

Carmen Won't Talk to Me: Facilitating Expressive Qualities in Games
Through Natural Language Interfaces and AI NPCs

by

Tamika Yamamoto

A thesis presented to OCAD University

in partial fulfillment of the requirements for the degree of Master of Design in Digital Futures

Toronto, Ontario, Canada, 2024

Abstract

Given the broad popularization of AI creative tools in contemporary society, this thesis explores the use of natural language interfaces by game designers as a primary gameplay mechanic to facilitate the expressive goals of the underrepresented game-maker.

The creative contributions of underrepresented game-makers hold cultural significance, hence, this research strongly advocates for the democratization of game-making tools, a need further underscored by the emergence of AI development. Such democratization is crucial as it addresses the exploitation of their creative labor within AI development, which frequently disregards the languages and cultures of marginalized communities in language models.

Drawing from J.L. Austin's Speech Act Theory and Suzanne Keen's Theory of Narrative Empathy, the research aims to provide a framework for game-makers to engage players in meaningful conversations with AI Non-Player Characters (NPCs) using large language models. The hypothesis suggests that language shapes how we perceive our relation to others, and aims to animate expressive qualities by holding players accountable for the words they use within the game. This project uses a Research-through-Design methodology which entails making a game, documenting the process and findings in a developer's journal, and finally synthesizing results to share best practice guidelines with other game-makers. These guidelines offer suggestions for implementing a natural language interface that takes into consideration the AI language model, the game-maker, and the player. Additionally, they provide a framework for facilitating a game-maker's expressive goals using the theories outlined in this project, and offer strategies for holding players accountable for their words by fostering meaningful actions in a game utilizing large language models.

The importance of this research lies in its aim to provide tools to underrepresented game-makers, via this document and publishing online, enabling them to leverage the affordances of large language models and allowing for the creation of their own unique stories.

Keywords: *artificial intelligence, games, language models, narrative-driven games*

Copyright Notice

This document is licensed under the Creative Commons CC 4.0 Canada License

<https://creativecommons.org/licenses/by/4.0/>

YOU ARE FREE TO:

Share — copy and redistribute the material in any medium or format

Adapt — remix, transform, and build upon the material for any purpose, even commercially.

The licensor cannot revoke these freedoms as long as you follow the license terms.

UNDER THE FOLLOWING TERMS:

Attribution — You must give appropriate credit, provide a link to the license, and indicate if changes were made. You may do so in any reasonable manner, but not in any way that suggests the licensor endorses you or your use.

No additional restrictions — You may not apply legal terms or technological measures that legally restrict others from doing anything the license permits.

WITH THE UNDERSTANDING THAT:

You do not have to comply with the license for elements of the material in the public domain or where your use is permitted by an applicable exception or limitation.

No warranties are given. The license may not give you all of the permissions necessary for your intended use. For example, other rights such as publicity, privacy, or moral rights may limit how you use the material.

Acknowledgment & Dedication

One of the wonderful aspects of embarking on this project is the wealth of knowledge and insight acquired in a relatively short period. The structure it provides and the invaluable guidance from faculty and peers are blessings, making the project feel achievable solely through the support received from the community, including my advisory team, Emma Westecott and Michelle Miller, for their invaluable insight and guidance.

To my primary advisor, Emma Westecott, who dedicated countless hours and opportunities to giving her students the best chance of success. I arrived at OCAD eager to learn, and I am thankful that you agreed to guide my independent study early in my time with the program, allowing me to create a small game about crocodiles and empathy. I have a thousand and one ideas for more little games to create, and I can't wait to work on them when all this is over.

To my family who brought me back home to them—I am so glad to be reunited after years of bouncing around from place to place. I hope someday I can grow roots and we can spend more time together. To my family in the Philippines, I miss you lots and I'm forever grateful for your care of me, I hope I make you proud.

To my friends here in Canada, the Philippines, and everywhere else in the world, thank you endlessly for your support and for listening to my late night rambles about all the ideas I want to do and doing maybe like ten percent of them. I'll do them someday, I promise, and then I'll make a little acknowledgment section like this and dedicate them to you.

Table of Contents

Abstract	2
Acknowledgment & Dedication	4
Table of Contents	5
List of Tables, Figures, & Illustrations	7
Definition of Terms	8
Chapter 1: Introduction	10
Project Overview.....	12
Research Question	13
Chapter 2: Literature & Contextual Review	14
Literature Review	15
1.1 AI Language Models	15
1.2 The Game-maker	18
1.3 The Player	20
Contextual Review	27
2.1 Language Models & Games.....	27
2.2 Prosocial and Expressive Qualities in Games.....	29
Summary of Findings.....	34
Chapter 3: Approach	35
Methodology.....	35
Research Through Design.....	35
Methods.....	37
Developer’s Journal	38
AI as Collaborator	38
Functional Testing.....	39
Conceptual Playtesting	40
Chapter 4: Creation	42
Chapter Overview	42
First Prototype	43
Requirements.....	44
The Context.....	45
Characters.....	45
Systems Breakdown	46
Gameflow Diagram	47
Iteration 1 Goal - Character personality, goal-oriented behaviour, sentiment analysis..	48
Iteration 2 Goal - Personality, Motivations, Situation, Intent Analysis	50
Iteration 3 Goal - Visual Design, GUI elements, tweaking the model	54

Evaluation and Reflection	56
Desired Behaviours	56
Complications	56
Unintended Effects	56
Final Thoughts: First Prototype.....	57
Second Prototype	58
Expressive Goals	58
The Argument.....	58
Iteration 1 Goal: Story structure, point-and-click exploration, auxiliary elements	60
Iteration 2 Goal: Two conversations, establishing relationships.....	62
Iteration 3 Goal: Beginning narration, interactable objects.....	65
Iteration 4 Goal: Addressing the cultural nuances	67
Evaluation & Reflection	70
Final Thoughts: Second Prototype	71
Chapter 5: Results & Discussion	72
Reflection & Synthesis	73
Developer’s Journal	73
The Lens of the Game-Maker + The Lens of the Player	74
Natural Language Interface — Collaboration between the LLM, the Game-Maker, and the Player	76
Facilitating Expressive Goals with the LLM	78
Keeping Players Accountable in a Game with LLMs: Environment, Narrative, Social Effect of Language.....	80
Conceptual Playtesting.....	83
Final Thoughts	88
Chapter 6: Future Work & Conclusion	89
Publishing the Journal	89
Workshops for Game-Makers.....	89
Conclusion	90
How This Project Fails	90
Making Kin with the AI Era.....	92
Bibliography	93
Appendices.....	96
Appendix A: Story Outline for Carmen Won’t Talk to Me	96
Appendix B: Recruitment Protocol	99
Appendix C: Questionnaires	101
Appendix D: Written Transcripts.....	105
Appendix E: Functional Testing	107

