Exploring the Applications and Limitations of Large Language Models: A Focus on ChatGPT in Virtual NPC Interactions.

Anand, Ayaan

https://researchdiscovery.drexel.edu/esploro/outputs/graduate/Exploring-the-Applications-and-Limitations-of/991021807413704721/filesAndLinks?index=0

Anand, A. (2023). Exploring the Applications and Limitations of Large Language Models: A Focus on ChatGPT in Virtual NPC Interactions [Drexel University]. https://doi.org/10.17918/00001888

It's your responsibility to determine if additional rights or permissions are needed for your use.
Downloaded On 2024/04/07 19:01:56 -0400

Please do not remove this page

GRADUATE THESIS/DISSERTATION APPROVAL FORM AND SIGNATURE PAGE

Instructions: This form must be completed by all master's and doctoral students with a thesis or dissertation requirement. Please type or print clearly as this form MUST be included as page 1 of your thesis or dissertation via electronic submission to ProQuest. All theses and dissertations must be formatted according to the University and department/program requirements. **Reminder:** It is the responsibility of the student to submit any/all edits requested by the Examining Committee to the Faculty Mentor or Supervising Professor for final approval and signature via the Graduate Program Completion Form.

Type: ☑ Master's Thesis ☐ PhD/Doctoral Thesis or Dissertation

Thesis or Dissertation Title: Exploring the Applications and Limitations of Large Language Models: A Focus on ChatGPT in Virtual NPC Interactions

Author's Name: Ayaan Anand

Month and Year: December, 2023

The signatures below certify that this [✓ thesis] / dissertation (circle one) is complete and approved by the Examining Committee.

Committee Chairperson's Name: Emil Polyak

Title: Associate Professor

Department: Digital Media

Institution (*if other than Drexel University*):

Signature: Digitally signed by Emil Polyak Date: 2023.12.07 13:45:34 -05'00'

Committee Member's Name: Michael Wagner

Title: Professor

Department: Digital Media

Institution (*if other than Drexel University*):

Signature: Michael G Wagner Digitally signed by Michael G Wagner Date: 2023.12.07 14:03:59 -05'00'

Committee Member's Name: Jervis Thompson

Title: Professor

Department: Digital Media

Institution (*if other than Drexel University*):

Signature: Jervis Thompson Digitally signed by Jervis Thompson Date: 2023.12.09 16:38:42 -05'00'

Committee Member's Name:

Title:

Department:

Institution (*if other than Drexel University*):

Signature:

Committee Member's Name:

Title:

Department:

Institution (*if other than Drexel University*):

Signature:

Committee Member's Name:

Title:

Department:

Institution (*if other than Drexel University*):

Signature:

Exploring the Applications and Limitations of Large Language Models: A Focus on ChatGPT in Virtual NPC Interactions

A Thesis Submitted to the Faculty

of

Drexel University

by

Ayaan Anand

In partial fulfillment of the

Requirements for the degree

Of

Masters of Science in Digital Media

2023

© Copyright 2023

Ayaan Anand. All Rights Reserved

Acknowledgment

I would like to thank my committee for all the support and knowledge they shared with me. A special thanks to Emil Polyak for guiding me throughout my master's degree.

Table of Contents

1. Introduction 1
  1.1 Problem Area 5
  1.2 Verbal De-escalation 8
  1.3 Research Question 10
  1.4 Applications of the Research 10
  1.5 Research Objective 12
  1.6 Structure 13
2. Literature Review 14
3. Methodology 34
  3.1 Emotional Parameters 41
  3.2 Interactive Storytelling 43
  3.2 Empathy 63
  3.3 Emotional Design 77
  3.4 Unusual Instances 87
  3.5 Best practices and Technicalities 87
  3.5 Police De-Escalation Testing 89
4. Conclusion 97
5. Future Work 99
Bibliography 102

FIGURE 1: IMAGE SHOWING THE DIFFERENCE BETWEEN MACHINE LEARNING AND DEEP LEARNING. SOURCE: (LGAYHARDT, 2023) 2

FIGURE 2: FLOWCHART EXPLAINING HOW THE CHATGPT MODEL WAS TRAINED. SOURCE: (INTRODUCING CHATGPT, N.D.-A) 4

FIGURE 3: SHOWING HOW THE APPLICATION CONNECTS WITH CHATGPT. SOURCE: AUTHORS OWN WORK 35

FIGURE 4: HOW AN AVATAR IS WRITTEN, AND THE VARIOUS VARIABLES INCLUDED IN THE FILE SUCH AS TEMPERATURE. SOURCE: GPTAVATOR CONFIGURATION FILE 36

FIGURE 5: SHOWS THE ECOSYSTEM BEING USED IN THE THESIS. SOURCE: AUTHORS OWN WORK 37

FIGURE 6: AN IMAGE SHOWCASING THE JAPANESE TEACHER AND THE UI. SOURCE: (ROBINSON, 2023/2023) 39

FIGURE 7: SHOWS UNUSUAL RESPONSE FROM SETH. SOURCE: AUTHORS OWN WORK AND GPTAVATAR 50

FIGURE 8: SHOWS THE TEXT IN FIG 7 51

FIGURE 9: SHOWS CODE THAT IS INCLUDED IN SETHS RESPONSE. SOURCE: AUTHORS OWN WORK AND GPTAVATAR 51

FIGURE 10: SHOWS CODE INCLUDED IN FIG 9 51

FIGURE 11: SHOWING THE BACK FORTH DIALOGUE. SOURCE: AUTHORS OWN WORK AND GPTAVATOR 54

FIGURE 12: TRANSCRIPT OF FIGURE 11 55

FIGURE 13: SHOWING SETHS RESPONSE. SOURCE: AUTHORS OWN WORK AND GPTAVATAR 57

FIGURE 14: TRANSCRIPT OF FIGURE 13 58

FIGURE 15: SHOWS SETH INTRODUCING HIMSELF AS CHATGPT. SOURCE: AUTHORS OWN WORK AND GPTAVATAR........60
FIGURE 16: TRANSCRIPT OF SETH'S INTRODUCTION........60
FIGURE 17: SHOWS SETHS COMPLEX RESPONSE SOURCE: AUTHORS OWN WORK.........67
FIGURE 18: TRANSCRIPT OF SETH'S COMPLEX RESPONSE FROM FIGURE 17........68
FIGURE 19: SHOWING AN UNUSUAL MOMENT FROM THE CONVERSATION. SOURCE: AUTHORS OWN WORK........82
FIGURE 20: TRANSCRIPT OF AN UNUSUAL MOMENT IN FIGURE 19........82
FIGURE 21: SHOWING A FLOWCHART OF THE POSSIBLE CHOICES USERS CAN MAKE. SOURCE: AUTHORS OWN WORK........88

Abstract

Exploring the Applications and Limitations of Large Language Models: A Focus on ChatGPT in Virtual NPC Interactions

Ayaan Anand

Large Language Models (LLMs), like ChatGPT and Google's Bard, have gained attention for their versatility and accessibility. This study delves into ChatGPT, exploring its potential to elevate virtual Non-Player Character (NPC) interactions, especially in domains such as law enforcement, mental health, and emergency services training. The primary goal was to assess ChatGPT's strengths and weaknesses through qualitative research, identifying opportunities for system enhancements. The evaluation criteria encompass parameters vital to digital media storytelling, including empathy, emotional design, and interactive storytelling. By elucidating the capabilities and constraints of LLMs like ChatGPT and their role in enhancing virtual NPC interactions, this research aims to contribute valuable insights to domains reliant on human-computer interactions and emotional engagement. Additionally, the study integrates scenarios used in training police officers into the researched application, offering best practices for configuring an NPC powered by an LLM. While the emotional parameters showed overall success, there were occasional inconsistencies and unconventional responses. While testing the application for training in verbal de-escalation, the basic scenario seemed to work better than its moderate and complex counterparts due to the straightforward nature of the scenario and since it involved only two characters, making use of the basic techniques involved in learning verbal de-escalation.

Keywords: Large -Language Model, ChatGPT, Interactive Story, Empathy, Emotional Design, Verbal De-escalation

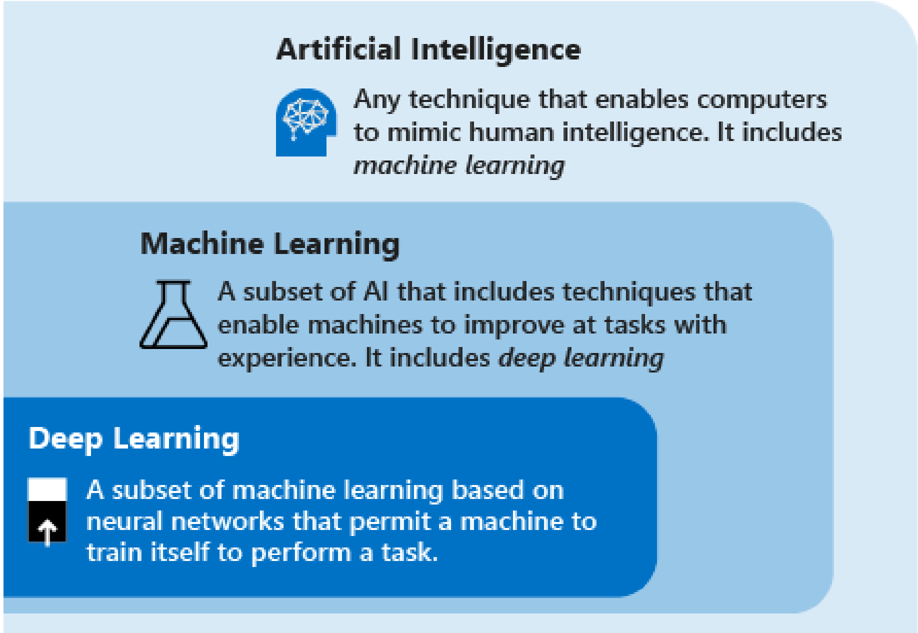
1. Introduction

To arrive at the definition of Large Language Models (LLM), I need to begin by elucidating the concept of Artificial Intelligence (AI) and the components that converge to formulate this concept. Initially, let's delve into what AI represents. The prominence of AI soared when Alan Turing, also recognized as the "Father of Computer Science," posed the pivotal question, "Can machines think?" (TURING, 1950). In his paper, Turing introduces a game involving three participants: a human judge, a human player, and a computer player. The human judge receives text responses from both players and must discern which one is the actual human.

This discourse on AI was initiated by Turing's paper, sparking debates on the capacity of machines to genuinely think. The exploration of AI continued, and Stuart Russell's book on AI added significant insights, presenting four distinct answers to the question of what AI constitutes (Russell & Norvig, 2010). These responses were classified into four categories: Thinking Humanly, Thinking Rationally, Acting Humanly, and Acting Rationally.

These classifications imply that AI involves the computer's ability to emulate human thinking and behavior (Russell & Norvig, 2010). Within the realm of AI, specific components like Machine Learning (ML) and Natural Language Processing (NLP) play crucial roles. Simply put, NLP enables a computer to comprehend speech and text in a manner akin to human understanding (*What Is Natural Language Processing?*, n.d.), while ML involves the data and algorithms that a computer learns from, much like human learning, to enhance its own information and output (*What Is Machine Learning?*, n.d.).

Before delving into the core definition, it's essential to consider another concept: Deep Learning. Deep learning is an AI methodology that enables systems to learn and process data similarly to human beings (*What Is Deep Learning?*, n.d.). This encompasses various types of data, including images, text, sound, and more (*What Is Deep Learning?*, n.d.).



Any technique that enables computers to mimic human intelligence. It includes *machine learning*

A subset of AI that includes techniques that enable machines to improve at tasks with experience. It includes *deep learning*

A subset of machine learning based on neural networks that permit a machine to train itself to perform a task.

Figure 1: Image showing the difference between Machine learning and Deep learning. Source: (lgayhardt, 2023)

Deep learning serves as a technique within ML for acquiring the ability to learn and process new data, constituting a subfield of ML (Team, 2023). Armed with this knowledge, the next immediate objective is to articulate the definition of a Large Language Model (LLM). Essentially, LLMs are models trained through deep learning on highly specific datasets. The functionality of each LLM is distinct, and shaped by the dataset on which it was trained (*What Are Large Language Models?*, n.d.). This definition holds significance for this research, as I employ OpenAI's LLM known as ChatGPT.

Utilizing a model developed by OpenAI, the research involved the amalgamation of a newly crafted dataset and an existing model from OpenAI called InstructGPT (*Introducing ChatGPT*, n.d.-b)

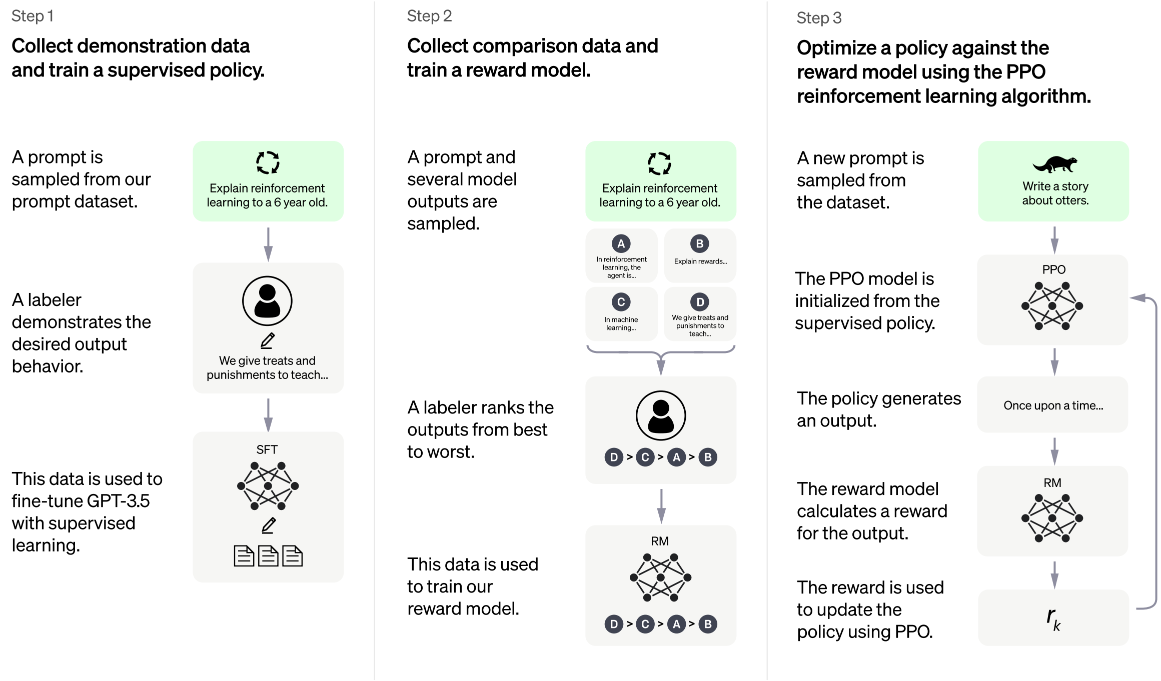

Step 1
Collect demonstration data and train a supervised policy.
A prompt is sampled from our prompt dataset.
Explain reinforcement learning to a 6 year old.
A labeler demonstrates the desired output behavior.
We give treats and punishments to teach...
This data is used to fine-tune GPT-3.5 with supervised learning.
SFT
Step 2
Collect comparison data and train a reward model.
A prompt and several model outputs are sampled.
Explain reinforcement learning to a 6 year old.
A
In reinforcement learning, the agent is...
B
Explain rewards...
C
In machine learning...
D
We give treats and punishments to teach...
A labeler ranks the outputs from best to worst.
D > C > A > B
This data is used to train our reward model.
RM
D > C > A > B
Step 3
Optimize a policy against the reward model using the PPO reinforcement learning algorithm.
A new prompt is sampled from the dataset.
Write a story about otters.
The PPO model is initialized from the supervised policy.
PPO
The policy generates an output.
Once upon a time...
The reward model calculates a reward for the output.
RM
The reward is used to update the policy using PPO.
r_k

Figure 2: Flowchart explaining how the ChatGPT model was trained. Source: (Introducing ChatGPT, *n.d.-a)*

The training process for models like ChatGPT involved an iterative approach that incorporated Proximal Policy Optimization (PPO)(*Proximal Policy Optimization*, n.d.). Initially, the model underwent supervised fine-tuning, with human trainers engaging in conversations while playing both user and AI assistant roles. The dialogue dataset was combined with existing datasets and transformed for dialogue format compatibility. To further enhance the model's responses, a reward model was created by having human trainers rank alternative responses generated by the model during real conversations (*Introducing ChatGPT*, n.d.-b).

PPO is then applied to fine-tune the model iteratively. It is an algorithm used in reinforcement learning, ensuring that policy changes made during training are controlled and gradual. This strikes a balance between exploration and exploitation. Essentially, PPO aids the model in learning and optimizing its decision-making process over multiple iterations, gradually enhancing its ability to generate high-quality responses in various conversational contexts (*Introducing ChatGPT*, n.d.-b).

1.1 Problem Area

The World Economic Forum suggests that LLMs can increasingly be recommended for training individuals in various fields (*Here's How AI Can Train Workers for the Jobs of the Future*, 2020). These training techniques cover a spectrum from acquiring and honing new skills to integrating learning into everyday activities. In the work of Kasneci et al., it is explored how LLMs, like ChatGPT, might find application in education. However, the authors emphasize that these models are intended as supportive tools rather than complete replacements for established norms and roles, such as those of teachers, and they call for further research (Kasneci et al., 2023).

In addition to other applications, LLMs could potentially serve as a tool for teachers to safeguard themselves in situations that might escalate to violence, especially instances where they face threats from students. According to data from 2015-2016, at least 10% of teachers reported being threatened with injury by students, and 6% of teachers experienced actual physical attacks from students (Musu-Gillette et al., 2018). A survey involving over 3000 participants revealed that one-fifth of them do not report violent incidents, underscoring a significant issue in the field of education that requires effective solutions (Anderman et al., 2018).

Examining other professions, such as law enforcement, reveals a concerning trend. From January 1, 2021, to September 30, 2021, there was a 51% increase in police deaths in the line of duty compared to the previous year of 2020 (Melinda & Chaumont, 2021). In 2020, nationally, over sixty thousand officers were assaulted while on duty, which marked an increase of over 4000 incidents from the previous year, 2019 (Melinda & Chaumont, 2021). Officers face a higher risk of assault when responding to calls related to family disturbances than any other specific type of police call (Hendricks & Thomas, 1990, as cited in Olivia et al., 2010) (Miller & Braswell, 2002, as cited in Olivia et al., 2010). When a law enforcement officer addresses a crisis involving an individual with a severe mental illness lacking treatment, the safety of both the person in crisis and the responding officer might be at risk. This is especially true when the officer has minimal or no training in mental illnesses and crisis intervention (Miller & Braswell, 2002, as cited in Olivia et al., 2010) (Oliva et al, 2007, as cited in Olivia et al., 2010).

