Post Game Questionnaire, question 2 continuation.

Debriefing

Muito obrigado por ter participado neste estudo. Conforme adiantado no início da sua participação, o estudo incide sobre estudar que elementos de jogo são importantes para a experiência dos jogadores em jogos que incluem componentes de negociação. Reforçamos os dados de contacto que pode utilizar caso deseje colocar uma dúvida, partilhar algum comentário, ou assinalar a sua intenção de receber informação sobre os principais resultados e conclusões do estudo: Daniel Morgado – dpmoa@iscte-iul.pt; Pedro Mariano – pedro.mariano@iscte-iul.pt; Pedro Santana – pedro.santana@iscte-iul.pt. Mais uma vez, obrigado pela sua participação.

12

Do you have any comments that you would like to point out about the game you just played?

Enter your answer

FIGURE C.11. Debriefing and question to ask optional feedback.

TABLE C.1. Feedback given by each participant and the type it was assigned to for evaluation.

Feedback	I enjoyed the game	I enjoyed interacting with natural language	The game could be improved
"A quantia de 1000\$ pareceu demasiado elevada considerando que era perfeitamente possível comprar todos os itens da lista com esse dinheiro, e acabar com imenso de sobra."	-	-	X
"Acho que preferi a interação quando posso escrever. Quando tenho de definir só o número, acho que não sinto muito o challenge de alcançar um objetivo. Foi uma boa experiencia."	X	X	-
"Achei interessante falar por texto para negociar, ao contrario de propor apenas o preço"	-	X	-
"Quando escrevia a mensagem para os NPC, eu queria clicar no enter para submeter, em vez de clicar especificamente no botão com o botão do rato."	-	-	X
"Eu achei o jogo bastante interessante e acho que a partir do modo de jogo em que consegui escrever senti uma maior ligação com os vendedores."	X	X	-
"Quando entro no menu para comprar deveria dar para voltar para tras, para poder negociar. Gostei imenso do jogo continuaria a jogar."	X	-	X
"eu gostei do jogo"	X	-	-
"o facto de podermos responder por palavras proprias e termos as respostas de acordo com o que dizemos ajuda a melhorar a experiencia do usuario, e o facto de os sellers negociarem e cederem ou nao também torna a experiencia mais interessante e desafiadora"	-	X	-
"Eu gostei imenso do modo de jogo em que os NPC's respondiam diretamente ao que eu escrevia e tinham muito mais emoção que os outros modos. Apenas acrescentava a possibilidade de pedir uma dose concreta separadamente ao invés de renegociar outra vez a quantidade/preço"	-	X	X
"gostei da liberdade na escolha da resposta a partir dos ultimos dois modos(modos onde podia escrever)."	-	X	-
"Gostei mais dos modos de jogo em que o senhor/senhora da loja falavam (aqueles em que eu tinha de escrever as falas)."	-	X	-
"A utilização de LLMs em jogos pesados em texto parece uma área de investigação com potencial"	X	-	-
"O jogo estava esteticamente apelativo, notava-se uma clara evolução dos intervenientes do jogo ao longo dos niveis assim como algum aumento na sua complexidade. Jogo simples de entender e acessivel, com um objetivo claro e muito bom a levar a sua mensagem e missao a cabo. Poderia criar um nivel seguinte com um grau de envolvimento maior dos intervenientes com uma simulação de reação, background story e algumas reaões quer adversas quer positivas. No geral, boa experiencia."	X	-	X
"Gostei de explorar os diferentes comportamentos dos NPCs consuante as minhas respostas, senti que podia haver menos repetição e talvez mais margem de manobra em algumas situações. Gostei mais de escrever respostas do que valores"	-	X	-
"Gostei da interação por escrito com os npc's, dava mais possibilidades de interação"	-	X	-
"Eu gostei de poder escrever e interagir com o AI, os constantes apertos de mão foram um pouco irritantes, gostei do espectro de emoções que o AI demonstrou."	-	X	-
"Gostei especialmente dos modos de jogo em que pudemos escrever e interagir de uma maneira mais "real" com os negociantes."	-	X	-
"Gostei mais do modo de jogo no qual era possível escrever"	-	X	-
<i>Total (18)</i>	<i>6</i>	<i>12</i>	<i>5</i>

[This page is intentionally left blank.]

APPENDIX D

Extra Tables

TABLE D.1. Wilcoxon signed rank test with Bonferroni correction's p -value for all comparisons on questions 1-18.

Question	A-OFF_L-OFF A-OFF_L-ON	A-OFF_L-OFF A-ON_L-OFF	A-OFF_L-OFF A-ON_L-ON	A-OFF_L-ON A-ON_L-OFF	A-OFF_L-ON A-ON_L-ON	A-ON_L-OFF A-ON_L-ON
1	< 0.005*	0.97	< 0.005*	0.06	0.70	< 0.005*
2	0.13	1.00	0.07	0.77	1.00	0.78
3	1.00	1.00	0.12	1.00	0.65	0.35
4	< 0.005*	1.00	< 0.005*	0.03*	1.00	< 0.005*
5	0.01*	1.00	< 0.005*	0.01*	1.00	< 0.005*
6	0.16	1.00	0.04*	0.04*	1.00	0.01*
8	0.01*	0.35	< 0.005*	0.61	1.00	0.20
9	< 0.005*	1.00	0.01*	1.00	1.00	0.02*
10	0.44	1.00	0.01*	1.00	0.28	0.04*
11	1.00	0.34	0.01*	1.00	0.02*	0.06
12	< 0.005*	1.00	< 0.005*	0.53	0.32	0.02*
13	0.04*	1.00	0.03*	0.08	1.00	0.10
14	0.17	0.52	0.12	1.00	1.00	1.00
18	< 0.005*	1.00	< 0.005*	0.04*	0.44	< 0.005*

TABLE D.2. Wilcoxon signed rank test with Bonferroni correction's p -value for all comparisons on the two game statistics.

Game Statistic	A-OFF_L-OFF A-OFF_L-ON	A-OFF_L-OFF A-ON_L-OFF	A-OFF_L-OFF A-ON_L-ON	A-OFF_L-ON A-ON_L-OFF	A-OFF_L-ON A-ON_L-ON	A-ON_L-OFF A-ON_L-ON
Times Negotiated	0.41	1.00*	0.05*	0.04*	0.73	0.03*
Time Negotiating	< 0.005*	1.00	< 0.005*	< 0.005*	1.00	< 0.005*