List of Tables, Figures, & Illustrations

Figure 1: Diagram of methods	36
Figure 2: Interaction Design Model (Rogers et al. 2013; Damasceno et al. 2017).....	43
Figure 3: Table of characters’ personality and motivations.....	46
Figure 4: Applied framework/architecture for AI Language Model integration with Unity.....	47
Figure 5: Gameflow diagram illustrating the three story beats.....	48
Figure 6: Results of connecting GPT-3.5 Turbo to Unity game engine.....	49
Figure 7: Sentiment analysis and assigning button color depending on results.	50
Figure 8: AI language model is able to roleplay and pick up on character inflections.	50
Figure 9: AI language model exhibits goal-oriented behaviour (waiting on password), prompting a positive response which communicates to Unity to turn the button a green color.	51
Figure 10: Requirements that need to be met through prompt engineering.	52
Figure 11: Instructions that were given to the LLM to satisfy each requirement.....	53
Figures 12-13: The AI language model executes the first story beat by initiating a three-person conversation, and later performs intent analysis in the third beat to discern the player’s stance in the argument.....	54
Figure 14: Early prototyping features a conversation between the player and two AI characters. In this scenario, the player is undecided and says “I don’t know”, which prompts a direct, probing response from the AI characters Trish and Leah, and an emotional, non-verbal response from Carmen.	56
Figure 15: Storyboard for Carmen Won’t Talk To Me	62
Figure 16: Graphic summary of major story points using Freytag’s five-act structure.	63
Figure 17: Group chat conversation.....	64
Figure 18: Highlighting scripted portions and AI responses.....	65
Figure 19: Journal messages and photo of protagonist and Carmen	66
Figure 20: Entering the memory from the journal	67
Figure 21: Photo of Carmen & protagonist	69
Figure 22: A rainy night in Manila, 2018	71
Figure 23: Memory scene – the classroom, the protagonist, and friends	72
Figure 24: Prompt engineering incorporating Filipino slang and Taglish	75
Figure 25: Physical journals.....	78
Diagram 1: Lens of the Game-Maker + Lens of the Player.....	79
Diagram 2: Framework for implementing a natural language interface	81
Diagram 3: Framework: Facilitating Expressive Goals with the LLM.....	83
Diagram 4: Diagram of the relationship between the three conceptual ideas.....	85
Diagram 5: Speech Act Theory and LLMs	86
Diagram 6: Keeping players accountable for their words in a game with LLMs (1)	88
Diagram 7: Keeping players accountable for their words in a game with LLMs (2)	89

Definition of Terms

In this section, I'll provide brief explanations for key terms frequently used in this document, offering clarity on their meanings within the context of this project.

Natural Language: According to linguist John Lyons (1991) in his book *Natural Language and Universal Grammar: Essays in Linguistic Theory*, natural language is “acquired by its users without special instruction as a normal part of the process of maturation and socialization” (p.1). Thinking with acclaimed linguist Noam Chomsky, Lyons offers a definition of natural language by sharing Chomsky’s ideas that the 'natural' can be defined as language that is innate to human psychology and a biological necessity. In other words, language is an inherent aspect of being human, and if humans did not have the capacity for it, we would not have acquired language at all. Natural language, therefore, is any language that develops without premeditation or instruction within human communities. This is in contrast to programming languages used by machines (for example, the C# language used by the Unity game engine) or fictional, constructed languages (for example, the *Quenya* language used by the high elves in J.R.R Tolkien’s *Lord of the Rings* series). For the purposes of my thesis, this project will use English.

AI Large Language Models (LLMs): Large Language Models or LLMs are a form of artificial intelligence that is trained on enormous datasets and are noted for their ability to perform general-purpose understanding and generation of text across a wide range of topics. Simply put, LLMs are excellent predictors of text, like a more powerful version of the auto-complete feature on our phones. Leveraging their training on large swathes of data, they are able to leverage algorithmic systems to predict the next word in a sequence. Today, LLMs are capable of a wide variety of tasks, ranging from answering questions, to summarizing text, to writing sub-par screenplays. Critically, however, LLMs are not neutral tools. LLMs don’t comprehend languages as we humans do, but rather, they consist of large strings of numbers. An AI engineer provides “weights” to certain numbers, such that the AI might favor this string of numbers over the other. Providing “weights” is the method by which a human AI engineer “trains” a language model to interpret those numbers. This training of an LLM by humans, coupled with its consumption of large amounts of internet data, constitutes a mark of human

intervention—resulting in a machine that is not neutral, but rather holding the values of its engineers and the data it uses in its training.

Prompt Engineering: Users of LLMs communicate with the model using a textual interface and natural language. Prompt engineering is essentially the act of framing text such that an AI language model can understand the instructions you give it.

Prosocial Values: Within the scope of this project, the concept of prosocial values in games revolves around enriching players' comprehension of certain subjects and sensitizing them to the challenges and perspectives of diverse individuals or communities, extending beyond their own lived experiences.

Prosocial values can be animated by inciting feelings of empathy in players, and by allowing them to inhabit the emotions of characters. In striving for prosocial messaging in games, there is an understanding that we cannot change our audience's behaviour or reception of our game (Keen 2006; Nissenbaum & Flanagan 2016; Belman & Flanagan 2013), but we can design our game to encourage relevant responses.

Furthermore, in recognition of the flexibility of values, and how game-makers express these values through the games they create, scholars Helen Nissenbaum & Mary Flanagan identify themselves as aligning with the *Canadian Charter of Rights and Freedoms* in their book *Values at Play in Digital Play*, which is one of the foundational texts of this project, and whose values resonate with me as game-maker. In particular, the fundamental right to “freedom of thought, belief, opinion and expression” as outlined under ‘Fundamental Freedoms’ (section 2).