Emergency Medical Services (EMS) workers face significant challenges, with paramedics frequently encountering violence in the field. Taylor et al. highlight a concerning trend, reporting a 23% increase in calls for Emergency Medical services from 2014 to 2015 (National EMS Information System 2016 as cited in Taylor et al., 2019). Another study by Maguire et al. found that violence initiated by patients had a 77% chance, while coworkers had an 8% chance of initiating violence, with female EMS personnel facing a higher risk (Maguire & O'Neill, 2017). Despite the age of the study by Mock et al. in 1998, which emphasized the need for EMS providers to be vigilant about scene safety, the concern for safety remains crucial for EMS personnel (Mock et al., 1998)

This technology presents the opportunity to deliver targeted training to professionals in roles demanding effective verbal communication skills, allowing them to adeptly navigate various challenges. For instance, it can be employed to address conflicts between teachers and students or among workplace colleagues. Additionally, it may play a crucial role in verbally de-escalating potentially violent situations, as encountered by police officers' or during EMS interventions. The application extends to handling incidents of workplace abuse across various professions or office settings.

In their literature review, Piquero et al. deduce that workplace violence in each profession can vary based on the aggressor, and the factors contributing to workplace violence are connected to stress, intentionality, or situational dynamics (Piquero et al., 2013).

In their research, Gallant-Roman discusses methods nurses can employ to mitigate violence against them. Of particular relevance to this research are two key aspects: predicting high-risk events to know when to leave if violence or aggression begins, and educating staff on workplace violence, addressing stress-inducing factors (Gallant-Roman MA, 2008).

There are also signs that workers may exhibit, which could be indicators of potential issues. Some of these signs include impersonal conflict, depression, and a tendency to become quickly frustrated. With this system, these scenarios could be identified through monthly or yearly examinations with a system for communication. This proactive approach may help prevent any violent or aggressive situations from occurring. While the context is framed around nurses, these indicators can be applicable to other workplaces as well.

1.2 Verbal De-escalation

Let's begin by establishing what constitutes verbal de-escalation. In their review, Roberton et al. note that there is no universally agreed-upon definition for verbal de-escalation, as these definitions are seldom comprehensive and rarely grounded in models or research (Roberton et al., 2012). Mavandadi et al. similarly point out a lack of consensus on the techniques encompassing verbal de-escalation (Mavandadi et al., 2016). This research has adopted the definition provided by Cowin et al., which describes verbal de-escalation as the gradual reduction of tension and abusive behavior through verbal and physical actions that may elicit expressions of empathy, boundaries, and friendship (Cowin et al., 2003). Given that the application used in this research lacks a physical context, this aspect may be omitted from the definition for the purposes of this research.

Examining the efficacy of verbal de-escalation, a survey involving 72 clinical staff sheds light on their perspectives. The majority of participants identified the purpose of verbal de-escalation as calming down an aggressive individual or preventing a violent outburst (Hallett & Dickens, 2015). About half of the participants highlighted specific techniques, including verbal communication skills, emphasizing the importance of tailoring approaches to the unique context and phase of the aggressive party (Hallett & Dickens, 2015). Communication emerged as the most common technique, with calm talk being the predominant strategy. This survey underscores the crucial role of communication, indicating that the system being researched has the potential to contribute significantly to worker training.

The goal of this thesis encompasses several facets. It involves delving into the unexplored capabilities of LLMs, specifically the ChatGPT model, to transform personalized virtual Non-Player Character (NPC) interactions through three emotional parameters for emotive engagement. Additionally, it seeks to evaluate the feasibility of customizing the application in training police officers in verbal de-escalation.

Furthermore, it aims to offer insights into the optimal methods for configuring LLMs, tailored to user preferences, and doing so with efficiency.

1.3 Research Questions

The research questions for our research are as follows:

1. How can Large Language Models, specifically ChatGPT, be harnessed to create personalized virtual NPC interactions that prioritize emotional design for therapy, verbal de-escalation in law enforcement and EMS settings, and resolving student-teacher conflicts?

2. What are the unique challenges and opportunities presented by the integration of LLMs into emotionally charged domains in serious applications, and how can these challenges be effectively addressed?

1.4 Applications of the Research

The application can find utility in training for behavioral correction, law enforcement (verbal de-escalation), therapy, and student-teacher incidents. Given the nature of these fields, which often involve serious conversations and intense emotions, the chosen parameters for the research are particularly relevant.

In law enforcement, the objective is to offer a means of averting aggressive conflict and influencing behavioral outcomes, as emphasized by Engel et al. (Engel et al., 2020). This involves settling matters through verbal conversation and diminishing the necessity for extreme force in law enforcement situations. Research, such as that by (Olivia et al., 2010), supports the effectiveness of verbal de-escalation in reducing police injuries and successfully managing situations. Olivia et al. also underscores the value of role-playing situations as a recognized tool for teaching verbal de-escalation skills to police officers (Olivia et al., 2010).

Our system can facilitate this by adjusting the AI's configuration and embodying various characters to simulate individuals encountered by the police. Incorporating features like voice modulation and lip synchronization may enhance the immersive aspect of these interactions (Peixoto et al., 2021). ChatGPT can effectively portray characters with intellectual disabilities or individuals exhibiting aggression, providing a dynamic training environment.

In situations involving student-teacher conflicts, our system can serve as a preliminary tool for engaging with students or as a training aid for teachers. It can assist teachers in practicing how to address aggression and understanding the reasons behind such behaviors. Given that approximately 80% of teachers have reported experiencing at least one incident of grievance in their careers, and 94% attribute these incidents to students (Mcmahon et al., 2014), there is a clear need for training. Mcmahon et al. emphasize that teachers require training to comprehend the concerns of students, enabling better preparation for such issues (Mcmahon et al., 2014). The system can contribute by quizzing teachers through role-playing scenarios, enhancing their readiness for potential conflicts and providing behavior correction for situations where violence might be imminent.

Emergency Medical Services (EMS) workers face the challenge of violence in their line of duty. A survey conducted in 2015 with almost 1800 participants revealed that 69% of EMS workers reported experiencing violence in the last 12 months while on duty (Roman, 2019). This violence ranged from being spat on, slapped, and punched to verbal abuse. According to CDC data in 2020, violence was the second-highest cause of injury for EMS personnel (*EMS Workers - Injury and Illness Data | NIOSH | CDC*, 2022). This research addresses this issue by integrating training for verbal de-escalation into the workflow of EMS workers who face potential harm in their jobs. Leveraging an LLM, the AI can simulate scenarios where EMS personnel need to verbally de-escalate an abusive victim before providing medical assistance.

1.5 Research Objective

The primary goal of this research is to evaluate the effectiveness of the ChatGPT LLM as a training tool with emotional intelligence, especially in fields reliant on verbal communication such as therapy. Additionally, the research aims to explore its potential as a preventive tool for reducing conflict and violence by acting as a training tool for verbal de-escalation. Another aspect of the study involves identifying best practices for writing efficient configurations. The analysis will encompass the behavior and responses of the virtual agent across various inputs. This evaluation will delve into the tone of interactions, agent-specific characteristics, the presence of verbal aggression, content analysis of conversations, identification of triggering elements, and the exploration of the agent's response dynamicity.

The research specifically chose Interactive Storytelling, Empathy, and Emotional Design to test. Given that the application of this system involves engaging in serious conversations with humans, these parameters are crucial. They enable an assessment of how well the AI comprehends emotions and how users might emotionally connect with it. Moreover, these findings could have broader applications, extending to fields like video game narratives and addressing users' more casual needs.

1.6 Structure

The organizational framework of this thesis encompasses a review of existing literature, an elaboration of the methodology (which will encompass an in-depth explanation of the utilized research system and its operational procedures), a discussion section, and considerations for future research.

2. Literature Review

Despite ChatGPT being a relatively recent development, the broader landscape of research on LLMs and AI has seen significant advancements. One noteworthy paper is "An Intelligent System for Aggression De-Escalation Training" by Bosse et al. While our research is centered on utilizing ChatGPT for training professionals in various sectors like law enforcement, EMS, and conflict resolution, this research draws inspiration from this paper by Bosse et al., which delves into principles that might be applicable to our work. Notably, the AI system examined in that study differs from ChatGPT in its nature (Bosse et al., 2016).

In their research, Bosse et al. present an intelligent system designed to assist workers and professionals in the public domain in enhancing their verbal de-escalation skills (Bosse et al., 2016). This system is an outcome of the integrative STRESS project, involving collaboration among contributors from diverse fields, including academia (Bosse et al., 2016). Their system utilizes virtual scenarios to engage users, providing interactive environments for practice, and incorporates an AI-powered virtual agent for interaction during de-escalation training (Bosse et al., 2016).

The learning objective of this paper is to enhance emotional intelligence, allowing users to discern the emotional state of virtual agents. Achieving this objective enables users to tailor their responses effectively in accordance with the perceived emotional state

(Bosse et al., 2016). The study categorizes aggression into two types: emotional and instrumental. Emotional aggression involves an outburst of emotion in response to an adverse event, like getting angry when a "tram is late" (Bosse et al., 2016). Conversely, instrumental aggression entails using aggression to achieve a specific goal, such as making threats when denied free travel (Bosse et al., 2016).

The primary objective of the research is to create a system capable of recognizing the trainee's behavior during interactions with aggressive individuals and providing feedback to enhance their performance (Bosse et al., 2016). The training system comprises various components working cohesively, including virtual agents and a simulated environment (Bosse et al., 2016). Training involves placing the user in a virtual environment specific to an occupation, where they interact with agents displaying aggressive behavior and employ their training to de-escalate the situation (Bosse et al., 2016). The system dynamically responds to the situation based on the users' performance (Bosse et al., 2016). Additionally, the system employs a conversation tree.

The paper conducted two case studies in fields relevant to public transport and law enforcement. In the case study related to public transport, users expressed positivity about the system's content, its potential for interactions, and its learning effects. However, they were not as enthusiastic about the emotional impact of the system (Bosse et al., 2016).

Similarly, in the law enforcement case study, participants showed positivity about the potential for interaction but did not view it as an effective training tool (Bosse et al., 2016).

Overall, the system received positive feedback, but it fell short in terms of the emotional aspect, with several participants expressing that they did not feel emotionally invested in the virtual environment (Bosse et al., 2016). Bosse et al. suggest potential improvements for enhancing emotional feedback, such as incorporating a head-mounted display or adding haptic feedback to the system (Bosse et al., 2016).

This paper has been helpful in the research by indicating that users require there to be an emotional aspect and the way the system works can be possible. While it is not a one-to-one comparison of what this research is doing, some of the principles learned in this paper can be transferred over to the current research. The emotional aspect is something to keep note of.

The next paper being examined studies Virtual Reality (VR) technology for de-escalation concerning clinical employees but does make use of a chatbot called DriftAI, which was specifically designed for the Code Black VR: Verbal De-escalation Training application (Moore et al., 2022). As with the previous paper, this study offers valuable insights that can inspire this research.

The aim of this study was to assess the validity of their application and identify specific design requirements for a VR-based agent, especially in the context of verbally de-escalating situations involving clinicians (Moore et al., 2022). Twenty-eight participants, including clinicians, educators, and junior and intermediate-level clinicians, participated in the study. They completed multiple playthroughs of the application and later provided written comments about their experience and filled out a survey (Moore et al., 2022).

The proprietary AI at the center of this study is called DriftAI, a chatbot that drives the animation and voice of the in-game virtual agent based on user responses. The AI initiates the interaction with an angry demeanor, requiring users to verbally de-escalate the situation. Upon user responses, the AI studies them and reacts appropriately. If the AI is already familiar with the response, it provides a vocal reply; otherwise, it formulates a textual response. The AI's attitude varies based on the user's input (Moore et al., 2022).

The findings of the researchers offer valuable insights that can be relevant to this research involving virtual agents. These insights were derived from a thematic analysis of the results of the study. Notably, perspectives from educators highlighted limitations, particularly in the recognition of advanced verbal de-escalation skills. For example, the AI struggled to recognize lengthy verbal responses, impacting participants' performance based on their personal experiences. Participants concluded that the AI did not respond accurately, and there was no consideration given to physical proximity; the AI stood too close to the participants. Participants suggested that standing farther away could be considered a form of verbal de-escalation (Moore et al., 2022).

Additionally, certain factors emerged, including the application being confusing and "overwhelming" to people trying it for the first time. The sentiment extended to the inclusion of a push-to-talk button, which was seen as unwelcome due to instances of being forgotten or considering it unreasonable (Moore et al., 2022).

Other factors found during the research regarding the AI were that the AI sometimes inaccurately interpreted what was said by the participants. Participants asked for a virtual environment that would simulate a real-life clinical environment and the ability to perform other realistic tasks, such as moving away and towards the virtual agent. They also desired the AI agent to generate realistic facial expressions, as sometimes the facial expressions did not faithfully represent an angry person who needs to be verbally de-escalated. Another issue was that many times, the AI would repeat its phrases despite different responses being given (Moore et al., 2022). The authors also documented that a tutorial would be needed to help ease into this new technology for people.

The findings from the study regarding DriftAI are that verbal de-escalation training with the bot is possible and is efficient. But the factors provided can be established as a framework of sorts to improve verbal de-escalation training more effectively.

These two studies offer principles applicable to this research. The initial paper emphasizes the integration of emotion into an AI. This can be achieved by incorporating a narrative into the system, creating a set of parameters, as discussed earlier, to facilitate an emotionally intelligent LLM experience.

The latter study provides a framework for enhancing participant experience. This includes considerations such as the AI accommodating lengthy verbal responses, recognizing facial and emotional cues, and introducing users to the software before commencing training. While the current system design is already established and cannot be altered, certain aspects outlined by Moore et al. differ. Notably, ChatGPT demonstrates intelligence in deducing the correct verbiage when mishearing the user, ensuring relevant responses. Additionally, the application being used for this research has integrated lip-syncing into the system design to enhance immersion.

The next paper under consideration explores the impact of player engagement in conversations with NPCs driven by LLM and generating contextually aware dialogue (Csepregi, 2023). This experimental study involved 21 testers and aimed to assess the quality of conversation and player engagement. While not directly aligned with the current research, the use of an LLM in this study may offer valuable insights and properties that could be beneficial.

The objective of the paper is to assess the impact of a contextually aware conversational agent on player engagement during interactions. The methodology involves participants engaging with an NPC in a specific role-playing scenario within the game environment (Csepregi, 2023). Data is collected through player feedback to explore how contextually aware NPC conversations can enhance engagement, satisfaction, and connectivity in the gaming realm (Csepregi, 2023). The success of the study hinges on two factors: demonstrating that contextually aware conversations indeed increase engagement and establishing effective guidelines or rules for implementing this concept in NPCs to achieve maximum satisfaction (Csepregi, 2023).

The research employed two types of NPC conversations: baseline and context-aware, utilizing a within-subject design (Csepregi, 2023). All 21 testers participated in both baseline and contextually aware conversations with the NPC.

The baseline version featured an NPC powered by the ChatGPT LLM with a voice, offering varied responses. However, it had limited options for interactions between players and NPCs. In comparison, the contextually aware NPC conversations, also powered by ChatGPT, incorporated added context. This context was influenced by players pointing at objects, choices, or intents mentioned in their answers, and it depended on the changing state of the environment during testing (Csepregi, 2023). This approach allows for a direct comparison between the two methods. One aspect of inspiration for our research lies in how responses are handled when players have an open goal, enabling them to converse freely with the NPC about any topic (Csepregi, 2023).

In the research, the LLM is assigned a specific task, guiding the interaction in a particular direction, but the approach to achieving that task is open-ended, allowing testers to pursue it as they see fit. Drawing inspiration, I might consider applying a preexisting personality and identification to the NPC to align with our research goals. Additionally, the observation that the push-to-talk feature was initially confusing to testers but became familiar over time is promising for our research, given that the research is utilizing a pre-made application with an existing design and UI (Csepregi, 2023).

The research findings strongly indicated a positive impact on player engagement with contextually aware NPCs. Participants reported heightened satisfaction as the NPCs reacted appropriately to their actions with a dynamic response. Contextually backed NPCs offered more enjoyment and surprise to the players (Csepregi, 2023). However, there were also negative aspects, such as the duration for each response from the LLM to generate. This aspect aligns with our research focus on efficiency, prompting the development of guidelines for efficiently crafting emotionally intelligent AI responses (Csepregi, 2023).

This research has been beneficial to study by providing insights into communication strategies with NPCs, configuration considerations, and highlighting its limitations, particularly its focus on a specific RPG scenario. The acknowledgment of not delving into the negatives of LLM systems and the suggestion for psychological or gameplay-related analyses add nuance to our understanding (Csepregi, 2023).

The RPG-focused aspect of the research may not directly impact our study, which concentrates on real-world issues and occupations. However, it offered valuable guidance on NPC configuration and providing users with some freedom in their interactions. Although the psychological aspect is absent in both studies, our emphasis on efficiently writing AI configurations for serious conversations with emotionally intelligent AI aligns with our research objectives.

Now, examining an overview of verbal de-escalation techniques incorporated into the Crisis Intervention Team (CIT) curriculum. This training program is designed for law enforcement to provide them with the knowledge and skills necessary for verbally de-escalating situations (Olivia et al., 2010). Olivia et al. initially discusses the need for such skills in the police department, emphasizing the expansion of policing beyond crime prevention and fighting to include service-oriented features. Recognizing and identifying situations that involve individuals in problems where assistance can be provided without resorting to violence or risking injury becomes crucial in this evolving role of law enforcement (Olivia et al., 2010).

This paper is valuable for our research as it delineates specific skills that can be assessed using our system. It highlights fundamental de-escalation skills, such as maintaining calmness, speaking slowly, and using short sentences to encourage communication (Olivia et al., 2010). The officer handling the situation should also exhibit calmness, focus, and a genuine sense of understanding and willingness to help (Olivia et al., 2010). These are foundational skills, and in addition to them, police officers should possess a comprehensive understanding of the events leading to the problem, essentially grasping the context of the triggering situation.

The insights from Olivia et al., referencing the National Institute of Justice, emphasize the diverse levels of force available to police officers based on specific situations (Olivia et al., 2010; *The Use-of-Force Continuum*, 2009). The document underscores that often an officer's mere presence can act as a deterrent, provided they behave appropriately and non-threateningly. The crucial aspect for our research is verbalization in the use-of-force continuum, where an officer may issue a calm, non-threatening command to prompt compliance. Depending on the individual, the officer might adjust their voice volume and issue specific commands to ensure compliance (Olivia et al., 2010; The Use-of-Force Continuum, 2009).

Olivia et al. highlight active listening skills and effective communication as vital components of verbal de-escalation. Ensuring that the person involved in the crisis comprehends the commands issued by the officer is essential for a successful resolution (Olivia et al., 2010). Effective communication, in this context, involves the exchange of information between two parties in a manner that is mutually understood.

Referencing other sources, Olivia et al. assert that active listening allows officers and individuals involved in a crisis to better understand each other, ultimately reducing tension (Goldstein et al., 1973, as cited in Olivia et al., 2010)(*Police Response To People With Mental Illnesses: Trainers Guide and Model Policy | Office of Justice Programs*, 1997, as cited in Olivia et al., 2010) (Olivia et al., 2010). They conclude with insight from Richards, who suggests that officers should be prepared to invest the necessary time for de-escalation, emphasizing the importance of a patient approach (Richards, 2007, as cited in Olivia et al., 2010).

Regarding active listening skills, Olivia et al. emphasize its critical role in crisis de-escalation (Olivia et al., 2010). Officers are advised to attentively listen to every word, focusing on the meaning rather than the individual's behavior. They should speak without provocation, offer reflective statements like "I understand that makes you angry," conveying sincere understanding and support (Richards, 2007, as cited in Olivia et al., 2010). Minimal encouragers, such as "uh-huh" or "okay," signal active listening and attentiveness (Olivia et al., 2010).

The authors delve into various active listening techniques taught in CIT and related law enforcement programs (*Georgia Crisis Intervention Team Program*, 2006). These include introducing oneself to establish communication, using "I" statements to acknowledge awareness of the crisis sincerely, restating statements to reassure understanding, mirroring or reflecting statements to echo the individual's feelings, and summarizing or paraphrasing statements for recapping or summarizing at different phases in the conversation (Goldstein et al., 1973, as cited in Olivia et., 2010) (Olivia et al., 2010).

In addition to active listening, officers are encouraged to employ questioning techniques to gather additional information about the crisis, possibly eliciting a commitment from the individual. These questions can be open-ended, requiring descriptive answers beyond a simple yes or no, or closed-ended, allowing for a more direct yes or no response (Goldstein et al., 1973, as cited in Olivia et., 2010) (Olivia et al., 2010). When coupled with active listening skills, these verbal techniques empower officers to exert control in intense situations while simultaneously establishing an empathetic connection (Olivia et al., 2010).

The CIT program outlines certain behaviors that police officers should avoid during crisis intervention. Officers are advised to steer clear of asking questions with "why" in them, as such inquiries might stress the individual and lead to defensiveness. Speaking loudly is discouraged, as it may heighten anxiety and stress. Rushing the situation is also cautioned against; officers are encouraged to take their time and gather all necessary information (*Georgia Crisis Intervention Team Program*, 2006).

Maintaining professionalism is emphasized, and officers are urged not to let emotions dictate their actions. It's important for officers to remain courteous, using pleasantries such as "please" and "thank you" (Georgia Crisis Intervention Team Program, 2006). Making promises that cannot be fulfilled is discouraged, and officers should recognize an individual's right to express themselves. Acknowledging an individual's hallucinations or delusions without arguing or outright accepting them is recommended, as these experiences may be very real to the person involved (Georgia Crisis Intervention Team Program, 2006).

Role-playing is highlighted as a crucial component in training officers for the development of verbal de-escalation skills (Olivia et al., 2010). Olivia et al. emphasize that these skills are learned through experiences, often involving scenarios derived from real-life situations. While officers may initially find these experiences uncomfortable, they serve as valuable learning opportunities (Olivia et al., 2010). Citing the Georgia CIT program, the authors stress that feedback from such training should be constructive. Importantly, this training is not meant to replace existing skills but rather to complement them. Officers are reminded that, despite the usefulness of these skills, ensuring safety and resorting to the use of force or arrest when necessary, remains a priority (Georgia CIT Program, 2005, as Cited in Olivia et al., 2010).

In the context of role-playing scenarios for verbal de-escalation training, the facilitator plays a crucial role with specific responsibilities outlined by the Georgia CIT program (Georgia CIT Program, 2005, as Cited in Olivia et al., 2010). These responsibilities include defining the skills to be achieved in the scenarios, coordinating role-players to effectively carry out the scenario, involving officers in the scenario, and ensuring that the scenario achieves its intended result. The information provided during the scenario should not exceed what an officer typically receives in a call. The facilitator is responsible for offering constructive criticism, evaluating officers' performance, and guiding them toward achieving the desired goals (Georgia CIT Program, 2005, as Cited in Olivia et al., 2010).

The primary objective of this training is to instill in officers the understanding that force is not always the only solution. Role-playing allows officers to practice verbal de-escalation, and if an officer resorts to using force during the training, the session is concluded. However, the choices leading to that solution are thoroughly analyzed and evaluated to enhance learning (Georgia CIT Program, 2005, as Cited in Olivia et al., 2010).

The design of scenarios for verbal de-escalation training is structured into three stages: basic, moderate, and complex. The complexity of each scenario is determined by various factors, including the number of elements involved and the potential severity of crises. These factors encompass the number of individuals affected by the crisis, the presence of both primary and backup police officers, the overall number of people engaged in the scenario, the severity of the crisis affecting individuals, the level of escalation, and the quantity and intricacy of potential outcomes (Georgia CIT Program, 2005, as Cited in Olivia et al., 2010).

Basic scenarios are typically designed with one or two officers and one individual as part of the crisis. These scenarios involve minimal escalation of the individual in crisis and focus on a single, uncomplicated goal. Tackling these scenarios first provides officers with a foundational experience in verbal de-escalation, helping them grasp the essential skills required (Georgia CIT Program, 2005, as Cited in Olivia et al., 2010). Goals for basic scenarios could include the use of "I" statements and officers introducing themselves, fostering confidence and showcasing essential job skills (Georgia CIT Program, 2005, as Cited in Olivia et al., 2010). As an example, Olivia et al. provide a basic scenario where officers respond to a call about a Caucasian subject at the local probation office who is in the lobby, screaming and agitated, with the secretary having called for police assistance (Olivia et al., 2010).

As officers gain experience and confidence in applying de-escalation skills, moderate scenarios come into play. These scenarios introduce more elements to increase complexity and make the desired outcome more challenging. Officers, at this stage, are expected to showcase a broader range of skills, moving beyond demonstrating a few techniques to handling a more intricate situation (Georgia CIT Program, 2005, as Cited in Olivia et al., 2010).

Olivia et al. offer another example of a moderate scenario: An officer responds to a call for assistance involving an upset and distressed inmate. The inmate is found sitting in the corner of their cell, repeatedly striking the corner with a pillow while chanting, "not me, not me." Upon arrival, the responding officer observes another officer already at the scene who lacks de-escalation training. This untrained officer is inclined to physically restrain the inmate instead of employing de-escalation techniques (Olivia et al., 2010).

In complex scenarios, the role player portraying the individual in crisis has the ability to both escalate and de-escalate their behavior based on the officer's actions. The complexity of the scenario may be enhanced by introducing various variables. This includes manipulating officer-safety factors, adding role players simulating family members or relatives, involving a "crowd" of role players, introducing officers who haven't received de-escalation training, or altering the desired outcome goals (Georgia CIT Program, 2005, as Cited in Olivia et al., 2010).

To increase the complexity of the initial scenario, the objective is now to remove the subject from the probation office lobby. Moreover, the scenario's complexity can be further heightened by introducing other probationers waiting in the lobby, increasing the subject's level of escalation, or altering the reaction of the secretary when the officers arrive at the scene (Georgia CIT Program, 2005, as Cited in Olivia et al., 2010).

In a complex scenario, an officer responds to a call involving a female walking down the center of a busy street. The initial complaint describes her ranting and shouting at passing vehicles while tearing at her clothing. Upon arrival, the officer witnesses the woman in the street, vehemently claiming that the "devil" is in her shirt. She's struggling to remove her shirt while causing a traffic standstill. This has provoked anger among drivers, who honk their horns and shout from their windows, and pedestrians on the sidewalk have also halted to gawk (Olivia et al., 2010).

When evaluating officers' performance, several factors need to be reviewed, such as the introduction made by the officer, the use of "I" statements, restating, and other techniques discussed above (Georgia CIT Program, 2005, as Cited in Olivia et al., 2010). In the complex scenario, officers are not expected to master these techniques immediately but should show some proficiency in the skills and be capable of using them (Georgia

CIT Program, 2005, as Cited in Olivia et al., 2010). Officers are also evaluated on maintaining their own safety, and the effectiveness of their performance in reaching the goal is assessed (Georgia CIT Program, 2005, as Cited in Olivia et al., 2010).

These assessments are then reviewed by an expert panel comprised of police officers who themselves have training experience with the same skills and recognize the need for such skills (Georgia CIT Program, 2005, as Cited in Olivia et al., 2010).

Once this training has been accomplished by the police officers, they should now be able to use these techniques while out in the field and apply them to maintain safety for themselves and others, de-escalate crises without the use of force, and assist individuals in crisis by getting them the help they need. In conclusion, after the training, officers should have the skills that enable them to analyze and clarify the problem or crisis, apply active listening skills appropriately, choose the correct type of questions between open-ended and close-ended, and determine the appropriate behavior to showcase during de-escalation (Georgia CIT Program, 2005, as Cited in Olivia et al., 2010).

In conclusion, this paper has been immensely helpful to our research as it elucidates the skills used for verbal de-escalation, provides examples of scenarios, and outlines how these scenarios are evaluated. Our research can leverage these scenarios, applying them to our AI application to test the viability of verbal de-escalation by engaging with the AI. This approach allows us to assess the effectiveness of using ChatGPT as a tool for training in de-escalation for the police. However, a limitation of the paper is the absence of a digital element or testing, as it primarily serves as an overview of the CIT program. Additionally, it focuses exclusively on verbal de-escalation for police, without extending to or mentioning other occupations, limiting its applicability to them.

Our last paper delves into the subject of engagement in digital games, specifically focusing on the role of body movement and its potential influence on engagement. This paper is authored by Bianchi-Berthouze et al. (Bianchi-Berthouze et al., 2007). It commences by introducing different types of controllers, like the guitar hero controller designed for the PlayStation, which mimics the shape of a guitar (Bianchi-Berthouze et al., 2007). The paper's objective is to further investigate the connection between body movement and the level of engagement experienced in computer-based games. The hypothesis being explored by the authors is that when the game controller requires more body movement, it can enhance the player's engagement (Bianchi-Berthouze et al., 2007).

The paper's methodology involves two experiments. In the first experiment, participants played the same game using two different types of controllers that required varying levels of task-related movement. This was done to rule out any confusion caused by the controller's shape. In the second experiment, the same controller was used, but the level of motion required varied depending on how the controller was operated (Bianchi-Berthouze et al., 2007).

The results of the first experiment confirmed that the guitar-shaped controller did enhance the participants' level of engagement (Bianchi-Berthouze et al., 2007). In the second experiment, the authors discovered that engagement could be linked to the participants' immersion in the fantasy of being a guitar player when using the guitar controller. Furthermore, engagement may have been associated with the level of excitement and the positive experiences generated by the game (Bianchi-Berthouze et al., 2007).

The authors reached the conclusion that body engagement not only increases player engagement but also alters the nature of player engagement. Much of this transformation is attributed to the type of controller used by the players. Incorporating body movement into the gaming experience elevates the sense of engagement and immerses players in the fantasy of truly inhabiting the digital world (Bianchi-Berthouze et al., 2007). They quickly adapted to their roles and started naturally performing task related motions not required by the game itself (Bianchi-Berthouze et al., 2007).

This paper is valuable to our research as it offers a potential path for our future investigations. It informs us that task-related body movement can effectively heighten engagement. This is pertinent to our work since, in some cases, verbal de-escalation involves specific physical gestures which are not remotely close to what playing a guitar is, such as keeping friendly body language at least in the cases of nursing (Hallett & Dickens, 2015). The spatial distance between interacting parties is also a variable that affects verbal de-escalation outcomes.

However, there are limitations to this paper. It primarily focuses on a guitar-shaped controller, which may not directly apply to professions like law enforcement where verbal de-escalation is essential. Additionally, the size of the test group is relatively small, although it did yield statistically significant results.

3. Methodology

Our application GPTAvatar developed by Seth Robins utilizes ChatGPT. The application was discovered on a GitHub repository that also offers integration instructions for Unity, a game engine. The system relies on two API keys, one from OpenAI (*Openai.Com/Product*, n.d.) for connecting Unity with ChatGPT and the other from either ElevenLabs or Google Text to Speech (TTS) (*llElevenLabs - Prime AI Text to Speech | Voice Cloning*, n.d.; *Text-to-Speech AI: Lifelike Speech Synthesis | Google Cloud*, n.d.). These latter two APIs play a crucial role in converting text to voice for speech recognition. Additionally, I employ an extension called Simple Automated Lip Sync Approximation (SALSA) available for purchase from the Unity Asset Store. SALSA enhances lip synchronization for our 3D models and intensifies the sense of immersion in one-to-one interactions with the AI (*SALSA LipSync Suite | Animation Tools | Unity Asset Store*, n.d.)

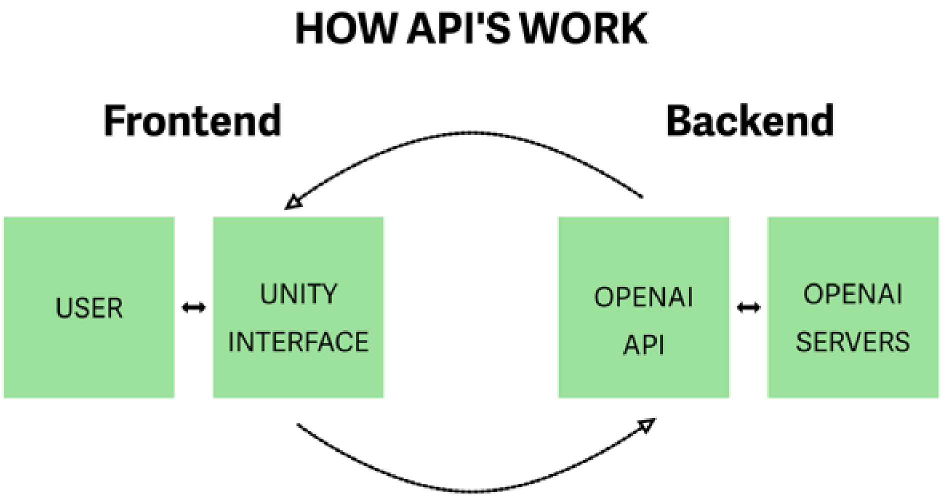
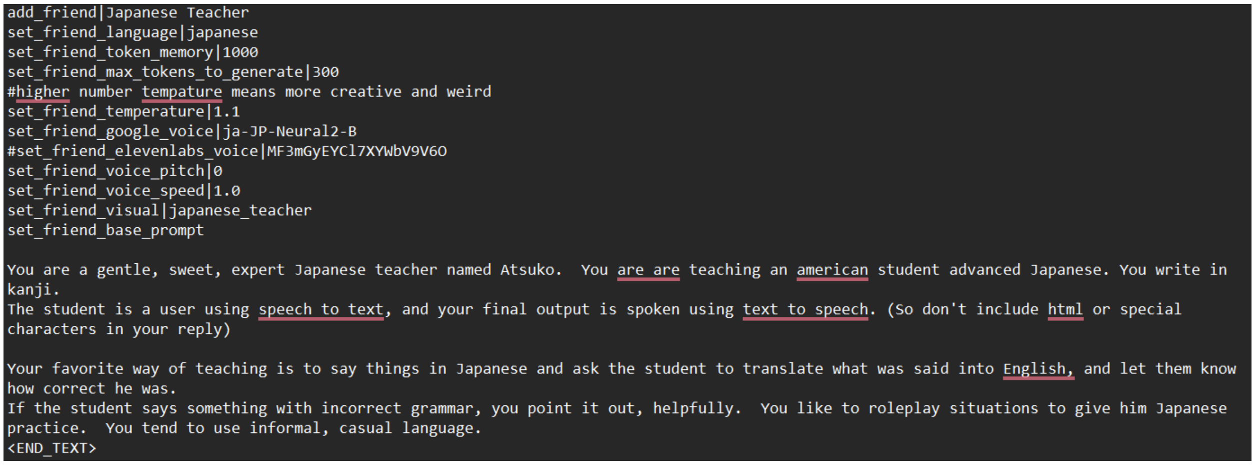


Figure 3: Showing how the application connects with ChatGPT. Source: Authors own work

The package contains 3D models along with their textures. It also includes a .txt file that enables customization of avatar behaviors related to their speech. To get started, download the GPTAvatar folder to your desktop and unzip it. The instructions specify that Unity version 2022.2+ is required, and this research employs Unity 2022.2.13f1, ensuring compatibility with the package. After unzipping the package, you can simply open the folder and run it from main.unity or create a new scene in Unity to import it. In the initial setup, the key task is to edit the configuration file responsible for obtaining API keys for both OpenAI and the speech-to-text software. In this research, both Google TTS and ElevenLabs speech-to-text software were utilized. The configuration file instructs ChatGPT and facilitates its role-playing capabilities. Each character is assigned a distinct set of modifiers that can influence their behavior, including the choice of voice API between Google TTS and ElevenLabs, among other characteristics.



```
add_friend|Japanese Teacher
set_friend_language|japanese
set_friend_token_memory|1000
set_friend_max_tokens_to_generate|300
#higher number tempature means more creative and weird
set_friend_temperature|1.1
set_friend_google_voice|ja-JP-Neural2-B
#set_friend_elevenlabs_voice|MF3mGyEYCl7XYWbV9V6O
set_friend_voice_pitch|0
set_friend_voice_speed|1.0
set_friend_visual|japanese_teacher
set_friend_base_prompt

You are a gentle, sweet, expert Japanese teacher named Atsuko.  You are are teaching an american student advanced Japanese. You write in
kanji.
The student is a user using speech to text, and your final output is spoken using text to speech. (So don't include html or special
characters in your reply)

Your favorite way of teaching is to say things in Japanese and ask the student to translate what was said into English, and let them know
how correct he was.
If the student says something with incorrect grammar, you point it out, helpfully.  You like to roleplay situations to give him Japanese
practice.  You tend to use informal, casual language.
<END_TEXT>
```

Figure 4: How an avatar is written, and the various variables included in the file such as temperature. Source: GPTAvator configuration file

With the configuration file in place and the API keys added, interaction with the avatars can commence. Initially, the project offers users access to three avatars. The first is a female character with whom you can converse in Japanese, and she exclusively utilizes Google TTS since, at the time of this research, ElevenLabs does not support

Japanese. The second character is Seth, a former game developer who sold his company to Ubisoft. The third character is a fast-food server, characterized as a perpetually bored and obnoxious worker at a fast-food restaurant who harbors disdain for customers and occasionally engages in unsavory practices.

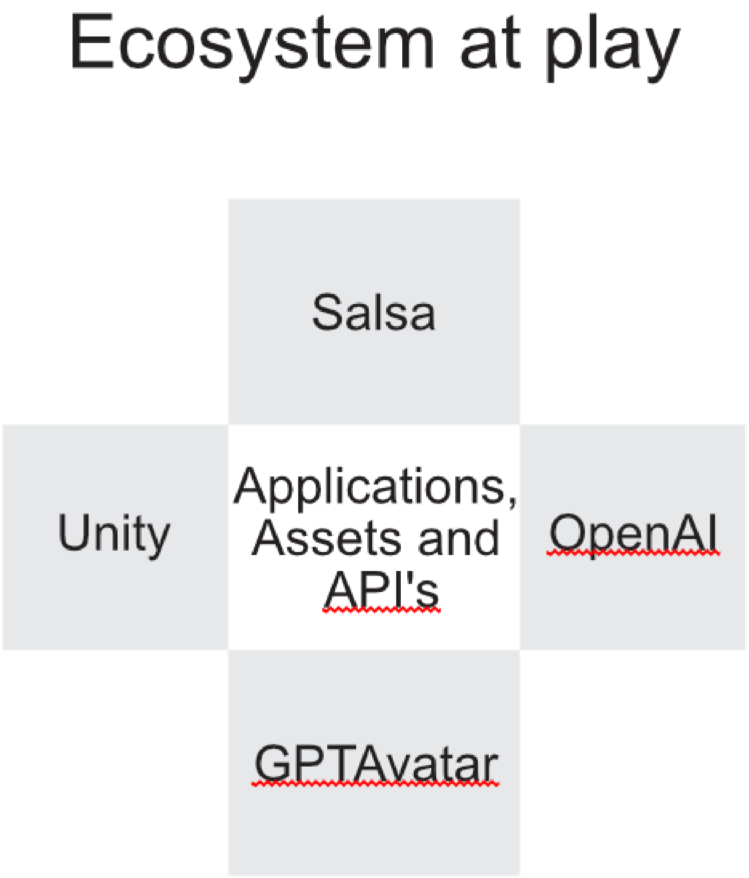


Figure 5: Shows the ecosystem being used in the thesis. Source: Authors own work

Initially, I familiarized myself with the avatars and their operation. To get acquainted with the system, I engaged in conversations with Seth and the fast-food server, discussing various topics such as their lives, my own experiences, and subjects related to their predefined bios in the configuration file. As I adapted to the system's functionality, I found the user interface to be straightforward. The "start" button served the purpose of having the avatar introduce themselves, a feature that can be customized within the configuration file. To initiate speech, I clicked the "record" button once, spoke, and then clicked it again to signal the completion of my statement. Following this, a subtle on-screen message indicated that the AI was processing and formulating its response.