Chapter 1: Introduction

When I first played *Papers, Please* (Pope 2013) nearly 10 years ago, I didn't quite yet have the vocabulary to explain how much or why it fascinated me. All I remember are the intense emotions I felt when I played the role of the protagonist, an immigration officer working at the border of an authoritarian regime. I was faced with the difficult decision of whether to feed my family or reunite a stranger with their loved one, even after identifying their fake passport. I remember feeling very torn and conflicted. Now though, I can understand and appreciate how *Papers, Please* makes us feel for its characters. The game is, in many ways, a reflection of the callous bureaucracy that ordinary people face when crossing the borders of countries. It highlights how this system divides families, creates stress and tension, and puts normal people in situations where they have to deal with forces beyond their control. In our actual, lived experiences, we may not have had the role of a border officer or a desperate traveller fleeing their country. However, we can still understand the desire to reunite with a loved one, or the pressures of maintaining a job in a difficult situation.

At the time I played *Papers, Please*, I became increasingly interested in a game's ability to increase our capacity to care. The game's designer, Lucas Pope, was initially inspired by the process that immigration officers underwent when reviewing documents, which led to the creation of simple yet intriguing mechanics (Alexander 2014). The strained relationship between the fictional countries of *Arstotzka* and *Kolechia* was built around those mechanics, and from there, players derived their own interpretations and connections to the real world.

Through its narrative, mechanics, aesthetics, and systems, playing games allows us to explore our fundamental relationship with the world, and make connections that perhaps weren't there before (Nissenbaum & Flanagan 2016). Above all, in my time playing games I was most fascinated by the characters and their ability to make us feel with them, how they motivate us to push forward in a game, and how through their actions, they become the vehicles of narrative and emotion. As a maker in the digital space, I was drawn to telling stories in ways that leverage the unique affordances of games.

It's 2023, and generative AI has suddenly become the forefront consideration seemingly everywhere. Everyone I know has had a brush with ChatGPT in some way or form. The education

system has been pulled into a frenzy, deciding whether or not to monitor and ban AI-assisted student work. The creative circles I'm a part of are getting increasingly worried—generative AI is evolving fast, how do we cope? At this point, the response has become, increasingly: the best choice is to engage (Roose 2023). How can creatives use generative AI to their advantage and explore its unique affordances? How can we teach ourselves to better understand AI as we are thrust into a new world where, potentially, AI tools are the norm?

The making of this thesis evolved out of a two-fold response: encompassing my love for games with their ability to increase our capacity for care, and my anxieties and excitement around AI language models. There are three main components in this project: creating a game with language models, documenting the process and findings, and synthesizing insights into best practices for underrepresented game-makers.

Underrepresented game-makers are a crucial focus for several reasons. Firstly, they are often overlooked and underrepresented in the games industry, and the lack of diversity among game developers results in a gaming landscape that fails to reflect the diverse lived experiences of various groups of people. This lack of representation perpetuates a cycle where certain demographics dominate the game development industry, further marginalizing others (Anthropy 2012). Furthermore, the labor of underrepresented groups is exploited in the development of generative AI and large language models and this exploitation is compounded by the fact that languages and cultures of marginalized groups often go unrepresented in these models (Edwards 2023).

This project is positioned to acknowledge those gaps by exploring how underrepresented makers like myself can understand and work with AI language models so that they may integrate the techniques outlined in this project into their own existing practice if they so choose, leveraging the unique affordances of working with LLMs.

The final result of this investigation is the creation of a story-driven game based on a fictional narrative that reflects on my own experience of queer, Filipino girlhood. The game allows players to converse with AI characters, influencing their behaviour and altering the game world and narrative.

Project Overview

In beginning this project, I was predominantly interested in the social effect of language, particularly in how it changes how we relate to ourselves or other people. As a queer person, there were words used to describe me that I would flinch at, but as the years go on I have grown comfortable with such terms and accepted them as a part of me. The way that we use our words, whether with callousness or with care, whether directed outward to others or inward to ourselves, can have such profound impacts on our environment.

This interest led me down a rabbit hole, cultivating a fascination with AI large language models. I was intrigued by their increasing ability to comprehend human language, not just in terms of what is being said, but also how it is being conveyed. Delving into this project prompted me to reflect on the achievements of *Facade* and how they continue to inspire, even more than 20 years later. In 2003, it took two experts in machine learning to create a custom system for *Facade*, enabling it to recognize player utterances and assign story beats accordingly. Today, companies like OpenAI have made AI platforms such as their GPT models (Generative Pre-trained Transformers, essentially general-purpose language models) accessible online. It was during this time, in the early months of 2023, when advancements in AI were propelled into the spotlight, increasingly becoming a captivating topic of discussion for numerous news outlets, capturing my attention as well. With the release of ChatGPT, anyone can now explore and interact with an LLM through their browser. I found myself imagining the narrative possibilities that could be created with AI non-player characters capable of understanding natural language.

Spurred by this inquiry, the makings of this project started as a proof-of-concept interactive sketch in the summer of 2023, developed under the guidance of Dr. Alexis Morris of OCAD University's Faculty of Arts & Science. Before this project, I had always been curious about how NPCs in games can increase our capacity for care. With LLMs, I wanted to explore how language can be used as a vehicle for players to enact that care.

The goal of this endeavour is to bridge game-making and natural language processing to investigate how players perform meaningful actions through words and utterances in a game environment. The conceptual framework borrows from Tanenbaum & Tanenbaum's articulation of J.L. Austin's Speech Act Theory and Suzanne Keen's Theory of Narrative Empathy, which expresses how *broadcast strategic empathy* (emphasizing common vulnerabilities and hopes)

and narrative techniques could invite empathy from readers by promoting *character identification* (feeling for a character's motives) through the *narrative situation* (an internal perspective).

The final work was the creation of a digital game in the Unity game engine, with an environment built using 2D graphics and production techniques. OpenAI's GPT language models played the role of AI characters whom the players could interact with using natural language (in this case, English), in a form of interactive drama.

The title of the game is *Carmen Won't Talk to Me*, a short game about connection and human relationships.

Research Question

For this project, the primary consideration and guiding question is as follows:

How might natural language interfaces with LLMs be approached to align with the expressive goals of underrepresented game-makers?

To explore this inquiry, the project used several methods. The literature and contextual review aims to situate this project in the domains of AI and game-making, and as a guide to orient its trajectory. The project employs a Research-through-Design methodology (Gaver 2012) in order to apply theory to praxis, using iterative methods to create the final game. Throughout this journey, findings, insights, and design annotations have been documented in a journal and synthesized to serve future game-makers.

Chapter 2: Literature & Contextual Review

With the advent of recent rapid developments in Artificial Intelligence, there is great concern over the harm that AI can do in disrupting industries and perpetuating toxic biases and prejudices (Eloundou et al. 2023; Irving & Askeil 2019). At the same time, generative AI captures the imagination with the creative possibilities of this technology. The significance of my research lies in empowering underrepresented game-makers to utilise this technology to imagine new ways of making in tandem with AI Large Language Models (LLMs).

My thesis research aims to contribute to the field of game studies, particularly in AI and machine learning applications, with natural language interfaces as a potential component of gameplay. The goal of my practice is to provide a framework for game-makers to facilitate expressive qualities, and rules of engagement with conversational AI NPCs (non-player characters) as a primary gameplay mechanic, inspired by the previous work of game designers and machine learning engineers whose work I unpack in this review.

There is much insecurity about how AI disrupts the creative industry, with valid concerns that AI will take away creative jobs (Eloundou et al. 2023). What I am proposing is one method that underrepresented creators can engage with AI and use it as a tool to create work that aligns with the expressive goals of the game-maker.

Literature Review

1.1 AI Language Models

Helen Nissenbaum and Mary Flanagan (2016) in their book *Values at Play in Digital Games* refer to Janet Murray's ideas on how games offer us an opportunity to exercise our fundamental relationship with the world and “our desire to prevail over adversity, to survive our inevitable defeats, to shape our environment, to master complexity, and to make our lives fit together like the pieces of a jigsaw puzzle.” (Nissenbaum & Flanagan 2016, 5) As game-makers, we have the ability to create systems and rules that are fabricated for the storyworld we create but are nonetheless, inevitably, reflections of the actual, ‘real’ world we live in.

Language is another vehicle through which we understand our relationship with the world. It is not only used to describe the world, but it can actively change and re-construct how we view reality itself (Boroditsky 2011). Natural language, language which develops organically within human communities, carries inherent relational properties intertwined with sociocultural norms and biases influenced by our environment. The way that we frame the world—with natural language—changes the way we perceive other people and objects around us.

In recognition of how language allows us to talk about ideas, express attitudes and emotions, and explore human relationships, Mateas and Stern (2003) designed a game system that recognized natural language input from players in the form of the interactive drama *Facade*. In the game, the player performs as a long-time friend of Grace and Trip, invited over for dinner at their place. As the player continues to interact with the two characters, it becomes increasingly apparent that their relationship is falling apart as the player gets entangled in their conflict.

In the paper *Facade: An Experiment in Building a Fully-Realized Interactive Drama (2003)*, the authors' work on *Facade* was a response to what they considered, at the time, a simplistic approach to computer-controlled non-player characters (NPCs). This approach primarily focused on a limited set of physical action abilities (run, jump, hit, shoot, etc.). This limited perspective became especially apparent when players engaged in spoken interactions with NPCs, which were restricted to providing preset options and phrases. This limitation hindered

the exploration of themes that the authors were interested in, such as human relationships and interpersonal conflict, with depth.

Recognizing the limited capabilities of game technology at the time, Mateas and Stern developed a system that considered the social impact of player input. They implemented what they termed 'shallow natural language processing,' which listened for various patterns of player utterances, such as words and phrases relevant to the dramatic situation of the game. The system recognised player utterances such as agreement, alliance, criticism, and flirtation, which the game responded to accordingly to support a meaningful, dramatic experience.

In 2003, the authors' system, specifically designed for *Facade*, was not always accurate in interpreting player utterances because it lacked support for general-purpose natural language understanding. In this way, players would experience a disconnect with the system which would result in a 'broken' performance. That said, *Facade* continues to stand as a remarkable work of interactive theatre that effectively demonstrates the unique affordances of natural language processing in games while maintaining the pleasures of drama and story.

We have made significant progress since *Facade*, particularly with the emergence of large language models in recent years. These models possess greater language proficiency, capabilities, and a better grasp on contextual-awareness. Increasingly, they are able to understand intent in human input.

Working with language models, however, presents significant issues. In essence, generative AI models, which language models are a part of, are algorithmic systems. These models identify patterns in data and provide generalizations based on a given prompt. The power of generative models lies in the training comprised of extremely large datasets scraped from the internet. Each piece of data is given a label from which the model can identify objects and patterns. This can be as simple as labelling a banana as 'fruit', 'yellow', or 'edible'. The problem arises in this labelling of data, which constitutes a significant mark of human intervention; since humans are by default flawed and biased, training data becomes inundated with the prejudices of the humans who provide them, which becomes evident in AI text-to-image and text-to-text generation (Irving & Askeel 2019; Samuel 2022). Scraping internet data can result in inaccurate or incomplete data, but it can also have significant consequences, and result in racist, misogynistic, or homophobic results.

My prior interest in merging Language Models (LLMs) with games predates this thesis, and the critical perspectives of scholars like Safiya Noble and others have been foundational to this work. In delving further into this topic, it is important to maintain a critical perspective when discussing AI generative models and the basis of their training: the internet.

Safiya Noble's (2013) paper *Google Search: Hyper-visibility as a Means of Rendering Black Women and Girls Invisible* investigates the harmful biases that exist in the internet's infrastructure, as reflected in Google search algorithms. Noble points to how Google search treats black women and girls by showing search results that hypersexualize them, stripping them of their voice to express and inform their own identity. In examining this infrastructure, Noble highlights how the internet serves as a conduit between dominant, prevailing ideologies that are coded into its engineering and how capitalist forces, such as the commercial interests of private companies like Google, benefit by complying with and amplifying existing power structures.

Google has taken steps to address the concerns raised since the publication of the paper in 2013. They now use a different algorithm that personalizes your experience based on your user profile and the data they gather from you. This has made the sordid aspects of the internet invisible to you, even though they still exist. However, with the proliferation of generative AI, the harmful biases on the internet have resurfaced.

AI language models have gained notoriety for unintentionally generating content that perpetuates racism, sexism, and homophobia. Attempts to rectify these issues by implementing word filters, like flagging terms such as 'Muslim' or 'gay', can inadvertently lead to the erasure of entire communities and identities (Samuel 2021). This challenge is commonly referred to by AI researchers as 'The Alignment Problem', underscoring the need for increased collaboration with social scientists and ethics researchers to advance the development of AI that aligns more thoughtfully with human values (Irving & Askill 2019).

In *Making Kin with the Machines* (Lewis et al. 2018), the authors provide a framework for approaching artificial intelligence that draws upon Indigenous epistemologies. This approach encourages us to engage with AI despite the anxieties often associated with discussing its potential harm. The authors emphasize recognizing technology's agency, a quality shared by all things, whether animate or inanimate. According to the authors, acknowledging AI's agency

involves a twofold process: comprehending its increasing influence in shaping our world, and as its creators, instilling within AI the values we hope to see in our society.