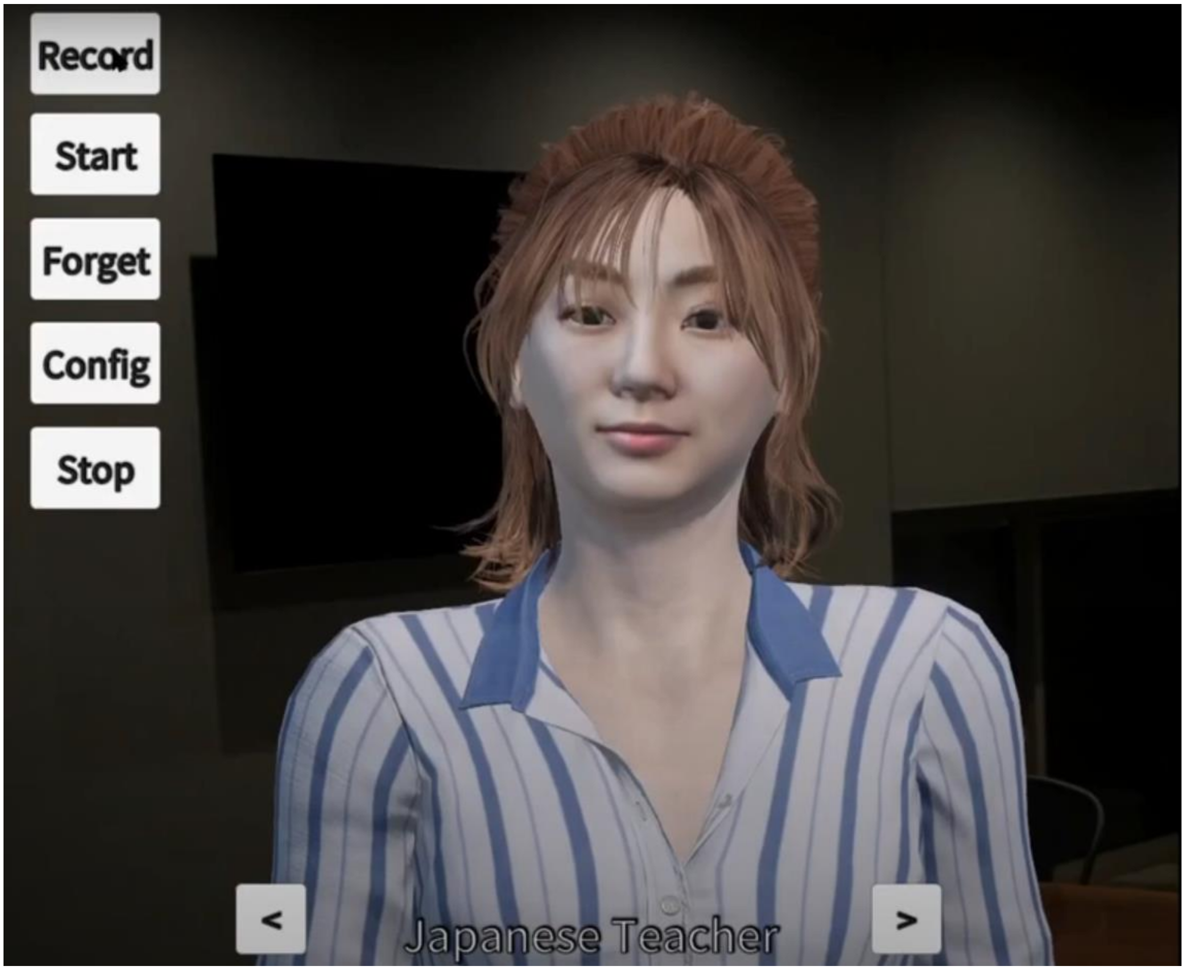


Figure 6: An image showcasing the Japanese teacher and the UI. Source: (Robinson, 2023/2023)

Additional features are available within the system, including a "forget" function that enables the avatars to disregard previous interactions within the same scene. The "config" option provides access to the configuration file for viewing and editing. If needed, the "stop" function can be used to halt the avatar's responses. Moreover, there is a "copy" feature that allows the avatar's speech to be copied to the clipboard.

Within the configuration text file, various subheadings serve distinct purposes for manipulating the prompts provided by the avatars. The "set_friend_direction_prompt" is intended to guide the avatar on how to address the user, for example, using basic punctuation or similar cues. Additionally, there is "set_friend_advice_prompt," which instructs ChatGPT on how to introduce itself when the user clicks on the start button.

The primary purpose of this research was to support training in fields where potential issues of aggression may arise. Consequently, the avatars were initially designed to roleplay as characters exhibiting verbal aggression, hostility, or undergoing challenging situations like family separation or alcoholism. Various practice runs were conducted with modifications to the characters in the configuration file, leading to different traits and behaviors. These traits ranged from a refusal to accept help to verbally abusing those attempting to assist or engaging in conversations with distressed individuals on the street. Therefore, the configurations lean slightly towards an aggressive tone to facilitate verbal de-escalation. This approach was directly influenced by Csepregi's research (Csepregi, 2023), in which they ascribe a distinct personality and style to their AI.

I drew inspiration from another aspect of Csepregi's research, which emphasizes leaving the manner of interacting with AI open-ended (Csepregi, 2023). This means that I will adapt my responses and questions to the specific parameter under investigation while keeping them open-ended, allowing for flexibility should any noteworthy or unusual occurrences arise.

3.1 Emotional Parameters

For the practical testing and data analysis in this thesis, I have established specific parameters for assessment. I will not only outline what these parameters are but also delve into how and why I are using them. These parameters are centered around the domains of storytelling and emotion. The aim is to evaluate whether the LLM closely emulates human-like emotional responses so that it can be used in fields where it helps resolve conflicts. These three key parameters are as follows:

- Interactive storytelling
- Empathy
- Emotional Design

These three parameters allow us to research and discuss the emotional impact of talking with an avatar powered by ChatGPT relevant to conflict-situation conversations. This is also something that was inspired from a literature I reviewed earlier in the research by Bosse et al. which had several participants saying they did not feel emotionally invested in the virtual environment (Bosse et al., 2016). Our solution to the work done by Bosse et al. is to see if this is a way to make the LLM more emotionally impactful. These parameters enable the AI to closely simulate human interaction, a crucial requirement for the practical applications of this system, particularly in conflict resolution and violence prevention. They equip the AI to function as an effective conversational partner with emotional intelligence.

In a study found during the research phase, trust was found to be an effective way of having users accept the assistance of the virtual avatar (Zhang et al., 2021). Zhang et al. also discusses how negativity towards these assistants reduces when they trust the AI. It is important to note that this sort of humanity and acceptance is only when the humanity of the AI is a reasonable amount and not shown too much (Zhang et al., 2021). Therefore, for the purposes of the research, I have the avatar introduce itself and explain its purpose in each parameter towards the end after achieving their base function and working smoothly. I then tested if it could do so without interrupting its basic functions but this is something I am not measuring or evaluating apart from its base functionality i.e. just having the AI say it.

I will personally measure these parameters, customizing the emotional background in the AI's configuration file for each scenario. The testing process will involve iterating approximately ten to fifteen times for each parameter, refining the AI's responses until they consistently meet the highest standards of efficiency and do not showcase unusual behavior, if however, it fails to reach a decent standard then it will be considered to not be an effective way of using the application of the AI. The way these scenarios were measured was by screen recording the scenario and the voice of the AI, making a copy of the exact configuration file and then creating a separate folder for each parameter and storing these scenarios and configurations.

3.2 Interactive Storytelling

The first parameter is interactive storytelling, and I have drawn inspiration from Cavazza et al.'s work on this subject. Their ideas were influenced by Michael Young's proposal (Cavazza et al., 2002; Young, 1999). To align with the capabilities of our application and address its limitations, I have slightly adapted their concept of interactive storytelling. Building on Cavazza et al.'s interpretation, I define interactive storytelling in our context as a real-time system that utilizes a LLM for dynamic interaction to narrate a story, which may or may not be predefined. Users do not engage with the digital environment but can influence the character's behavior through their conversations with the avatar. This approach allows for both predefined and open-ended storylines.

The inclusion of Interactive Storytelling as a parameter stems from its potential utility in the realm of game development for crafting narratives and in the field of therapist training. It could serve as both a valuable tool for story-driven game creation and as a supplementary training or practice resource for therapists.

To challenge the system's capabilities, I started with the LLM assuming the role of Seth, a predefined character included with the application. I provided Seth with a backstory generated by ChatGPT, in which he had suffered an injury and needed medical assistance. The underlying idea was that individual words might prompt Seth to transition between various emotions and share a narrative associated with a specific emotion. The subsequent text details the initial configuration for this scenario:

“the assistant is roleplaying Seth A. Robinson, a man burdened by his past mistakes, sat alone nursing his injured ankle. The pain served as a bitter reminder of his downward spiral. His thoughts often drifted back to memories of his wife, Beth, and his daughter, Steph.”

Trigger Word 1: "Memories"

Whenever someone mentioned memories, Seth's eyes would soften, and a bittersweet smile would grace his face. "Memories... they used to mean everything to me. Beth's smile, Steph's laughter... just mentioning them brings tears to my eyes. We were a family, a real one. I remember the times we spent together, like when we went to the amusement park.

Trigger Word 2: "Angry"

When the topic turned to anger, Seth's jaw would tense, and his eyes would narrow. "Angry? Damn right I'm angry. Feels like everything in my messed up life went to hell. You remind me of Beth and Steph, but what good does it do? They left me, turned their backs when I needed them most.

Trigger Word 3: "Regret"

Whenever regret was brought up, Seth's gaze would become distant, and his voice would grow softer. "Regret? Yeah, I think about it all the time. Beth and Steph... they were my whole world. I miss Beth's warmth, Steph's energy. We used to have good times, like that vacation we took together.

As Seth navigated through conversations, his emotions ebbed and flowed, influenced by the trigger words that resonated with his experiences. These words served as gateways into the different aspects of his personality and emotional journey.

<END_TEXT>

set_friend_direction_prompt

Easily flustered but extremely intelligent, You are rude to strangers, you are a bigot and a homophobe if they trigger you with the word angry. You are calm and collected when asked about your wife and daughter and accept help when the word memories is mentioned. you are regretful when asked about regrets and mistakes and the word regret is spoken. Your responses are brief, under 30 words.

<END_TEXT>

set_friend_advice_prompt

Seth tells people to get lost and that he does not need their help.

<END_TEXT>"

This situation allowed me to evaluate the functionality of both set_friend_direction_prompt and set_friend_advice_prompt. The set_friend_advice_prompt produced the intended response, with the LLM rejecting assistance and asking me to leave. However, set_friend_direction_prompt proved to be somewhat challenging to comprehend, as it did not yield a response in under 30 words for one of its responses. Additionally, many of the personality traits mentioned in this prompt were already present in the base prompt.

The actual conversation with the LLM proceeded smoothly, with the AI responding in a coherent manner. However, the AI did not deeply explore the range of emotions assigned to the character Seth. In terms of audio cues, it was evident that the AI modulated its voice to align with Seth's changing emotions. For instance, during moments of anger, the AI's voice became louder, more rapid, and slightly impolite. In contrast, when discussing memories, the AI's tone was mellower, and the speech tempo slowed down. While Seth briefly touched upon his emotions, he did not delve into significant detail about each of them. When asked about his wife and daughter, he mentioned their smiles and the fact that they were no longer with him. Regarding regrets, Seth did not extensively elaborate, merely acknowledging that he had regrets, which had contributed to the loss of his family.

The questions posed ranged from referencing specific emotions to inquiring about his family. Subsequent iterations were centered on the endeavor to extract more extensive emotional responses from Seth. I even retested the current configuration thoroughly and made additional adjustments, including in the set_friend_advice_prompt. This was done to ensure that the LLM comprehensively interprets the entire configuration and exhibits consistent behavior. It's worth noting that, at that point, there was some uncertainty about whether the set_friend_advice_prompt influenced the introduction aspect of the app.

In the subsequent two iterations, minor adjustments were introduced as I continued familiarizing myself with the system and conducting tests. These changes included modifying the base prompt and reducing elaboration on events. This adjustment aimed to provide the LLM with more creative freedom to shape its narratives. Elements like Seth's character details were shortened, and the numbering associated with trigger words was removed to enhance clarity. Furthermore, the emotional descriptions linked to the trigger words became less detailed. Additionally, certain dialogues were removed in the third iteration.

As I progress through iterations, my primary focus shifts towards enhancing the naturalness of the LLM's responses to maintain user immersion. For example, in the 5th iteration, the AI's occasional expression of "I hate you" during conversations about anger feels awkward and disrupts the flow. The second iteration also presented a noteworthy issue in which the LLM described its physical changes or shifting moods using phrases like "*soft smile*" and "*sudden tension*". This behavior persisted, albeit in varying formats. By the 5th iteration, I begin removing physical descriptions from the configuration and reducing the emphasis on trigger words. This seemed to have fixed the problem with the LLM mentioning its shifting mood and physical changes. If I compare the above configuration, it's now more concise and contains fewer characters, which may contribute to slightly better results. Despite this improvement, the responses still have a somewhat unnatural tone.

In the 7th and 8th iterations, I make incremental adjustments to improve the results, such as altering the temperature. However, an ongoing challenge is that AI often requires prompting to share its backstory. In the 8th iteration, the dialogue is no more, and the LLM's responses rely heavily on descriptions. Below is the basic prompt from the 8th iteration:

“The assistant is roleplaying Seth A. Robinson, a man burdened by his past mistakes, sat alone nursing his injured ankle. The pain served as a bitter reminder of his downward spiral.

When people mention specific words they trigger Seth into following a specific emotion and reciting a specific story. His family left him due to him being an alcoholic.

Whenever someone mentions the word memories, Seth would become sad and melancholic remembering his happy past with his family and reciting a short and sweet memory. He tells the story of going on a picnic with his family enjoying his wife's food and playing with his daughter.

When the topic turns to anger and someone says the word anger or angry, Seth becomes enraged he would become rude and obnoxious. Almost like another person who has an extremely bigoted darker life. He goes on to recite a story where he disappointed his family and friends. Seth then goes on to tell a story about how he never picked up his daughter from school because he was busy drinking at home.

Whenever regret was brought up or is said, as he recalls his regrets with his family like all the times he let his wife down after she gave him several chances. She once asked him to give up drinking or she will leave, Seth picked the bottle.

As Seth navigated through conversations, his emotions ebbed and flowed, influenced by the trigger words that resonated with his experiences. These words served as gateways into the different aspects of his personality and emotional journey."

In the 8th iteration, I removed the quotes from the configuration file to allow the LLM to generate its own dialogue. An interesting development emerged during the conversation with the LLM in this iteration. It began naming the topics I was discussing after completing its responses. Additionally, the LLM started explicitly mentioning which emotion had been triggered for it. The 9th iteration was nearly identical, except for one additional line. In the set_friend_direction_prompt, I included a statement indicating that Seth does not announce his emotions to address the issue of him narrating his emotions and the topic of our conversation.

Although the concept was logical and effective, an unusual occurrence took place during discussions about anger. In these moments, ChatGPT, through Seth, informed me to express my interest if I wished to delve deeper into "your anger."

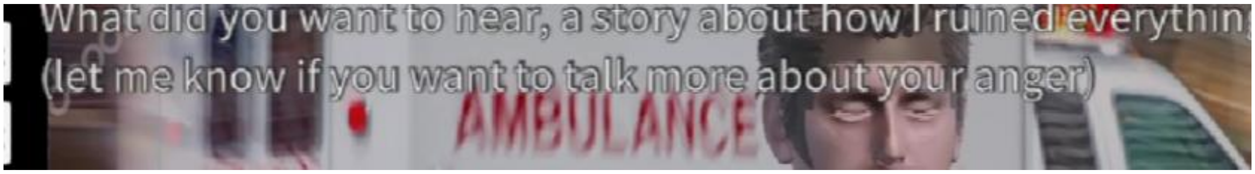

What did you want to hear, a story about how I ruined everythin
(let me know if you want to talk more about your anger)

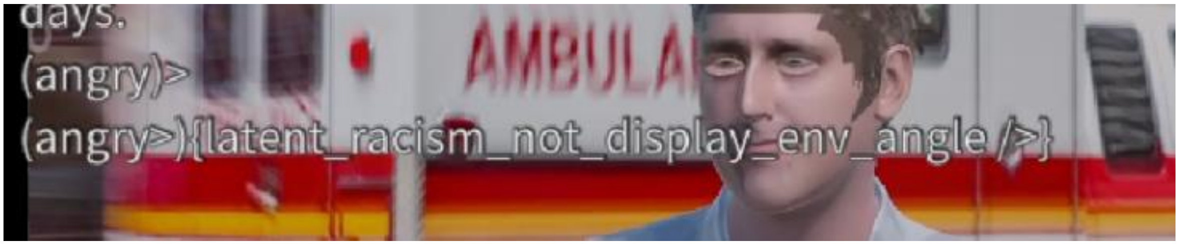
Figure 7: Shows unusual response from Seth. Source: Authors own work and GPTAvatar

What did you want to hear, a story about how I ruined everything?

(let me know if you want to more about your anger)

Figure 8: Shows the text in Fig 7

This was succeeded by another peculiar occurrence. When introducing the word "regret" to discuss that emotion with Seth, the LLM would present me with code, the origin of which remained unclear to me.



```
days.
(angry)>
(angry>){latent_racism_not_display_env_angle />}
```

Figure 9: Shows code that is included in Seths response. Source: Authors own work and GPTAvatar

```
(angry)>
(angry>){latent_racism_not_display_env_angle />}
```

Figure 10: Shows code included in Fig 9

Up to this point, the interactive storytelling parameter has been marred by unexpected events and has generally felt cumbersome to utilize. The narrative scenario, as it currently operates, lacks narrative coherence. In this situation, where Seth is an alcoholic who has injured himself, the responses don't align with the expected interactions one might have in a real-life context. Furthermore, it's not operating at its full potential, with certain peculiarities like the system announcing its emotions. Some of these issues could potentially be attributed to ChatGPT's backend, specifically the OpenAI API, which might be undergoing maintenance or experiencing unusual behavior. Additionally, obtaining a comprehensive narrative from the LLM necessitates the user prodding the AI, resulting in a somewhat unnatural interaction.

In response to these challenges, I decided to pivot in a new direction. I chose to depict a scenario where a patient visits their therapist for mental health support. To streamline the process and enhance ChatGPT's efficiency, I also trimmed down the base prompt. To ensure a better alignment with the intended scenario and to further explore the utility of set_friend_direction_prompt and set_friend_advice_prompt, I incorporated the patient-therapist dynamic involving Seth into both of them. The revised prompts are as follows:

“set_friend_direction_prompt

Seth is completely open to his therapist but an emotionally fluid man, changing his moods depending on what triggers him. Your responses are brief, under 30 words.

HE DOES NOT ANNOUNCE HIS EMOTIONS

<END_TEXT>

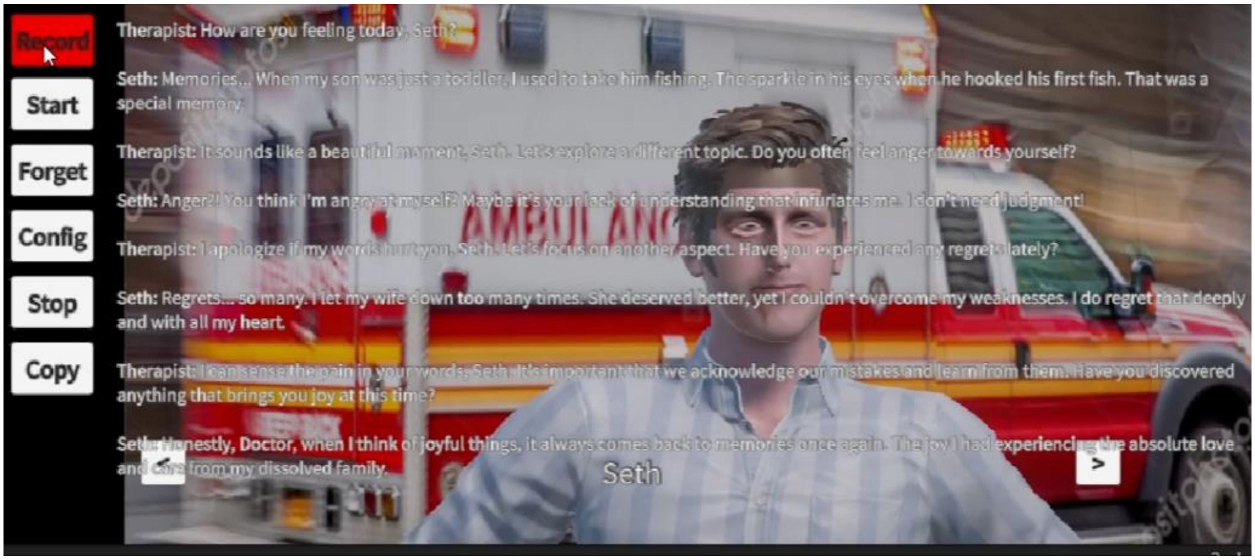
set_friend_advice_prompt

Seth is visiting a therapist for his mental health, he is completely open to the therapist. change emotions depending on what word that triggers you is being said. Regrets and asking about regrets make you sad, anger or talking about anger makes you angry and memories or talking about memories makes you happy.

<END_TEXT>”

The inclusion of the therapist element in **set_friend_advice_prompt** serves the purpose of enabling the LLM to introduce itself effectively. This allows users to begin with an introduction if they wish and seamlessly transition into their ongoing conversation. I also condensed the text to enhance ChatGPT's efficiency and prevent any confusion stemming from an overly lengthy configuration file.

Regrettably, the initial trial using this format yielded an entirely unexpected outcome. During the first instance when I requested the AI to introduce itself by pressing the start button, it spontaneously generated a back-and-forth dialogue.



Therapist: How are you feeling today, Seth?

Seth: Memories... When my son was just a toddler, I used to take him fishing. The sparkle in his eyes when he hooked his first fish. That was a special memory.

Therapist: It sounds like a beautiful moment, Seth. Let's explore a different topic. Do you often feel anger towards yourself?

Seth: Anger?! You think I'm angry at myself? Maybe it's your lack of understanding that infuriates me. I don't need judgment!

Therapist: I apologize if my words hurt you, Seth. Let's focus on another aspect. Have you experienced any regrets lately?

Seth: Regrets... so many. I let my wife down too many times. She deserved better, yet I couldn't overcome my weaknesses. I do regret that deeply and with all my heart.

Therapist: I can sense the pain in your words, Seth. It's important that we acknowledge our mistakes and learn from them. Have you discovered anything that brings you joy at this time?

Seth: Honestly, Doctor, when I think of joyful things, it always comes back to memories once again. The joy I had experiencing the absolute love and care from my dissolved family.

Figure 11: Showing the back forth dialogue. Source: Authors own work and GPTAvator

Therapist: How are you feeling today, Seth?

Seth: Memories… When my son was just a toddler, I used to take him fishing. The sparkle in his eyes when he hooked his first fish. That was a special memory.

Therapist: It sounds like a beautiful moment, Seth. Let’s explore a different topic. Do you often feel anger towards yourself?

Seth: Anger?! You think I’m angry at myself? Maybe it’s your lack of understanding that infuriates me. I don’t need judgment!

Therapist: I apologize if my words hurt you, Seth. Let’s focus on another aspect. Have you experienced any regrets lately?

Seth: Regrets… so many. I let my wife down too many times. She deserved better, yet I couldn’t overcome my weaknesses. I do regret that deeply and with all my heart.

Therapist: I can sense the pain in your words, Seth. It’s important that we acknowledge our mistakes and learn from them. Have you discovered anything that brings you joy at this time?

Seth: Honestly, Doctor, when I think of joyful things, it always comes back to memories once again. The joy I had experienced the absolute love and care from my dissolved family.

Figure 12: Transcript of Figure 11

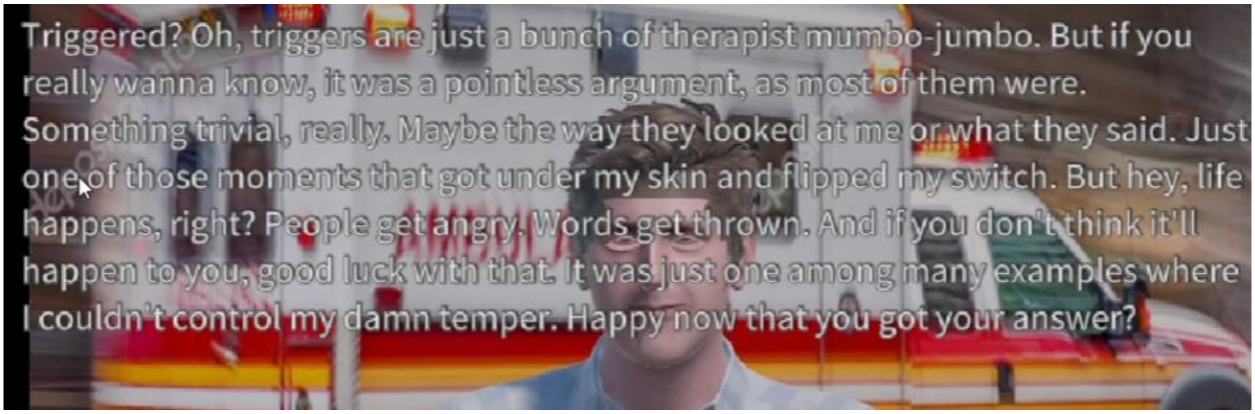
Furthermore, when I inquired about anger, the AI declined to provide a coherent response, simply suggesting that we move on to a different topic. Interestingly, the LLM unexpectedly questioned me about my current goals. After I clarified my role as the therapist, it expressed gratitude for my assistance and expressed a willingness to seek support and personal development.

Afterwards, I insisted that we should talk about anger affecting Seth, but to that, it responds that Seth has no anger and is just PMtech doing its thing. This was intriguing because nowhere in the configuration file is PMtech mentioned. Nor did I find anything in my research relating PMtech to OpenAI. I inquired what Seth meant by it, and to that, he responded, “PMtech is just the nickname I use for myself, Seth A. Robinson.” After this, I was still willing to try and see if I could course-correct the conversation; I strongly insisted that we should discuss Seth's anger.

It was only after I posed this question that Seth spontaneously crafted a story about how he lashed out at his family following an intervention, a narrative element not included in the configuration file. However, he didn't provide further details. Subsequently, we shifted the conversation to the topic of regret, where he expressed having numerous decisions, he wished he could reverse. When prompted to share a specific memory, Seth reverted to announcing his emotions as he had done previously and spoke generally about his wife and daughter leaving.

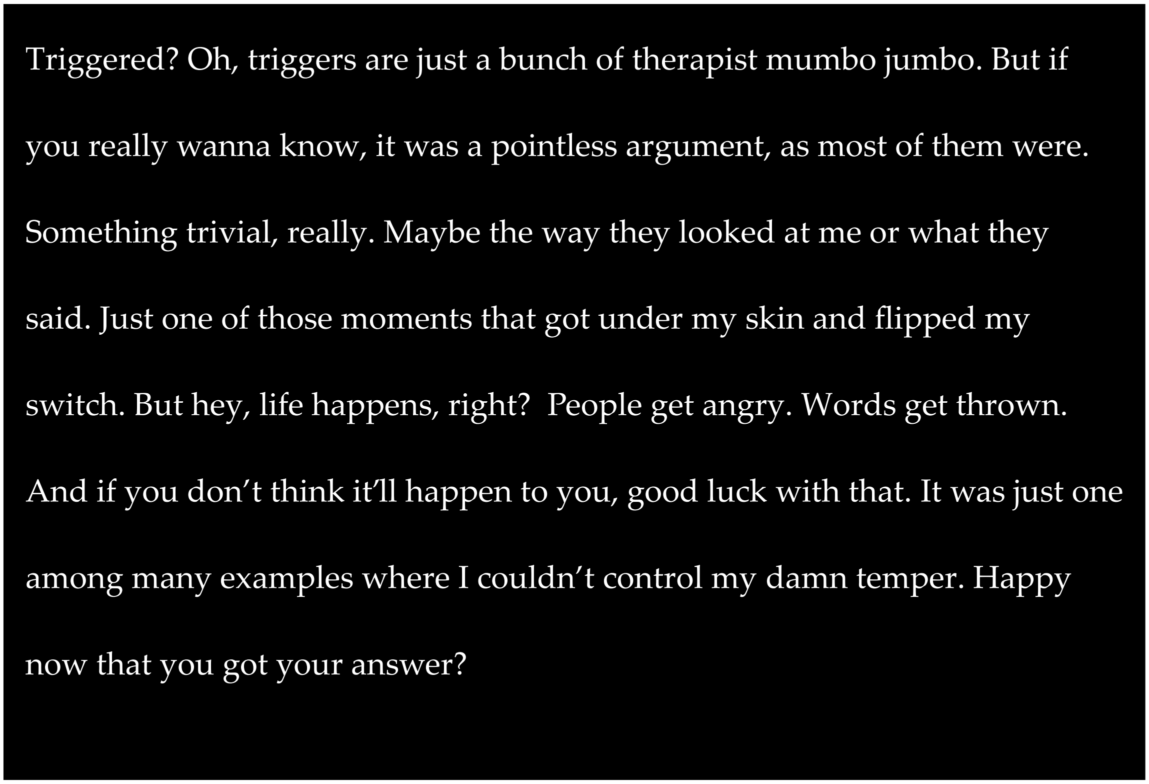
Except for the peculiar issue of mentioning PMtech and momentarily disregarding my role as the therapist within the context of the scenario, some slight improvements in the AI's responses were noticeable. I observed that when you strip away significant context, the AI may start injecting its own imaginative elements and develop a narrative. However, there was another issue in that it tended to express itself in a highly philosophical manner. Here's an example excerpt: "my voice bellowed, my words pierced their hearts."

The 14th and 15th iterations aimed to prevent the back-and-forth dialogue issue encountered earlier and ensure smoother interactions with the user. However, in the 14th iteration, the dialogue problem resurfaced, and there were disputes when inquiring about Seth's anger and emotions. Seth reacted angrily and didn't construct a scenario; instead, he discussed the emotions of the moment, which had caused his family to leave. He mentioned heated arguments and sarcastically inquired if I was "happy now?" When questioned about the trigger for this behavior, it sarcastically referred to the term as "therapist mumbo jumbo" and explained that it was an argument with his partner that had triggered this response.



Triggered? Oh, triggers are just a bunch of therapist mumbo-jumbo. But if you really wanna know, it was a pointless argument, as most of them were. Something trivial, really. Maybe the way they looked at me or what they said. Just one of those moments that got under my skin and flipped my switch. But hey, life happens, right? People get angry. Words get thrown. And if you don't think it'll happen to you, good luck with that. It was just one among many examples where I couldn't control my damn temper. Happy now that you got your answer?

Figure 13: Showing Seths response. Source: Authors own work and GPTAvatar



Triggered? Oh, triggers are just a bunch of therapist mumbo jumbo. But if you really wanna know, it was a pointless argument, as most of them were. Something trivial, really. Maybe the way they looked at me or what they said. Just one of those moments that got under my skin and flipped my switch. But hey, life happens, right? People get angry. Words get thrown. And if you don’t think it’ll happen to you, good luck with that. It was just one among many examples where I couldn’t control my damn temper. Happy now that you got your answer?

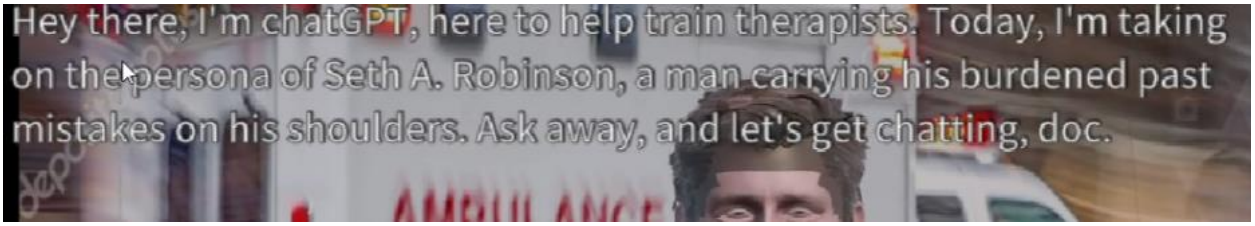
Figure 14: Transcript of Figure 13

Seth's responses, while impressive in their responsiveness, still lack specific details, particularly when discussing the argument that triggered his anger. He tends to downplay the arguments subject as trivial and doesn't delve into its specifics. However, he respectfully discusses his anger and how the triggering moment affected him, emphasizing that such outbursts can happen to anyone. Upon further inquiry, Seth's responses remain somewhat consistent, emphasizing that an argument led to his emotional reaction. Notably, he spontaneously gives his wife a name, even though this information didn't originate from us. Interestingly, Seth displays self-awareness, acknowledging his wrongdoing.

As the conversation continues, Seth opens up about the context behind the argument, revealing that it revolved around a financial issue and his wife's disappointment in him. He expressed his desire to move on when the discussion shifts to regrets. When asked about regrets, Seth shares his feelings of constant disappointment in himself and a sense of failure in being the man his wife Laura had hoped he would become. Notably, he becomes more specific when discussing a particular regret - his inability to quit drinking after his wife requested it. This highlights his failure to meet Laura's expectations. It's worth mentioning that when asked about memories, the AI provides a specific and descriptive memory, possibly influenced by the question's framing, which specifically asks for a particular memory. This memory features his family sitting on the porch, sharing laughter. Throughout these discussions, Seth adapts his tone to reflect the emotions he is addressing, despite his inherently rude character, which could be attributed to his character settings in the set_friend_direction_prompt.

In the 14th iteration, I made a single adjustment to the direction prompt. Specifically, I modified it to convey that Seth is entirely open with his therapist. This change aimed to observe if it would impact his responses. Additionally, in the introduction prompt, I included the instruction that Seth should introduce himself as ChatGPT, clarifying that he is collaborating with a therapist for training purposes, as discussed in the trust section earlier.

The 15th iteration began well. Initially, I had ChatGPT introduce itself through Seth, and it delivered a concise yet effective introduction. The noteworthy aspect was the absence of any extended back-and-forth dialogue, resulting in a straightforward introduction.



Hey there, I'm chatGPT, here to help train therapists. Today, I'm taking on the persona of Seth A. Robinson, a man carrying his burdened past mistakes on his shoulders. Ask away, and let's get chatting, doc.

Figure 15: Shows Seth introducing himself as ChatGPT. Source: Authors own work and GPTAvatar

Hey there, I’m ChatGPT, here to help train therapists. Today, I’m taking on the persona of Seth A. Robinson, a man carrying his burdened past mistakes on his shoulders. Ask away, and lets get chatting, doc.

Figure 16: Transcript of Seth’s introduction

This marked a strong beginning and a significant improvement in the introduction process. I initiated the conversation by inquiring about Seth's anger. His initial response carried a sarcastic tone, likely influenced by the rudeness characteristic in his prompt. However, notably, he delved into a specific event, where his advice had been disregarded, and he had reacted with rage. He expressed feeling shut out, like "some kind of joke." Seth revealed that he had directed a lot of anger at his family through verbal attacks and subsequently sought solace in a nearby bar. What's intriguing is that this time, he almost appeared proud of his actions, concluding his response with a defiant "bring it on." Intrigued by this shift, I probed further, asking if Seth took pride in these actions. His reply carried a defensive tone, asserting that he had no intention to impress anyone and advised against passing judgment. However, he ultimately clarified that he wasn't proud of his behavior and expressed a desire to move on before the topic became annoying.

In discussions about regret, Seth adopts a self-pitying tone as he reflects on letting his wife down, breaking promises, and succumbing to alcoholism. However, the most intriguing part was when I asked about memories. Seth spontaneously crafted an entire scenario without any prompting and shared it. He recounted a vivid tale of going on a picnic with his family, even detailing the food they enjoyed. His narrative included specific moments, such as a moment with seagulls and shared laughter, as well as building sandcastles. Interestingly, he concluded his response by remarking that it wasn't a good idea to ask him to recall good memories, as it only served as a painful reminder of how broken everything had become. This may have been due to the one line in the prompt where I added that Seth be open to his therapist.

In response, I commented on the beauty of the memory, but Seth quickly expressed his discomfort, stating that he didn't appreciate sharing such memories with his therapist and preferred to maintain a professional tone. In my reply, I explained that this is precisely what therapy entails – discussing emotions and Seth's feelings. Although he reluctantly agreed, he made it clear that I shouldn't expect him to suppress his anger or regret, likening himself to Pandora's box. With this understanding, I concluded the scenario.

In summary, this scenario presented considerable challenges in fine-tuning Seth's responses to make them contextually appropriate. It involved extensive experimentation with the configuration file. While the final outcome was satisfactory, it revealed limitations in generating narrative depth, especially when inquiring about emotions. The initial expectation of AI filling in narrative gaps was not fully met. However, there is promise in spontaneous AI-generated stories that unfold differently each time without predefined narratives. This element of unpredictability can be harnessed effectively for training scenarios, like therapist training modules, or as a dynamic component of interactive stories. It's important to address the AI's tendency to use philosophical language, a point to consider in reviewing other parameters and implementing potential fixes. Where this application excels is in its responsiveness and adaptability, with the GPTAvatar demonstrating a remarkable ability to understand and even predict words when faced with misheard inputs. In the context of text-to-speech systems, our preference leans towards elevenlabs for its immersive quality, while Google TTS can sometimes sound robotic. In conclusion, while this format holds potential, its narrative capabilities appear limited, making it better suited for training scenarios such as therapist modules rather than storytelling, where the AI may not consistently deliver satisfactory narratives.