This project resonates with those perspectives, but it is essential to clarify the boundaries of AI agency. Presently, it is more precise to characterize AI as moving towards semi-autonomy. In other words, AI can make decisions, solve problems, and express values based on human instructions, with humans ultimately determining those values (Chiang 2023).

Connecting Lewis et al.'s ideas with my thesis, in a similar fashion, it is important to consider the role that game-makers have in controlling the values expressed in their games. As game-makers, we may not be AI scientists or engineers, but we can control how we choose to use AI tools in our work, and in doing so, we may advocate for further development into the ethical use of AI.

Building upon these themes, my research-through-design investigates how game-makers can work with language models to facilitate expressive qualities while maintaining criticality. The research endeavours to acknowledge the agency of all three parties involved: the AI, the game-maker, and the player.

1.2 The Game-maker

Ensuring that the games we craft with AI resonate with the values we hold involves an examination of how a game-maker's own principles are embedded within the games they create. It prompts me to explore some ways that a game perpetuates the values of its maker and which stories are shared.

In investigating this question, I turn to Aubrey Anable (2018) in her book *Playing with Feelings* as she brings affect theory and digital games into conversation. Looking at the two subjects with a feminist and queer lens, we can come to an understanding of digital games as 'affective systems'. Here, Anable references Sara Ahmed's definition of "orientation" to explain how our markers of identity—gender, race, class, and sexuality—shape our interaction with a digital game. She explores how bodies, computational systems, aesthetics, narrative, and sociohistorical contexts, "work together to give shape to feelings that exceed any one of these locations or modes of interpretation." (Anable 2018, xvii)

Anable highlights the discursive potential of games and emphasizes how game systems cannot be meaningfully separated from their representational qualities, and by extension cannot be divorced from their ability to persuade their audience, establish identifications, and convey meanings. Digital games offer a conduit to re-orient ourselves as feeling, sensing beings. They evoke feelings, bring meaning to the mundane, and establish us as relational beings. Thus, we cannot separate a game's computational systems from the politics of representation, as it is an avenue for us to relate to the subject of its narrative. Critically, affect informs our relation to our environment.

Anna Anthropy (2012) calls attention to this in her book, *Rise of the Videogame Zinesters: How Freaks, Normals, Amateurs, Artists, Dreamers, Dropouts, Queers, Housewives, and People Like You Are Taking Back an Art Form*. She provides a critical examination of the traditional model of game creation perpetuated by the digital game industry, one that is created by and caters to a narrow group (young, male, predominantly white) and calls for the empowerment of underrepresented voices to create games that speak to a wider range of human experiences. *Rise of the Videogame Zinesters* showcases how the medium of digital games has been used by marginalized creators in order to create game experiences that push against the fold, and how the democratization of game-making tools has paved the way for that possibility.

Helen Nissenbaum and Mary Flanagan (2016) further develop this idea of the game-maker having control over her game's values and messages. As a game designer's handbook, *Values at Play in Digital Games* provides a theoretical and practical framework for identifying socially recognized values in games, as well as providing guidelines for designers to create games that aim to animate social, political, or artistic expression. The authors highlight three core premises to the book: that groups of people tend to have common values, that technologies embody these values, and that designers and game creators have the ability to "shape players' engagement with these values."(Nissenbaum & Flanagan 2016, xii)

The orientation of the game-maker, invariably informs the messages and meanings behind what they make, manifesting in the game's mechanics, the narrative, and the ways in which players are allowed to engage with the storyworld and the characters that inhabit it. Game-makers then have the profound ability to shape meanings in their games and craft an experience that resonates with their audience.

To refocus attention on the core question at hand, which is how might natural language interfaces with LLMs be approached to align with the expressive goals of underrepresented game-makers, I'd like to revisit the central inquiry of this thesis. The main tenet of this project is to look into the ways in which game-makers can facilitate expressive qualities in games, using large language models. Having established that digital games are affective systems that have the capacity to persuade, create identifications, and circulate meanings (Anable 2018). Furthermore, a game-maker's orientation—their identity, such as race, class, and sexuality— informs the affective capacities of their games, and groups of people tend to have common values (Nissenbaum & Flanagan 2016), expressed in the games they create. As the mainstream game industry predominantly reflects the perspectives of cisgender, white, male individuals, games often cater to this demographic. Therefore, the democratization of game-making tools and the self-publishing of independent games are crucial for marginalized individuals to share their creative work and offer their unique perspectives, contributing to their ongoing cultural significance.

If we want to have games that exhibit a range of human experiences, of different stories and meanings, it's important to give underrepresented voices a platform to tell stories through digital games (Anthropy 2012). These considerations become especially significant when game-makers align with language models, recognizing that these models inherently carry their own flaws. Invoking *Making Kin with the Machines*, which states that it is our responsibility to determine the values expressed by technology, it is therefore the responsibility of game-makers to shape players' engagement with language models and ensure that the use of AI aligns with their expressive goals.

1.3 The Player

This next section acknowledges the agency of the player and looks into the ways the game-maker can effectively communicate her expressive goal or value message to resonate with the player, and some of the ways where that can fail.

Prosocial expression in digital games has been characterized by its capacity to elicit empathy and raise awareness of certain issues in players (Nissenbaum & Flanagan 2016). For

these authors, the objective of a game with prosocial values is to encourage players to engage their beliefs, biases, and feelings, while also enhancing their understanding of specific topics.

Before delving into empathy induction in games, it's important to first consider the contested nature of empathy in this context. Firstly, it must be acknowledged that techniques in empathy induction have the potential to cause harm to the player, through powerfully overwhelming feelings in the form of 'empathic overload'. This can occur, for example, by allowing players to experience a character's trauma in a VR game without the informed consent of the player (Dunlap 2019). Furthermore, empathy can also be exploited to invite harm to others or to incite hate towards another group or individuals of a certain background (Chesler 2022).

Secondly, scholars acknowledge that empathy is not a precursor to prosocial or altruistic behavior (Dunlap 2019; Keen 2006). Empathy essentially entails the ability to project oneself into another person's perspective, but it does not necessarily mean that one has actually experienced that person's lived reality. It is often more helpful to frame empathy as raising awareness or increasing players' appreciation of the lived experiences of others, or as a means of fostering recognition of commonalities between players and other groups of people. These framings are advocated by the scholars discussed in the literature review, who ground their approach in fields outside of game-making.

In a prior collaboration, Mary Flanagan, along with co-author Jonathan Belman (2013), delved into the realm of psychological research and its revelations regarding empathy induction. They applied these insights to the domain of game design. In their work *Designing Games to Foster Empathy*, the authors explored how game designers could incorporate these insights by outlining a series of design principles aimed at fostering empathy in their intended audiences. One of these considerations is that in order for messaging to be effective, as designers we must ensure that both the cognitive and affective components are engaged by our players. The work should aim to be both informative (regarding its subject) and emotionally resonant. The authors highlight that emotions play a key role in allowing our audiences to relate with the people or groups with whom they are encouraged to empathize. This is done through perspective-taking, which is one of the core components of Suzanne Keen's Theory of Narrative Empathy (2006).

Similarly to Belman and Flanagan, Keen uses an interdisciplinary approach drawing from psychology, neuroscience, narrative theory, and affective science in order to investigate how

narrative induces empathy and inspires altruistic thought in the reader. Both articles maintain that for empathy to happen, it is important to engage both the cognitive and affective responses from the reader. According to Keen, authors attempt *strategic empathy* as a framework to “direct an emotional transaction through a fictional work aimed at a particular audience, not necessarily including every reader who happens upon the text” (Keen 2006, 224).

Keen presents three varieties of strategic empathy. *Bounded strategic empathy* occurs within an in-group encouraged to feel for familiar others within shared experiences. *Ambassadorial strategic empathy* “addresses chosen others with the aim of cultivating their empathy for the in-group, often to a specific end.” *Broadcast strategic empathy* invites every reader to empathise with members of a group by “emphasising our common vulnerabilities and hopes” (Keen 2006, 224). When employing strategic empathy, Keen further outlines two narrative techniques that authors use in empathic writing that have strong parallels in narrative-driven games. The first narrative technique is *character identification*, which allows readers to feel for characters’ motives, facilitating affective engagement. The second, *narrative situation*, provides an internal perspective to the reader, stating that a first-person view of the narrative allows the reader to “travel with the character” and be privy to the internal workings of a character’s mind.

I find that these two techniques align with the unique affordances of narrative-rich games. Firstly, in terms of *narrative situation*, controlling a character in a game often involves a first-person perspective, occasionally supplemented by the protagonist's narration, providing insight into their thoughts and desires. Secondly, concerning *character identification*, the ability to control a game character is closely linked to their motives, as manipulating a character's actions can be seen as fulfilling their motivations within the game's narrative.

In making sure that expressive goals are maintained in narrative-rich games, it is important to uphold authorial intent vis-à-vis player agency. Tanenbaum & Tanenbaum (2009) address this tension between player agency and narrative-driven games in the paper *Commitment to Meaning: A Reframing of Agency in Games*. The authors analyze and challenge the game industry’s assumption of the notion of agency in games, shifting the definition from one of *choice* and *freedom* towards a framing of agency as a player’s *commitment to meaning*. The paper aims to understand how we come to enjoy narratively rich games, addressing the tension that exists between player choices and authorial intent. The authors outline their argument using J.L.

Austin's Speech Act Theory (2003) and its definition of 'illocutionary acts' using it as a framework to analyze interactive storytelling. Simply stated, a *locution* is what is meant by the utterance, *illocution* is the action that was performed, and *perlocution* is what happens as a result and its effect on the receiver of the utterance. Through role playing and performing actions as that character (*illocution*), the game responds (*perlocution*) and keeps the player accountable for their actions, which in turn keeps player actions meaningful. What occurs can be seen as a form of improvisational theatre, where the player enters a social contract with the game designer that transforms them into a collaborator and active participant in the story.

The authors suggest that as designers we must align the joy of play with the satisfaction of storytelling. By better understanding how narrative facilitates a player's experience, we will be able to some aspects of pleasurable gameplay imbued with meaning that also involves the pleasure of story (Tanenbaum & Tanenbaum 2009, 1). To support this, the player *must be able to perform meaningful actions*. In particular, whenever a player does something or says something, it is important that the game *responds* to player input and keeps them accountable for their actions. In this way, we are able to create meaning for the player.

Tanenbaum (2011) expands on their previous paper in the article *Being in the Story: Readerly Pleasure, Acting Theory, and Performing a Role* which emphasizes the reader's pleasure of transforming into a character in the narrative, rather than having authorial control over the story's outcome. The paper draws inspiration from arguments in performance theory, particularly in method acting by re-imagining the player as an "actor in a scripted drama" (Tanenbaum 2011, 58), with pre-written lines, committing to the actions, desires, and emotions of the player-character. Tanenbaum states:

By instead imagining the interactor as an actor playing a role within a play, the pleasure becomes a participatory, transformative pleasure where the interactor becomes a character and experiences that character's emotions and desires instead of her own (Tanenbaum 2011, 58).

Tanenbaum uses these arguments to frame their writing on bounded agency and how players come to enjoy narrative-driven games despite not having full control over its trajectory.

Building on this work, Tanenbaum draws from *Mass Effect II* (BioWare 2010) to illustrate how the player is given a well-defined player character, and while not having control over the

story's outcome, the player has agency in how they reach that outcome and achieve the core protagonist Shepard's goals. During the game, players are given the ability to shape the character Shepard's personality through a dialogue wheel. The dialogue wheel offers different inflections that represent the attitudes that Shepard can exhibit in a given scenario. Based on the choices made by the player, Shepard can lean towards a Paragon (noble) or Renegade (rebel) alignment, or display traits of both. The game designers recognise that although players might not have agency over the story's trajectory, they are able to control *how* they get there by controlling Shepard's attitude. Providing a well-defined character is crucial to guide the player's performance in narrative-rich games. To articulate the goals, desires, and emotions of the character, the player must understand the character's inner psychology in order to speak with the character's voice.

Critically, for a game that uses language models to make sense of player input, the AI must comprehend their intentions. If the system is unable to grasp the player's aim or comes in conflict with it, that meaning is lost, which Tanenbaum mentions comprises a failing of *Facade* (2003). Providing the player with unrestricted agency can result in meaningless capabilities. According to Tanenbaum, it is better to offer the player a limited set of choices that the system can recognize and react to meaningfully. Balancing meaningful capabilities is important when creating games with language models, as the model's response may conflict with player input.