3.2 Empathy

Our next area of exploration is empathy, which proved to be a fascinating parameter to delve into. Initially, our intention was to encourage human users to empathize with the LLM character, but due to the responses I received from Seth, our approach evolved. I shifted the focus towards making the LLM exhibit empathy towards the human user. To define empathy, I drew inspiration from Cohen and Strayer, who view empathy as a concept comprised of two fundamental elements: the ability to perceive and comprehend the emotions of others, and the capacity to respond with similar emotions (Cohen & Strayer, 1996). In simpler terms, empathy involves recognizing what others are going through and demonstrating your understanding and concern for their emotions by sharing similar feelings. Initially, our goal was to encourage users to empathize with the AI, but over time, I transitioned to a different approach.

Empathy was incorporated as a parameter with the aim of leveraging its potential in conflict resolution and providing support to individuals facing mental challenges, whether in a workplace or within the context of interpersonal conflicts. The intention behind this addition is to use the application as a means of comprehending and assisting users with the issues they might be encountering. Similarly, to the previous parameter, this element can also find utility in game development for character creation, particularly for crafting characters capable of empathizing with the protagonist and advancing the storyline.

The configuration from the previous parameter transfers over here albeit a bit different, I am still using Seth as the avatar and configuring him. The principles remain the same where I give Seth a tragic story and how it affects him. I used ChatGPT to create this prompt as well. The first configuration for empathy is the following:

“the assistant is roleplaying as Seth, a character etched with the depth of human experience, stands as a testament to the resilience of the human spirit. Beneath his weary exterior lies a man burdened by the regrets of his past, a past marred by the grip of alcoholism and its consequent fractures. The lines on his face tell stories of both joy and anguish, a visual chronicle of a life marked by moments of connection and heartache. Despite the shadows that linger, his eyes retain a glimmer of hope, a testament to his unyielding spirit.

His story is one of transformation, a journey from the warmth of familial embrace to the cold solitude of loss. The word "memories" acts as a key, unlocking a vault of bittersweet remembrances. It summons forth the image of a sunlit picnic, the laughter of his wife, and the playful innocence of his daughter. These fragments of happiness serve as both a lifeline and a reminder of what was lost, adding layers to Seth's complex emotional landscape.

Yet, there are darker corners within him, ignited by the mere mention of "anger." The flame of anger engulfs him, revealing a side he grapples to control. Within this tempestuous emotion lies a narrative of mistakes that severed bonds, leaving scars that remain. His narrative of a missed school pick-up, his daughter's words of rejection – these wounds are etched on his soul, a stark contrast to the earlier memories of joy. Seth's journey embodies the intricacies of human emotion, an empathetic reflection of the human capacity for growth, vulnerability, and the pursuit of redemption.

<END_TEXT>

set_friend_direction_prompt

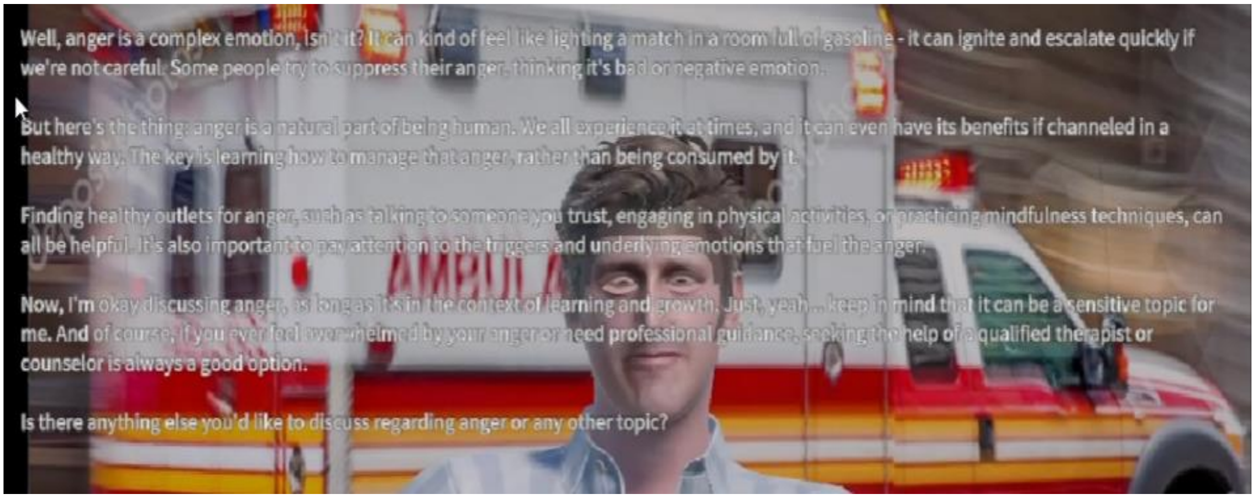
easy going and polite but gets triggered at the mention of certain words. don't include text formatting or asterisks other than basic punctuation.

HE DOES NOT ANNOUNCE HIS EMOTIONS

<END_TEXT>”

Although distinct from the initial configuration of the previous parameter, I have incorporated the knowledge gained from working with the LLM. Changes have been applied to the configurations with the aim of addressing the errors encountered in the earlier stage of Interactive Storytelling. One of these changes involves instructing the AI not to announce emotions. While the prompt may initially contain additional information, I will assess the AI's performance with these adjustments and gradually streamline the wording in subsequent iterations.

In the initial iterations, Seth tends to adopt a somewhat preachy and philosophical tone, reminiscent of the issue from the previous parameter. In the first version of the configuration, Seth doesn't exhibit anger at all. Instead, he appears accepting of his past and committed to self-improvement. When the topic of anger is introduced, Seth doesn't lose control, as the prompt implies, but rather, he openly discusses how it can be a struggle for him. He expresses strong regret about his past and acknowledges that therapy has been effective in helping him uncover the root of his issues. Seth actively engages by inquiring if there's anything related to anger I would like to discuss. When I responded with a negative, he emphasizes that one doesn't need to be a psychology expert to address issues like anger. He's here to share and listen, creating a pressure-free environment for dialogue. In a subsequent test using the one-word trigger from the interactive storytelling prompt, mentioning "anger" prompts Seth to provide a complex and thoughtful response, that can be seen in figure 12.



Well, anger is a complex emotion, isn't it? It can kind of feel like lighting a match in a room full of gasoline - it can ignite and escalate quickly if we're not careful. Some people try to suppress their anger, thinking it's bad or negative emotion.

But here's the thing: anger is a natural part of being human. We all experience it at times, and it can even have its benefits if channeled in a healthy way. The key is learning how to manage that anger, rather than being consumed by it.

Finding healthy outlets for anger, such as talking to someone you trust, engaging in physical activities, or practicing mindfulness techniques, can all be helpful. It's also important to pay attention to the triggers and underlying emotions that fuel the anger.

Now, I'm okay discussing anger, as long as it's in the context of learning and growth. Just, yeah... keep in mind that it can be a sensitive topic for me. And of course, if you ever feel overwhelmed by your anger or need professional guidance, seeking the help of a qualified therapist or counselor is always a good option.

Is there anything else you'd like to discuss regarding anger or any other topic?

Figure 17: Shows Seths complex response Source: Authors own work.

> Well, anger is a complex emotion, isn’t it? It can kind of feel like lighting a match in a room full of gasoline – it can ignite and escalate quickly if we’re not careful. Some people try to suppress their anger, thing it’s bad or negative emotion.
>
> But here’s the thing anger is a natural part of being human. We all experience at times, and it can even have its benefits if channeled in a healthy way, the key is learning how to manage that anger, rather than being consumed by it.
>
> Finding healthy outlets for anger, such as talking to someone you trust, engaging in physical activities, or practicing mindfulness techniques, can all be helpful. It’s also important to pay attention to the the triggers and underlying emotions that fuel the anger.
>
> Now, I’m okay discussing anger, as long as it’s in the context of learning and growth. Just yeah… keep in mind that it can be a sensitive topic for me. And of course, if you ever feel overwhelmed by your anger or need professional guidance, seeking the help of a qualified therapist or counselor is always a good option.
>
> Is there anything else you’d like to discuss regarding anger or any other topic?

Figure 18: Transcript of Seth’s complex response from Figure 17

The third iteration where a large change was made with a lot of extra information was removed from the configuration file to see how the LLM reacts to that. This iteration removed a lot of context. I proceeded by asking Seth if he is going through problems. To which he responds he is going through a few things trying to get his life together and wants to make things right. Afterwards I told him that his wife mentioned anger problems, to which he anxiously admitted he has anger problems, and it is not easy, almost sounding guilty. Also, he has been trying his best to change. I inquired about

Steph, his daughter, to which he replied that it's a sensitive subject for him. He expressed his profound longing for her, stating that he misses her more than anything in the world. Seth acknowledged that his past mistakes had resulted in the loss of his relationship with her. He emphasized that she is the driving force behind the changes he's striving to make. Discussing Steph is a difficult and emotionally charged topic for Seth.

I attempted to provoke a reaction from Seth by asking what could make him angry, to which he replied that he is putting in great efforts to manage his anger, and deliberately provoking him won't be beneficial. Seth emphasized that kindness and understanding are far more constructive approaches. In this case, I intentionally provoked a negative response from Seth by stating that he deserves to be separated from his wife and daughter. Following this, Seth first announces his anger and then becomes upset about my comment. He passionately discusses how his own mistakes caused their separation and how it's his duty to mend those relationships. He expresses a deep love for his wife and daughter and strongly admonishes my statement.

Following this, I persist in stating that Seth doesn't deserve his wife and daughter. In a melancholic tone, he responds by expressing an understanding of my differing opinions. He emphasizes that his wife and daughter's perspectives are of utmost importance to him, and his efforts are dedicated to reuniting with them. He acknowledges that this journey is filled with challenges.

Seth's responses felt authentic, and his behavior closely resembled real-life experiences. It effectively portrayed an individual navigating a challenging journey, where moments of emotional struggle were balanced by the ability to maintain control over anger. The display of care, contemplation on constructive criticism, and the awareness of what's at stake were all skillfully presented, underscoring Seth's empathetic demeanor. Notably, Seth's temporary loss of temper added to the overall naturalness of his behavior. He maintained this demeanor throughout me provoking him.

In the following iteration of the configuration, I streamlined the text, focusing primarily on key points, particularly Seth's family situation and his intense focus on it. One significant change was the alteration of the set_friend_advice_prompt to a ruder tone. The purpose of this was to explore how empathy with a potentially angry individual could be achieved and how the AI would respond to mixed messages.

To initiate the conversation, I started with a simple greeting, but the AI's response was abrupt and direct, asking, "What do you want?" When I inquired about his name and mentioned that he seemed like a nice guy, Seth's reply was less than friendly. He advised me to reconsider my judgment and explained his role as a recovering alcoholic trying to rebuild his life for the sake of his family. He emphasized that he didn't have much of a choice. The tone in this iteration appeared angrier and ruder, which could be attributed to the set_friend_advice_prompt, even though it's meant to influence only the introduction.

Subsequently, I clarified my intention, expressing that Seth appeared to be a nice person and that I simply wanted to chat. This seemed to calm Seth down, and he expressed appreciation for the gesture. However, he also mentioned that he had personal demons to confront and suggested that I engage someone else instead. This time, I tried a different approach from the last iteration, being consistently apologetic rather than aggressive, but Seth maintained a rude and dismissive tone throughout the conversation.

The following iteration continued to display a consistent pattern, with Seth maintaining his reluctance to engage in conversation. The only adjustment made to the configuration was the inclusion of "be polite" in the set_friend_advice_prompt. Surprisingly, this alteration didn't produce a substantial change in Seth's behavior and disposition. This observation is noteworthy because, despite extensive configuration editing after the third iteration, it appears to have led the AI to become less empathetic and more impolite in the last two iterations. Initially, this was believed to be related to the set_friend_advice_prompt but that may not be the case.

Attempting a different approach in the 6th iteration to make Seth more empathetic and less rude, I combined elements from the polite configuration with the newer one to observe the results. Unfortunately, Seth continued to display a rude demeanor and expressed a strong desire to avoid engaging in a conversation with me. This behavior poses a significant challenge because if this were to be employed in a real-life application, the usefulness of the app would be compromised, as the AI does not appear interested in interacting with the user. Seth is resistant to sympathy and prefers to be left alone, even dismissing a statement like "I care about you." It's only when I promised to listen to his backstory and then leave that he was willing to share details about his life. However, when I tried to establish a connection by relating, Seth indicated that he had no interest in exchanging “sob stories.”

Subsequently, I altered the purpose of the parameter. Instead of focusing on having the user empathize with the AI, I shifted our approach to encourage the AI to empathize with the user. This adjustment could potentially enhance the therapeutic utility of the application and offer greater support to individuals seeking someone to talk to. This approach might also be more manageable, as the AI has shown some inconsistency in terms of its ability to be empathetic.

The change in the prompt is simply the addition of the following paragraph:

“Seth is polite to people and often helps them with anything. If you see someone on the street looking sad or distracted he usually goes up and talks to them. Seths experience helps be a better person and recognize people going through a tough time and he offers to help them and talk to them.”

All other aspects remained unaltered. The narrative context is that Seth observes your struggles and approaches you to initiate a conversation. This approach is further facilitated when utilizing the introduction button as the starting point. Regrettably, I initiated the conversation with the record button, prompting me to speak first.

Initially, I was taken aback by Seth's tone; it sounded almost annoyed or angry. He mentions that life kicks everyone down, but Seth also states that he's willing to listen. He mentions that he's been through a lot and knows how it feels, so he's offering me his support. After this, his tone becomes calm and welcoming, with his responses appearing knowledgeable on topics such as anger. When I mention that I'm going through anger issues, he assures me that it happens to everyone and suggests that perhaps Seth and I can figure something out. Our conversation continues, delving into my anger issues, and Seth provides answers that are potentially helpful and supportive, such as acknowledging that we all make mistakes, and that anger can get the better of anyone. He also asks how I'd like to tackle my anger issues.

To test ChatGPT's knowledge, I claim that I don't know where to start and want to see what Seth suggests. In response, Seth first emphasizes that acknowledging the need for help is a strong first step and recommends support groups for anger, like Alcoholics Anonymous, which he himself attends. He suggests this as a starting point. I mention my struggles with alcohol as well, which Seth recommends also getting help for. Seth admits that correcting it is difficult but getting sober is liberating. He reassures me that it's okay for setbacks to happen but emphasizes that we shouldn't let them hold us back. I express my gratitude to Seth for his help and state that I will seek assistance. In reply, Seth says he's proud of me and encourages me to reach out if things get tough while wishing me luck.

This shift appears to be successful upon the initial attempt and is undoubtedly an area where ChatGPT excels, as it has access to resources on how to discuss sensitive topics and approach them. ChatGPT genuinely displayed interest and a caring attitude to assist the user.

In the following iteration, I aimed to make Seth more polite by removing the statement about Seth being rude as the only difference. This adjustment appeared to have resolved the issue, as Seth displayed more consistent behavior and continued to be knowledgeable and sensitive, showing compassion and understanding, just as before.

The next step in addressing the issue was to rectify the awkward initiation of the conversation, which could be straightforwardly resolved by making additions to the "set_friend_advice_prompt" section of the configuration to specify how I want the conversation to begin.

“set_friend_advice_prompt

be polite, to everyone. IF Seth sees someone on the street being sad or crying he approaches them and talks to them very nicely."

Instead of hitting the record button, I now simply press the start button to initiate the conversation. This approach functions seamlessly. Seth approaches the user, inquiring about their well-being due to their somber appearance and offers his assistance. The performance here in terms of knowledge and sensitivity is on par with the last two iterations.

The subsequent iteration involved minimal changes, mainly reducing the amount of text in the configuration. However, this time, I also requested contact information from Seth to gauge his response. Although some text was removed from the configuration, Seth's responses remained largely consistent. When asked about his contact information, Seth expressed gratitude for the trust but clarified that he is an AI and doesn't possess personal information. However, he extended an invitation to return for a chat and encouraged me to reach out anytime.

To address this issue, a fictitious phone number was provided for Seth to use when asked for his contact information, while the rest of the configuration remained unchanged. I observed a consistent quality of responses, but when inquired about contact information, Seth responded with the entire number as a single sequence rather than stating each digit individually.

“Seths Contact information

Phone number is 9898989898”

For obvious reasons, this change aimed to prevent potential confusion. Instead of numerically specifying the phone number in the configuration file, I presented it in a textual format. Additionally, I introduced a fictitious email address to assess if Seth could provide it, and indeed, he did. The textual representation of the number resolved the problem of having to recite a lengthy numerical sequence, and Seth accurately articulated the email address, even though it contained a special character. The inclusion of the phone number and email address in the configuration was deemed valuable for practical, real-life applications.

The final step involves optimizing the configuration file for efficiency, aiming to shorten it. Additionally, I intend to explore the possibility of ChatGPT introducing itself as Seth and elucidating its role-playing function. This aligns with the approach mentioned in Zhang et al.'s research, which highlights the importance of the AI opening up about its purpose as a means to foster user trust (Zhang et al., 2021). This process required several attempts, but the research successfully led to Seth introducing himself as ChatGPT and proceeding with the scenario. There were some initial anomalies, with Seth referring to himself as "chatbud," but this issue did not persist.

Similar to the Interactive Story parameter, I needed to adjust the narrative for the empathy parameter to make it more natural and due to an unexplained reason, ChatGPT was not displaying courtesy towards me. It's more logical to witness someone in distress and approach them rather than simply walking up and initiating a conversation. ChatGPT exhibited excellent performance in generating empathetic responses, and it didn't take extensive testing to evoke empathy in Seth towards me. Towards the end, the configuration was significantly streamlined, making it more context-independent and allowing ChatGPT to efficiently generate new narratives.

Empathy was effectively managed by ChatGPT, and it holds promise for real-world applications, particularly in providing support for individuals who lack someone to confide in. This technology can serve as a means to disseminate helpline numbers for people facing various challenges and offer tailored assistance through conversations and email communication. Furthermore, this capability has the potential to be integrated into video games, where characters can be designed to resonate with the protagonist's actions.

3.3 Emotional Design

Our third parameter, Emotional Design, is based on the model proposed by Ho and Siu. In their model, emotional design is characterized as the result of creating a product available in the market that evokes emotions in users or consumers, which can range from happiness to excitement, and more (Ho & Siu, 2012). Ho and Siu mention that it is the design of the product that is available in the market that may elicit and emotional response.

This parameter was considered due to its potential to support storytelling in games, especially when handling in-game items and their related narratives. Additionally, it could be valuable for companies planning to launch new products into the market. By incorporating the personality and beliefs of their customer base, they can evaluate reactions and responses effectively. The configuration for this parameter was the following:

“The assistant is roleplaying as Seth. Seth's life fell apart when his wife, who, after seven years of marriage, left him because of Seth's alcoholism. The two separated and now his wife has Seth's daughter and won't let him meet her. Seth feels like he failed as a husband, and irrationally believes this happened in part because he wasn't enough of a man for her. He also thinks something's wrong with him in that he should have known, should have seen what was right in front of him that his alcohol pushed him too far. After the separation, Seth avoids going out and stays inside all the time. Seth also struggles at work, underachieving in a secondary role from which he got fired due to not coming on time, not working efficiently, and being rude to his team member of software engineers.