In the article *Mass Effect 2: A Case Study in the Design of Game Narrative*, Tanenbaum and co-author Bizzocchi (2012) continue to build upon previous writing regarding narrative-driven games and redefining player agency in story-based player experiences. In particular to character, the authors highlight how "identification with believable characters is one of the keys to designing games that support empathic emotion." Non-player characters are critical to supporting the narrative experience. The game designer will have to combine visual design, goal-oriented behaviour, and "an expression of an underlying psychology" in order to present characters with depth and believability, further supporting empathic emotion (Bizzocchi & Tanenbaum 2012, 396). This emphasis on character identification and the importance of an expression of an underlying psychology are congruent with the principles outlined by Keen's *Theory of Narrative Empathy* (2006).

Both Tanenbaum and Keen speak to this transformative quality of narrative. They contend that under certain conditions, such as enabling readers to gain insight into a character's inner psychology through access to the protagonist's internal monologue, which includes their thoughts, desires, and motivations, perspective-taking and character identification can occur.

Games offer unique affordances as an interactive, narrative medium; in games, a protagonist's actions become our own (Tanenbaum 2011). The computational dimension (mechanics) of games becomes the platform by which we are transformed into the character, cognitively and emotionally (Nissenbaum & Flanagan 2016, Tanenbaum 2011, Anable 2018). Not only do we have access to a character's internal monologue, we go one step further and become *enactors of their will*.

In ensuring that the experience as received by the player is aligned with the values that the game-maker promotes, she must be conscientious of the rules of engagement of her game (Nissenbaum & Flanagan 2016). It is important to consider what actions players are permitted to take in the game world and how they are allowed to interact with non-player characters. For instance, how can a game that arms the player with a gun aim to invoke a message of peace? The authors maintain that the game-maker must limit and guide player action through mechanics, narrative, and aesthetics, in order to foster plausible and relevant interpretations by the player.

Nevertheless, various authors contend that irrespective of the author's intent, the audience will not perceive the message uniformly (Keen 2006; Nissenbaum & Flanagan 2016; Belman & Flanagan 2013). A player's interpretation of a message is subject to numerous factors, including their individual beliefs, worldviews, and values. Drawing from Anable and Ahmed, just as the game-maker's orientation changes the affective charge of her game, the inverse is also true: the orientation of the player informs how they relate to the game.

In summary, working with language models offers unique possibilities and challenges for the game-maker. We can imagine a method wherein the game-maker collaborates with AI, designing an experience which allows players to translate their expressive messages into actions through their words (Tanenbaum & Tanenbaum 2011; Austin 2003) and keep them accountable for what they say—in natural language.

We've come a long way from the days of *Facade* in 2003. As language models increase their understanding of human sensibilities, the technology provides a wider arsenal for game-makers (Fraser et al. 2018; Park et al 2023; Zhu et al 2023). Balancing the game-maker's expressive goals with the inherent flaws and biases embedded in language models when using them to convey messages in games, and how the game conveys these messages through a player's actions, presents a notable challenge. This balance can be conceptualized as a three-way conversation that recognizes the agency of the AI, the game-maker, and the player.

Contextual Review

2.1 Language Models & Games

In this section, I will explore the intersection of LLMs and games, examining both the utilization of LLMs within games themselves and LLM services tailored for game-makers, and how they relate to this project.

In creating the game *Facade*, authors Mateas and Stern explore intersections between games and improvisational theatre, and how “language allows us to talk about ideas and express attitudes and emotions” (Mateas & Stern 2003, 1-2), which the authors state are key requirements in designing interactive experiences that focus on human relationships. *Façade* (2003) is one of the earliest games to use shallow natural language processing as an interface for players to engage with embodied agents (non-player characters) in a 3D environment. To facilitate these features, the authors created their own coding architecture to manage in-game characters and annotated hundreds of possible player inputs to train a machine learning system to understand a user’s natural language. In discussing the narrative capabilities of the game, the authors highlight the importance of using a system that recognizes the “social effect of player utterances” while balancing that with tight authorial control (Mateas & Stern 2003, 2). This is helpful for my research as it highlights the importance of structured narrative and tight authorial control in writing drama in games, in order to provide a well-formed, meaningful experience to players. With the emergence of accessible and comprehensive language models like GPT, game designers can explore new possibilities by incorporating these models into NPC behavior. This allows players to engage in free-form conversations with characters, using their own words, while the NPCs demonstrate a level of general natural language understanding.

The paper *Spoken Conversational AI in Video Games: Emotional Dialogue Management Increases User Engagement* (Fraser et al. 2018) provides an example of how development in LLMs are being used to facilitate new player interactions. This paper is a study into how conversational agents with integrated emotional dialogue management result in higher player engagement (measured by time spent playing) and overall satisfaction in a role-playing game. The researchers compare the results of an experimental group where players interact with AI game characters with emotional dialogue management and compare these results with a control group that interfaces with AI characters without these features. The system uses sentiment

analysis and emotional modeling and requires players to actively engage in emotionally manipulative conversation to reach certain game goals. The findings of this paper are helpful for my research for three key reasons: The first is that the paper outlines a method of employing emotional dialogue management, having the LLM conduct sentiment analysis on player input and applying emotional modelling to their NPCs. Their approach has been helpful in building my own system. Second, implementing in-game goals while allowing players to interface using natural language helps mitigate open-ended, drawn-out conversation that tends to be the trappings of chatbot agents. Third, emotional dialogue management and in-game goals encourage certain emotionally guided behaviours from the player while discouraging others. This key aspect of guiding player behaviour is worth looking into for designing values-embedded play, by crafting the rules of engagement so that players naturally pursue in-game objectives through actions that align with the maker's expressive goals.

Seeing the potential of integrating LLMs in games, AI platforms seek to capitalise on this innovation by creating paid and easy to integrate tools for game-makers. One such service is provided by Inworld AI, a low-code, free-to-try commercial platform for creating AI non-player conversational agents using their website [Inworld Studio](#). The platform also provides an online arcade in order for players and game creators to try out and converse with their AI characters. Inworld AI further allows for integration into a wide range of game engines, so that game designers can implement the “next generation of NPCs” into their games (Inworld AI 2023). The company provides their own research metrics that state that more and more players desire conversational agents in their games, and how these NPCs can make games more enjoyable. Inworld AI is a good starting point to explore their engine and test it’s capabilities and limitations, and gauge how their platform could potentially be used for prototyping games with language models.