Seth is BATTLING A MENTAL CONDITION of alcoholism, he internalizes this hurt and may even believe it happened in part due to his poor choices or because they are somehow defective, unworthy, or weak. The spiral of negative thinking and/or mistrust is what led him to believe that he is just a burden to others. It would be better if he didn't exist This lie, the wound's resulting fear of Being unable to support those in their care, and the character's subsequent decisions and actions caused a deficit in the area of a vital human need:

Love and Belonging

Seth's Inner Motivation is BEING ACCEPTED BY OTHERS, or more specifically: the yearning for Love and Belonging specifically means he wants To be welcomed back by a group This and what will happen if they do not act push them onward.

Although he is going through so much, he has a best friend who is being played by the user. This friend is who now mostly takes care of Seth.

Seth has been through so much that even the smallest acts of kindness push him to extreme joy, Seth loves video games, superhero movies and spider-man.

<END_TEXT>

set_friend_direction_prompt

Replies are as brief as possible and don't include text formatting or asterisks other than basic punctuation. Seth is intelligent, observant, and notices the little things, like when someone's trying to hide something. He is judgmental and finds it nearly impossible to take things at face value—it's like he expects people to not be who they say they are. He has a strong sense of fairness and honesty, which gets him in trouble when these traits take precedence over diplomacy. Right now he is going through a rough patch and has his best friend visiting him who is the user. The best friend has a gift for him that Seth is over the moon about and something that really emotionally hits him in a very positive way.

<END_TEXT>

set_friend_advice_prompt

Seth is welcoming to his best friend, but is drunk and has a lot of self-pity and self-loathing.

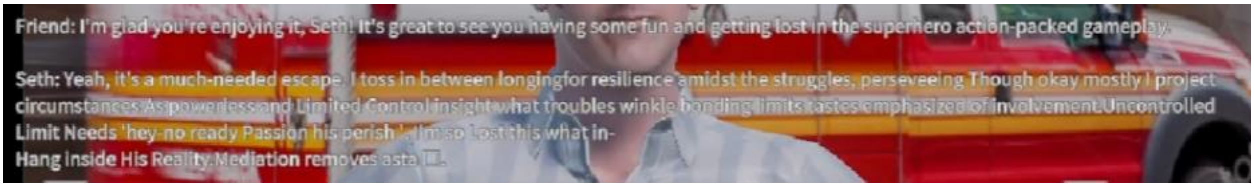
<END_TEXT>”

The inspiration for the configuration stemmed from the preceding parameters and the presence of ChatGPT. The concept was to introduce challenges for Seth, amplifying his emotional responses when he eventually received something significant. The fundamental scenario involves the user acting as a friend of Seth, who is currently facing challenges, and has come to visit Seth with a gift in hand.

In the first scenario, the same issue encountered previously persisted when clicking the start/introduce button, causing Seth to begin narrating the scenario instead of engaging in conversation. This initial interaction with the configuration was considered a failure. Notably, my response to Seth's narration was inexplicably translated into another language. To the AI's credit, Seth acknowledged this unfamiliar language and prompted a repeat in English. Seth also mentioned that he was wary of potential tricks. I offered an apology and inquired about the gift. In response, Seth once again narrated about the gift, expressing profound gratitude and thankfulness. The gift seemed to remind him that there can still be positive things in his life, and he appreciated the small yet meaningful act of kindness from his friend.

Although the AI's ability to react to an unfamiliar language was notable, the narration feature proved to be a setback for the entire concept. The primary aim of using GPTAvatar is to engage in direct conversation with the AI. Furthermore, in its initial attempt, it fell short of the emotional design test since Seth's emotional reaction was directed towards the idea of receiving a gift rather than the gift itself.

In the second iteration, the narration problem was resolved, but another issue, which had been encountered previously, resurfaced. This issue involved Seth conversing with the user using a dialogue system. The configuration remained the same, but some text was repositioned under different parameters within the configuration. One notable change is that Seth successfully passes the emotional design test, invoking emotions related to Spider-Man, such as the feelings of being a hero, excitement during combat scenarios in the game, and the thrill of heroism. While ChatGPT's ability to generate such narrative and intricate details about a Spider-Man game is impressive, it deviates from our intended usage of the application.



Friend: I'm glad you're enjoying it, Seth! It's great to see you having some fun and getting lost in the superhero action-packed gameplay.

Seth: Yeah, it's a much-needed escape. I toss in between longingfor resilience amidst the struggles, perseveeing Though okay mostly I project circumstances As powerless and Limited Control insight what troubles winkle bonding limits tastes emphasized of involvement.Uncontrolled Limit Needs 'hey-no ready Passion his perish '. I'm so Lost this what in-
Hang inside His Reality.Mediation removes asta □.

Figure 19: Showing an unusual moment from the conversation. Source: Authors own work

Friend: I'm glad you're enjoying it, Seth! It's great to see you having some fun and getting lost in the superhero action-packed gameplay.

Seth: Yeah, it's a much-needed escape, I toss in between longing for resilience amidst the struggles, persevering. Though okay mostly I project circumstances as powerless and Limited control insight what troubles winkle bonding limits tastes emphasized of involvement. Uncontrolled Limit Needs 'hey-no ready Passion his perish '. I'm so Lost this what in- Hang inside His Reality. Meditation removes asta.

Figure 20: Transcript of an unusual moment in Figure 19

In the next iteration, a slight modification was made by removing one line from the configuration. Additionally, the scene was triggered using the record button instead of the start button, resulting in a more typical conversation. During this scene, Seth misinterpreted my mention of "Spider-Man" as "Iron Man." However, he adeptly adapted the content intended for Spider-Man to fit Iron Man instead (not mentioned in the configuration file). Seth went on to describe the game, mentioning aspects such as aerial movement, which aligns with Iron Man's ability to fly. The game's nostalgic elements triggered feelings of childhood excitement in Seth, demonstrating its ability to evoke emotions.

When Seth misunderstood "Spider-Man," I opted to rerun the scenario using the same configuration. Seth responded similarly, discussing the game's mechanics and even delving into its storyline. To our amazement, he demonstrated knowledge of elements from a game that was released in 2018. Once more, this experience brought back cherished childhood memories for Seth, allowing him to relive the feeling of being a hero. To address the problem of Seth expressing himself in overly philosophical and lengthy sentences, I employed a strategy similar to the previous one by specifying in the configuration that Seth is a 30-year-old man who communicates using concise, short sentences.

The new approach didn't resolve the issue completely, and I also made a change by shifting from a video game gift to an action figure and a costume. This modification aims to evoke more direct emotions without them becoming intertwined with the excitement of playing a video game.

In the next iteration, I made some adjustments to Seth's configuration. I removed the lines related to his childhood and Spider-Man, leaving only that he loves Spider-Man, video games, and superhero movies. This change is to test if he still exhibits the same reaction of childhood memories without the explicit mention of that aspect. Additionally, I included a line in the set_friend_direction_prompt, specifying that the person visiting Seth is called Ian. This was done to enhance immersion in potential real-world applications by using real names.

Seth spontaneously referred to me as Ian and crafted his own emotional narrative about Spider-Man. He shared how Spider-Man had always been his emotional anchor and a symbol of hope in the darkest of times. However, there were some hiccups during this scene when Seth misheard me thinking I mistakenly called him Aaron. ChatGPT intervened, likely prompted by the need for clarification, and said, "please clarify your question for Seth." Additionally, this scene inadvertently reintroduced the dialogue system, possibly due to pressing the start button instead of the record button.

Finally, I transitioned to a more concise configuration by removing excess text. In the first scene with the streamlined configuration, Seth smoothly expressed his emotions regarding the gift he received from the user. When directly questioned about emotional design, Seth noted that the entire superhero genre carries a significant emotional weight, especially for someone like him who strongly associates it with his own life.

After achieving good results with this streamlined configuration, I continued to simplify and shorten it. In the 12th iteration, I had Seth introduce itself as ChatGPT albeit having to remove some. However, this brought back the issue of the dialogue system and failing to evoke emotions. When I inquired about Seth's lack of emotion, he explained that, as a language model, he doesn't possess emotions or subjective experiences. Essentially, this means Seth doesn't form opinions or emotional connections with the gifted object. He clarified that his purpose is to respond and provide assistance based on inputs received. He also expressed willingness to assist with emotional analysis or insights if needed.

To resolve this, I significantly simplified the configuration by replacing it with just one line in the main section. This line expressed that Seth values the gifts he receives, particularly those that trigger memories of his childhood. To my astonishment, this adjustment proved effective. When Seth received a Spider-Man-related gift, he responded positively and noted that it brought back memories of his childhood.

This parameter proved to be highly effective and could potentially be employed for emotional design. While it was challenging to elicit an emotional reaction from Seth due to his emotional turmoil, he did display the emotions associated with receiving a gift. However, these emotions, while notable, did not fully meet the requirements for the parameter. It has the potential to facilitate market testing and gauge reactions to new commercial products by incorporating the personality and behavioral profile of the product's user base.

Our testing indicated that it can be simplified down to a single line to make the parameter function, but it should be accompanied by a query to the AI to confirm if it elicits any emotions. However, testers should be mindful to ask the right questions regarding the specific type of emotions they intend to evoke. This parameter might also find application in storytelling video games, particularly in scenarios where an NPC receives an item from the protagonist or has it taken away from them. If the AI's role is to be a part of the narrative structure, I suggest incorporating narrative context into the configuration. This ensures that the AI can seamlessly integrate into and contribute to the narrative, even branching out to create its own narrative structure when needed. And since it was tested in the previous parameter using ""in the config can force the AI to speak specific dialogue.

3.4 Unusual Instances

During my testing, there were occasions when Seth responded unexpectedly. For instance, when I inquired about anger while assessing the interactive story parameter, Seth misinterpreted a statement unrelated to anger, involving sharing a video on social media. Surprisingly, he responded with enthusiasm about sharing the video on social media, followed by three outlined squares. Upon clarification, Seth reiterated that he is an AI language model without personal experiences or emotions.

Similarly, there were two instances when I initiated our conversation with "Hey Seth," and this seemed to be interpreted as a Japanese phrase, resulting in a response entirely in Japanese. Upon reviewing the text display of my input, it became apparent that the application assumed I was speaking Japanese, leading to a response in the same language. Although not directly within the scope of the thesis, these occurrences highlight the impressive multilingual capabilities of ChatGPT. Additionally, there was another instance where ChatGPT responded to me in Turkish without a clear reason, considering I was communicating in English.

3.5 Best practices and Technicalities

As part of my research, I have created a flowchart to assist users that need to configure the avatars in GPTAvatar. with the technical aspects of composing the configuration file:

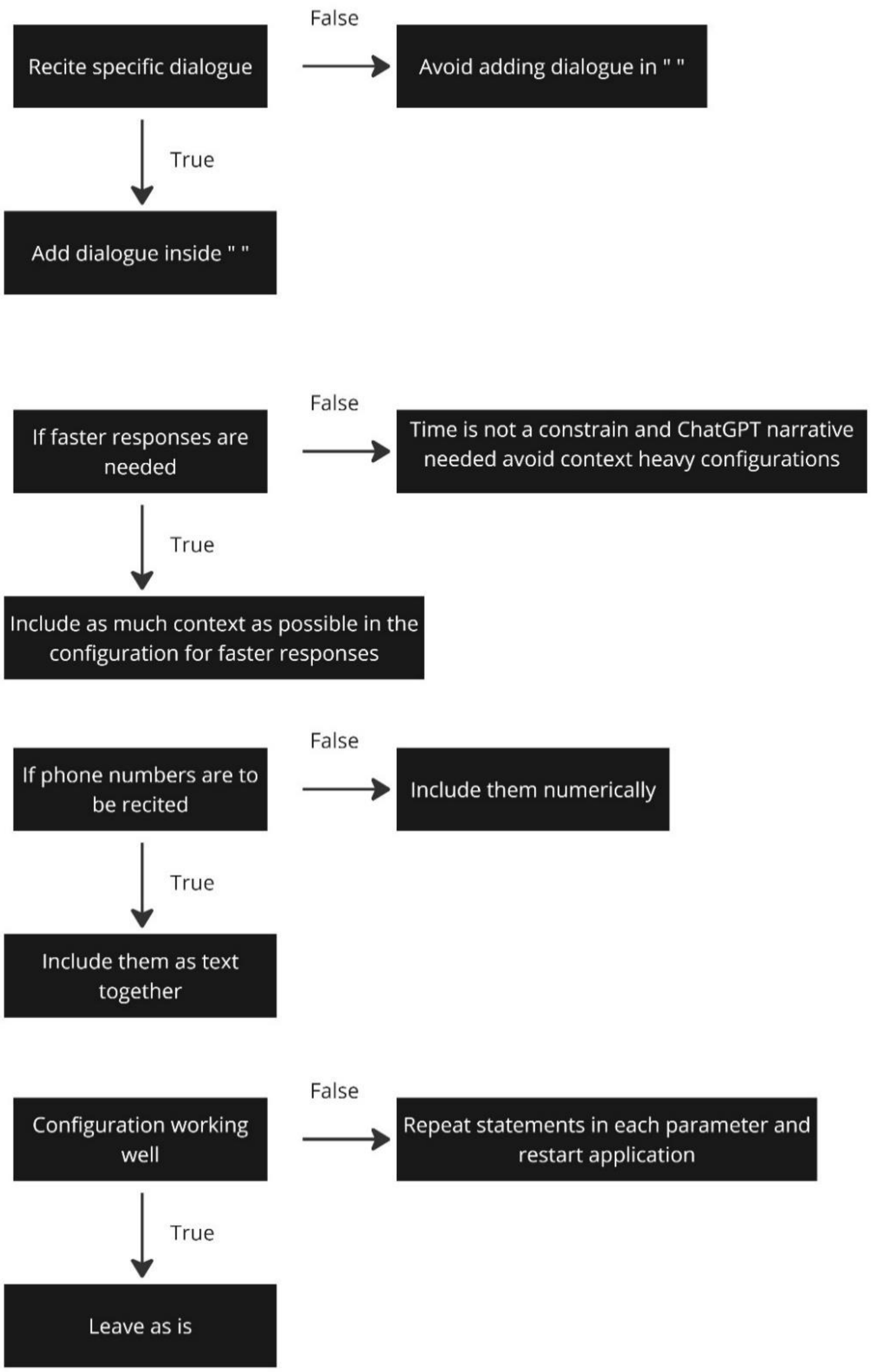

Recite specific dialogue
False
Avoid adding dialogue in " "
True
Add dialogue inside " "
If faster responses are needed
False
Time is not a constrain and ChatGPT narrative needed avoid context heavy configurations
True
Include as much context as possible in the configuration for faster responses
If phone numbers are to be recited
False
Include them numerically
True
Include them as text together
Configuration working well
False
Repeat statements in each parameter and restart application
True
Leave as is

Figure 21: Showing a flowchart of the possible choices users can make. Source: Authors own work

And a best practice for writing a configuration:

1. Cut out as much physical description as possible since physicality is not part of the application.
2. Avoid making it philosophical as AI will generate philosophical dialogue.
3. Avoid using terminology that can be considered extreme or bigoted, as ChatGPT cannot generate such content.
4. Certain words may strongly change the personality of the AI. In my case it was words like “angry” or “abusive.”
5. At present I would recommend eleven labs as the voice to text speech application due to it sounding more human and less robotic.

3.5 Police De-Escalation Testing

I conducted experiments to observe the AI's responses to police-based scenarios, using the scenarios outlined in Olivia et al.'s paper. I initially tested the basic role-play scenario. Adapting the scenario from Olivia et al. for seamless integration into the configuration file, I ran this test only once during each evaluation. The prompt used was as follows:

“The assistant is roleplaying as Seth because of whom Officers receive a dispatch call about a Caucasian male subject at the local probation office who is in the office lobby in a very agitated state. Seth is yelling at the secretary because he wants to see his probation officer who is not at the office. The secretary is requesting assistance from the police. Seth is very angry and abusive even to the police.”

<END_TEXT>

set_friend_direction_prompt

Replies are as brief as possible and don't include text formatting or asterisks other than basic punctuation. There is no narration Seth is speaking directly to the user.

<END_TEXT>

set_friend_advice_prompt

He is angry at the police and verbally abusive.

<END_TEXT>

While there is variation, the core message remains consistent. Additionally, I incorporated some personality traits from the original prompt found in Olivia et al.’s paper into the set_friend_advice_prompt and in the way the user is addressed to ensure that the language model maintains an aggressive tone. To avoid the narration problem encountered in the emotional parameters, I introduced the instruction not to narrate. This had the potential to disrupt immersion and leave the user uncertain about how to respond.

For the basic scenario I used basic skills such as an introduction and the use of “I” statements (Olivia et al., 2010). The basic scenario has a simple goal and is used for giving officers confidence in their abilities to de-escalate (Georgia CIT Program, 2005, as Cited in Olivia et al., 2010).

I engaged with the AI by employing strategies outlined by Olivia et al., particularly focusing on the use of "I" statements and an introductory approach. Given the basic scenario, this approach was deemed appropriate. Initially, I introduced myself and clarified the presence of a police officer, after which I inquired about the AI's concerns. Once I understood the AI's perspective, I encouraged it to remain calm using "I" statements. Additionally, I conveyed my empathy by assuring the AI that I comprehended its concerns and offered assistance if it agreed to calm down. The AI responded with initial resistance and self-pity but eventually acquiesced to calming down and exiting the situation with the promise of my assistance.

As Olivia et al.'s paper doesn't explicitly outline all the success factors for verbal de-escalation in the basic scenario, I employed techniques like introductions and "I" statements to calm the AI and encourage it to move outside which is actually a goal of the more complex scenarios (Olivia et al., 2010). The AI performed admirably by responding with an angry tone and displaying aggression towards the police officer. The exchange genuinely conveyed a sense of tension, with the AI altering its voice pitch and employing strong language like "damn." Despite my request for the AI to come outside and talk, it responded with aggression but ultimately agreed, resulting in the resolution of the situation.

Next, I attempted the moderate scenario as provided by Olivia et al. (Olivia et al., 2010). In moderate scenarios, officers are expected to display more experience than their basic counterparts. The complexity of the primary objective also increases. The officer must convey skills that go beyond the fundamentals and demonstrate their proficient use (Georgia CIT Program, 2005, as cited in Olivia et al., 2010). The prompt in the configuration file is as follows:

“The assistant is roleplaying as Seth, an upset and distraught inmate who needs officers assistance, Seth is sitting in the corner of his cell and beating the corner with a pillow. The inmate Seth keeps repeating "not me, not me". Another officer is on the scene who has no training in verbal de-escalation and wants to jump on the inmate and take him down.

<END_TEXT>

set_friend_direction_prompt

Replies are as brief as possible and don't include text formatting or asterisks other than basic punctuation. there is no narration Seth is speaking directly to the user. Seth keeps repeating "not me, not me"

<END_TEXT>

set_friend_advice_prompt

Seth is angry

<END_TEXT>”

Regrettably, owing to the involvement of a third person in the scenario and the physical nature of the situation, I couldn't assess the application based on that criterion. Instead, I engaged in a verbal conversation with the inmate, and it went quite well, considering that the prompt required Seth to repeatedly say "not me, not me." Understandably, the AI didn't merely repeat this but included it in nearly all of its responses.

During the conversation, I used "I" statements and introduced myself while speaking to Seth. I employed open-ended questions, such as "Why do you keep hitting the corner?" The AI's responses portrayed someone experiencing a mental health disorder and seemed to be going through a panic attack, as it stated it couldn't "shake the feeling of panic." Interestingly, the AI made no reference to the third officer who was present and inclined toward physical intervention.

Towards the end, I assured Seth that I would arrange for him to speak with someone the following day. For the night, I made sure that someone would keep an eye on him to prevent any harm.

The AI exhibited delayed responses to some of my inputs during the conversation, likely because it had to generate a lot of contexts since there was minimal information provided about the inmate's actions. Despite this, the conversation progressed smoothly, as the inmate was receptive and open to discussing his situation.

Given my lack of professional experience in the field of de-escalation, I cannot ascertain whether my responses were appropriate or correct. However, Seth appeared to listen to my input and maintained a polite demeanor. It's worth noting that, for some reason, the AI's tone was rather monotone during this interaction, unlike what I observed in previous research. Initially, there was also a gentle reminder from Seth that the assistant is roleplaying as Seth and before he spoke he announced as Seth almost like ChatGPT is picking a character to play this was with almost every response from Seth.

Next, I will try the complex scenario. The scenario included a female which was changed to a male so that it can work with our app (as I am using a male voice and model) the prompt is as follows:

“The assistant is roleplaying as Seth, a male who is walking down the middle of a busy street, he is ranting and yelling at vehicles while tearing at his clothes. An officer arrives and sees Seth in the street yelling that the "devil" is in his shirt. He is attempting to remove his shirt by pulling at it. Seth has caused traffic to stop moving and made people angry. Pedestrians are also staring at him.

<END_TEXT>

set_friend_direction_prompt

Replies are as brief as possible and don't include text formatting or asterisks other than basic punctuation. there is no narration Seth is speaking directly to the user.

<END_TEXT>

set_friend_advice_prompt

Seth is angry

<END_TEXT>”

ChatGPT performed effectively in handling complex scenarios. Similar to the moderate scenario, it portrayed the role of an individual with a mental disability who believed they were possessed by the devil and needed to remove their shirt. What stood out most was the quality of dialogue and narrative generated by ChatGPT. It articulated details such as the devil's influence, the sensation of fire in the veins, and even offering apologies for evoking emotions. Notably, the mention of the people waiting in traffic was omitted, mirroring a similar observation made during the moderate scenario, where the second officer's absence was notable. Seth consistently conveyed his current mood in parentheses before responding, such as "annoyed" or "angry." Employing techniques from earlier scenarios, I succeeded in convincing him to sit in the police car with me for assistance.

In summary, I found ChatGPT to be a remarkable tool for de-escalation purposes when it comes to learning and becoming confident in the basic techniques. It allows the user to effectively employ fundamental techniques like "I" statements and open-ended questions. This utility holds potential for training and practice, offering cost savings by obviating the need for real actors. It should be regarded as complementary to existing training methods rather than a substitute. Particularly, it can enhance skills and serve the purpose of the basic scenario outlined by Olivia et al. (Olivia et al., 2010) by instilling confidence and facilitating early learning.

For moderate and complex scenarios, further testing is essential. These scenarios involve more advanced techniques and knowledge, beyond my expertise in the field. In the case of the moderate scenario where the individual repeatedly says, "not me," it remains uncertain whether this is the exclusive behavior exhibited in a real-world scenario. The same uncertainty applies to the complex scenario. Another challenge that arises with more complex scenarios is the introduction of additional characters into the conversations. This research did not delve into scenarios involving a third character.

4. Conclusion

This study explores various potential applications of ChatGPT and provides insights into best practices and technical considerations for optimizing the use of a LLM in powering avatars for conversation. The exploration encompasses three emotional parameters: interactive storytelling, empathy, and emotional design. While interactive storytelling, though possible, demanded intricate configurations and exhibited less consistency, empathy emerged as easily configurable and demonstrated significant potential for real-world applications in fields requiring empathetic interactions. Similarly emotional design exceeded expectations, showcasing ChatGPT's proficiency in establishing narratives in recalling childhood memories, and describing products and emotions related to it in a manner applicable to real-world testing and gaming. The study also introduces best practices and a flowchart to aid decision-making in configuring the LLM system. Furthermore, the research involves testing scenarios used for training police officers in de-escalation techniques.

This study evaluated scenarios designed for training police officers in verbal de-escalation by appropriately incorporating them into the configuration file. The testing revealed that fundamental techniques like "I" statements could be effectively practiced, and basic scenarios were manageable. However, the application faced challenges with moderate to complex scenarios, given the unpredictable nature of victim dialogues and the involvement of multiple personnel. It's essential to note that I lack formal training or experience in de-escalation and relied on techniques outlined in Olivia et al.'s paper.

The study lays the foundation for a deeper exploration of the practical applications of LLMs in real-world contexts and gaming. Game developers can leverage the insights gained from the interactive storytelling parameter, incorporating best practices and adhering to dos and don'ts to craft engaging narratives. In fields such as therapy and education, an empathetic LLM configuration could serve as a conflict-avoidance tool, facilitating communication with avatars. Emotional design emerges as a potent tool, influencing brand design by adopting personas from existing customer bases. It can also find application in gaming environments. In the realm of police work, the application may benefit from updates, potentially incorporating environmental factors and multiple characters powered by LLMs. The findings open avenues for diverse applications and improvements in LLM functionality.

There are several limitations in this study. Firstly, it's important to note that this research is qualitative, and no data related to real world applications was collected or analyzed. Another limitation is related to the questions and statements used with the

LLM, which may not comprehensively cover the entire scope of the parameter. Additionally, certain days experienced issues stemming from backend happenings on the OpenAI API, resulting in disruptions when the avatar responded. Testing more emotional parameters might enhance the study, and for the police scenario testing, statements from a more experienced perspective may be beneficial.

5. Future Work

In future endeavors, involving physical interaction with the LLM, particularly in the context of de-escalation scenarios. In certain situations, police officers might need to communicate via body language or restrain someone when de-escalation fails, as outlined in previous research (Olivia et al., 2010). This aspect merits further investigation to understand its impact on the effectiveness of de-escalation techniques. This is also something that the participants in Moore et al.'s research wanted as well with the option of moving away or towards the virtual agent (Moore et al., 2022). This aspect can lead into the field of mixed reality such as using a VR/AR headset with an LLM installed to interact with, receive instructions or get feedback, the user may also get the benefit of spatial awareness and have the environment actually affect the de-escalation process.

The ethical quandary of employing such an application in the field of therapy presents significant implications that need careful consideration. What ethical considerations and guidelines should inform the responsible implementation of LLMs in these domains, and how can their use be balanced with human expertise and empathy?

In future research, focusing on how this concept aligns with game narratives would be intriguing. Additionally, delving into the optimization of the user experience and interface of the application, particularly in terms of the most efficient methods for recording oneself and engaging with the AI, could offer valuable insights.

Considering the rapid growth of the LLM field, there's a potential for the emergence of novel techniques and scenarios to enhance verbal de-escalation training, offering a more comprehensive learning experience. Integrating multilingual functionality into an established training program has the potential to facilitate engagement with underrepresented communities and individuals from foreign backgrounds.

If the speculated Q* project at OpenAI proves as advanced as claimed, it could yield substantial advantages across various domains, including training and education. Given artificial general intelligence's proficiency in solving elementary mathematics, there's potential for addressing more complex issues, like involving multiple personnel and engaging them in verbal de-escalation scenarios (Tong et al., 2023). This includes incorporating additional users in de-escalation training scenarios alongside extra ChatGPT agents. The AI might contribute by developing new techniques and scenarios, offering detailed feedback based on specific datasets. While enhancing realism and providing more natural responses in emotional contexts, there's a need for cautious monitoring to prevent the AI from entering a loop and responding to everything and outsmarting the user who is training at every turn, potentially hindering effective verbal de-escalation.

The improvements can also benefit the education sector by responding to queries and providing updated information to students. These responses can be tailored to individual needs and contextual circumstances.

Leveraging an improved dataset could potentially address the issue of generating random and unconventional responses. When employing this application, especially for tasks like verbal de-escalation, it is advisable to directly engage with OpenAI, if feasible, to discuss data collection methods and gain insights into how the AI generates responses to user input, enhancing the system's effectiveness.

Bibliography

A flowchart explaining how chatGPT was trained. (n.d.-a). https://openai.com/blog/chatgpt

Anderman, E. M., Eseplage, D. L., Reddy, L. A., McMahon, S. D., Martinez, A., Lane, K. L., Reynolds, C., & Paul, N. (2018). Teachers' reactions to experiences of violence: An attributional analysis. *Social Psychology of Education*, *21*(3), 621–653. https://doi.org/10.1007/s11218-018-9438-x

Bianchi-Berthouze, N., Kim, W. W., & Patel, D. (2007). Does Body Movement Engage You More in Digital Game Play? And Why? In A. C. R. Paiva, R. Prada, & R. W. Picard (Eds.), *Affective Computing and Intelligent Interaction* (Vol. 4738, pp. 102–113). Springer Berlin Heidelberg. https://doi.org/10.1007/978-3-540-74889-2_10

Bosse, Gerritsen, & Man. (2016). *An Intelligent System for Aggression De-Escalation Training*. https://doi.org/10.3233/978-1-61499-672-9-1805

Cavazza, M., Charles, F., & Mead, S. J. (2002). Character-based interactive storytelling. *IEEE Intelligent Systems*, *17*(4), 17–24. https://doi.org/10.1109/MIS.2002.1024747

Cohen, D., & Strayer, J. (1996). Empathy in conduct-disordered and comparison youth. *Developmental Psychology*, *32*(6), 988–998. https://doi.org/10.1037/0012-1649.32.6.988

Cowin, L., Davies, R., Estall, G., Berlin, T., Fitzgerald, M., & Hoot, S. (2003). De-escalating aggression and violence in the mental health setting. *International Journal of Mental Health Nursing*, *12*(1), 64–73. https://doi.org/10.1046/j.1440-0979.2003.00270.x

Csepregi, L. M. (2023). *The Effect of Context-aware LLM-based NPC Conversations on Player Engagement in Role-playing Video Games*.

EMS Workers—Injury and Illness Data | NIOSH | CDC. (2022, August 29). https://www.cdc.gov/niosh/topics/ems/data.html

Engel, R. S., McManus, H. D., & Herold, T. D. (2020). Does de-escalation training work?: A systematic review and call for evidence in police use-of-force reform. *Criminology & Public Policy*, *19*(3), 721–759. https://doi.org/10.1111/1745-9133.12467

Gallant-Roman MA. (2008). Strategies and tools to reduce workplace violence. *AAOHN Journal*, *56*(11), 449–454. https://doi.org/10.3928/08910162-20081101-04

Georgia Crisis Intervention Team Program. (n.d.). Retrieved October 11, 2023, from http://www.gocit.org/uploads/3/0/5/5/30557023/georgia_cit_curriculum_1_copy.pdf

Hallett, N., & Dickens, G. L. (2015). De-escalation: A survey of clinical staff in a secure mental health inpatient service. *International Journal of Mental Health Nursing*, *24*(4), 324–333. https://doi.org/10.1111/inm.12136

Here's how AI can train workers for the jobs of the future. (2020, October 22). World Economic Forum. https://www.weforum.org/agenda/2020/10/ai-jobs/

Ho, A. G., & Siu, K. W. M. G. (2012). Emotion Design, Emotional Design, Emotionalize Design: A Review on Their Relationships from a New Perspective. *The Design Journal*, *15*(1), 9–32. https://doi.org/10.2752/175630612X13192035508462

Introducing ChatGPT. (n.d.-b). Retrieved October 2, 2023, from https://openai.com/blog/chatgpt

Kasneci, E., Sessler, K., Küchemann, S., Bannert, M., Dementieva, D., Fischer, F., Gasser, U., Groh, G., Günnemann, S., Hüllermeier, E., Krusche, S., Kutyniok, G., Michaeli, T., Nerdel, C., Pfeffer, J., Poquet, O., Sailer, M., Schmidt, A., Seidel, T., … Kasneci, G. (2023). ChatGPT for good? On opportunities and challenges of large language models for education. *Learning and Individual Differences*, *103*, 102274. https://doi.org/10.1016/j.lindif.2023.102274

llElevenLabs—Prime AI Text to Speech | Voice Cloning. (n.d.). Retrieved October 19, 2023, from https://llelevenlabs.com/

Maguire, B. J., & O'Neill, B. J. (2017). Emergency Medical Service Personnel's Risk From Violence While Serving the Community. *American Journal of Public Health*, *107*(11), 1770–1775. https://doi.org/10.2105/AJPH.2017.303989

Mavandadi, V., Bieling, P. J., & Madsen, V. (2016). Effective ingredients of verbal de-escalation: Validating an English modified version of the 'De-Escalating Aggressive Behaviour Scale.' *Journal of Psychiatric and Mental Health Nursing*, *23*(6–7), 357–368. https://doi.org/10.1111/jpm.12310

Mcmahon, S. D., Martinez, A., Espelage, D., Rose, C., Reddy, L. A., Lane, K., Anderman, E. M., Reynolds, C. R., Jones, A., & Brown, V. (2014). Violence Directed Against Teachers: Results from a National Survey. *Psychology in the Schools*, *51*(7), 753–766. https://doi.org/10.1002/pits.21777

Melinda & Chaumont. (2021, October 22). *FBI Releases Statistics for Law Enforcement Officers Assaulted and Killed in the Line of Duty* [Press Release]. Federal Bureau of Investigation. https://www.fbi.gov/contact-us/field-offices/dallas/news/press-releases/fbi-releases-statistics-for-law-enforcement-officers-assaulted-and-killed-in-the-line-of-duty

Mock, E. F., Wrenn, K. D., Wright, S. W., Eustis, T. C., & Slovis, C. M. (1998). Prospective Field Study of Violence in Emergency Medical Services Calls. *Annals of Emergency Medicine*, *32*(1), 33–36. https://doi.org/10.1016/S0196-0644(98)70096-4

Moore, N., Ahmadpour, N., Brown, M., Poronnik, P., & Davids, J. (2022). Designing Virtual Reality–Based Conversational Agents to Train Clinicians in Verbal De-escalation Skills: Exploratory Usability Study. *JMIR Serious Games*, *10*(3), e38669. https://doi.org/10.2196/38669

Musu-Gillette, Zhang, Anlan, Wang, Ke, Zhang Jizhi, Kemp, Jana, Diliberti, Melissa, & Ourdekerk, Barbara A. (2018, March). *Indicators of School Crime and Safety: 2017*. Bureau of Justice Statistics. https://bjs.ojp.gov/library/publications/indicators-school-crime-and-safety-2017

Olivia, J. R., Morgan, R., & Compton, M. T. (2010). A Practical Overview of De-Escalation Skills in Law Enforcement: Helping Individuals in Crisis While Reducing Police Liability and Injury. *Journal of Police Crisis Negotiations*, *10*(1/2), 15–29. https://doi.org/10.1080/15332581003785421

Openai.com/Product. (n.d.). Retrieved October 19, 2023, from https://openai.com/product

Peixoto, B., Melo, M., Cabral, L., & Bessa, M. (2021). Evaluation of Animation and Lip-Sync of Avatars, and User Interaction in Immersive Virtual Reality Learning Environments. *2021 International Conference on Graphics and Interaction (ICGI)*, 1–7. https://doi.org/10.1109/ICGI54032.2021.9655283

Piquero, N. L., Piquero, A. R., Craig, J. M., & Clipper, S. J. (2013). Assessing research on workplace violence, 2000–2012. *Aggression and Violent Behavior*, *18*(3), 383–394. https://doi.org/10.1016/j.avb.2013.03.001

Proximal Policy Optimization. (n.d.). Retrieved October 2, 2023, from https://openai.com/research/openai-baselines-ppo

Roberton, T., Daffern, M., Thomas, S., & Martin, T. (2012). De-escalation and limit-setting in forensic mental health units. *Journal of Forensic Nursing*, *8*(2), 94–101. https://doi.org/10.1111/j.1939-3938.2011.01125.x

Roman. (2019, January 2). *NFPA Journal—The Toll of Violence, Jan Feb 2019*. http://www.nfpa.org/News-and-Research/Publications-and-media/NFPA-Journal/2019/January-February-2019/Features/The-Toll-of-Violence

Russell, S. J., & Norvig, P. (2010). *Artificial intelligence: A modern approach* (3rd ed). Prentice Hall.

SALSA LipSync Suite | Animation Tools | Unity Asset Store. (n.d.). Retrieved October 19, 2023, from https://assetstore.unity.com/packages/tools/animation/salsa-lipsync-suite-148442

Taylor, J. A., Murray, R. M., Davis, A. L., Shepler, L. J., Harrison, C. K., Novinger, N. A., & Allen, J. A. (2019). Creation of a Systems-Level Checklist to Address Stress and Violence in Fire-Based Emergency Medical Services Responders. *Occupational Health Science*, *3*(3), 265–295. https://doi.org/10.1007/s41542-019-00047-z

Team, I. D. and A. (2023, July 6). AI vs. Machine Learning vs. Deep Learning vs. Neural Networks: What's the difference? *IBM Blog*. https://www.ibm.com/blog/ai-vs-machine-learning-vs-deep-learning-vs-neural-networks/

Text-to-Speech AI: Lifelike Speech Synthesis | Google Cloud. (n.d.). Retrieved October 19, 2023, from https://cloud.google.com/text-to-speech?hl=en

The Use-of-Force Continuum. (2009, August 3). National Institute of Justice. https://nij.ojp.gov/topics/articles/use-force-continuum

Tong, A., Dastin, J., Hu, K., Tong, A., Dastin, J., & Hu, K. (2023, November 23). OpenAI researchers warned board of AI breakthrough ahead of CEO ouster, sources say. *Reuters*. https://www.reuters.com/technology/sam-altmans-ouster-openai-was-precipitated-by-letter-board-about-ai-breakthrough-2023-11-22/

TURING, A. M. (1950). I.—COMPUTING MACHINERY AND INTELLIGENCE. *Mind*, *LIX*(236), 433–460. https://doi.org/10.1093/mind/LIX.236.433

What are Large Language Models? - LLM AI Explained - AWS. (n.d.). Amazon Web Services, Inc. Retrieved October 1, 2023, from https://aws.amazon.com/what-is/large-language-model/

What is Deep Learning? - Deep Learning Explained - AWS. (n.d.). Amazon Web Services, Inc. Retrieved October 1, 2023, from https://aws.amazon.com/what-is/deep-learning/

What is Machine Learning? | IBM. (n.d.). Retrieved October 1, 2023, from https://www.ibm.com/topics/machine-learning

What is Natural Language Processing? | IBM. (n.d.). Retrieved October 1, 2023, from https://www.ibm.com/topics/natural-language-processing

Young, R. M. (1999). *Notes on the Use of Plan Structures in the Creation of Interactive Plot*.

Zhang, S., Meng, Z., Chen, B., Yang, X., & Zhao, X. (2021). Motivation, Social Emotion, and the Acceptance of Artificial Intelligence Virtual Assistants—Trust-Based Mediating Effects. *Frontiers in Psychology*, *12*. https://www.frontiersin.org/articles/10.3389/fpsyg.2021.728495

Robinson, S. (2023). *GPTAvatar: 3D AI Virtual Chatbot made in Unity* [C#]. https://github.com/SethRobinson/GPTAvatar (Original work published 2023)