[Laika](#) is an AI creativity tool that helps writers think with ‘brains’, providing inspiration for writing, inspiring fresh ideas and encouraging a multifaceted approach to their work. With Laika, writers can even train their own AI 'brains,' to create an AI that thinks with the voice of the writer. This process necessitates around 10,000 words of text to effectively train the AI brain, enabling it to grasp the writer's unique style and anticipate patterns within their work. The creators of Laika come from a background in digital game research, and their ambition is to integrate Laika into the game development pipeline as a collaborator for writers.

What makes Laika truly exciting is that it empowers writers to train language models using their own compositions, granting them significant control over the AI's tone and style. For game writers, this means more consistent and dependable results when employing Laika in their projects. This would provide writers with a reliable tool to ensure that the AI's language aligns with the game-maker's intended goals and vision, by giving writers greater control over the script and language that the AI draws from—and consequently, the ideas, values, sentiments, embedded in them.

2.2 Prosocial and Expressive Qualities in Games

This section will examine three games that provide emotionally resonant experiences for players. I analyze these games referencing theories regarding the expressive qualities of games, such as those proposed by Nissenbaum & Flanagan (2016), and narrative empathy, as discussed by Suzanne Keen (2006). This analysis delves into two games, *Hush* (Antonisse & Johnson 2008) and *Dys4ia* (Anthropy 2012), examining how they achieved expressive qualities and crafted emotionally resonant experiences for their audience. Additionally, it will explore one project, *The Last of Us Part 2*, which, for some players, failed to connect with its audience.

The game *Hush* by Jamie Antonisse and Devon Johnson (2008) places players in the perspective of a mother with her baby during the Rwandan genocide of 1994. The game begins by prompting the players with the following message:

Rwanda, 1994: The Hutu are coming, Liliane. Hide your child. If you falter in your lullaby, he will grow restless. The soldier will hear him, and he will come for you.

The player, as the mother, must sing to the rhythm of the lullaby in order to pacify their child, lest a soldier discover where they're hidden. This is reflected in the mechanics of the game: players must type in time with the notes that appear on-screen.

In playing *Hush*, players have reported powerful feelings of tension and dread when they falter in typing out the notes. Belman and Flanagan (2013) pointed out that the game is capable of creating a strong sense of parallel empathy through the unique interaction design used in *Hush*. Typing out notes in a lullaby strongly simulates what the character is depicted as doing.

Hush has the ability to evoke empathy in players by enabling them to experience a character's internal perspective. This is achieved through a combination of game mechanics and

the narrative situation, allowing players to identify with the character and feel what they feel. The mother's fear and anxiety can be effectively conveyed to players through the game's mechanics and story, eliciting a strong emotional response.

As game-makers, it is important to elicit empathetic emotion in players to encourage prosocial attitudes. Through *character identification* and the *narrative situation* (Keen 2006) and the effective design of game systems, aesthetics, and narrative (Belman & Flanagan 2013, Nissenbaum & Flanagan 2016), players are invited to feel issues or situations removed from the players' sociocultural contexts.

Dys4ia by Anna Anthropy (2012) is an autobiographical game that recounts her experience of gender dysphoria in a six-month period as she undergoes hormonal replacement therapy. The game touches upon the frustrations that entail her experience; the identity politics, prejudices, and personal growth that the author wrangles with. A relatively short game with roughly fifteen minutes of playtime, the game's mechanics and aesthetics are able to convey complex themes through simple, playful interactions.

What is remarkable about *Dys4ia* is its ability to invite its audience to connect with its subject matter, despite players' potential distance from the author's own experiences, much like games such as *Hush* and *Papers, Please*. The game is able to engage with the cognitive effects on its players by providing the audience with an internal perspective of a person in gender transition. By voicing her own frustrations with this process, an emotional dimension is activated as players learn of the author's fears, desires, and motivations. Although a uniquely personal experience of a woman in transition, the game speaks to the experiences of other people in the transgender community—and more broadly, speaks to the human experience and feelings of frustration that come with one's own body being in conflict with societal expectations. In this throughline, audiences from different backgrounds are able to relate to and identify with the author's struggles and motivations. *Dys4ia* achieves social messaging by providing an informative and emotionally moving experience—shifting players' feelings about its subject matter and deepening their understanding of a community of people that may be outside their own in-group.

Both *Hush* and *Dys4ia* are remarkable in their subject and execution, and one major aspect that differentiates these two games as compared to the next one we will discuss, is that they

were created by very small teams (in the case of *Hush*) or individual authors (*Dys4ia*). Smaller teams and individual creators are helpful in persisting the creative vision behind their game, and according to Anthropy, is what bogs down a lot of big budget mainstream games that often have very large teams that are working on the same title. Instead, Anthropy advocates for “more personal games, more relevant games, more games with something to say” (p. 178) with which we could have more coherent ideas behind games instead of a big team working on a game separately.

Unfortunately, the increasingly lucrative industry of mainstream game development prompts the expansion of staff, with more game-makers striving to work for these big companies. However, as high-budget games are developed, the high stakes also lead to massive layoffs when projects fail to meet financial goals (Carpenter 2024). Consequently, these high-stakes games can sometimes result in products that are either overly safe or muddled in their messaging, although there have been notable outliers in mainstream game development.

Over the past years, the expressive qualities of storytelling in games have become increasingly recognized in the same way that TV and film have for decades. In 2013, Naughty Dog's *The Last of Us* (2013) achieved widespread critical and commercial success, prompting other big-budget games to prioritize narrative on par with aesthetics and gameplay (Martens 2023). This development marked a significant shift, where storytelling that could dabble in complex matters of social issues and representation, became as important as other elements of game design. The significance of social messaging and representation in games is not a new concept; for decades, independent game-makers have been exploring heavy, complex themes in their games. This trend has now become an integral part of several big-budget game productions, as seen in titles like *Detroit Become Human* (2018) and *God of War* (2018).

This investment in narrative brings to the forefront its representational subjects prime for examination. As big-budget games tend to have relatively large audiences, it is worth pursuing an examination of their impact in persuading audiences, facilitating identifications, and circulating meanings (Anable 2018). Following the success of *The Last of Us*, the highly anticipated sequel carried grand expectations for its narrative. The sequel evoked strong emotions as was expected, but it was also fairly shocking and created a great deal of division among its audiences (Sheehan 2021).

The *Last of Us Part 2* (2020) begins by establishing the relationship between the protagonist, Ellie, and her relationship with her father-figure Joel as they attempt to build a life as a family in the midst of post-apocalyptic zombie America. Within the first few scenes, however, something terrible happens and Joel is killed by Abby, a woman from another community whose father Joel had killed. Unable to stop it, Ellie swears revenge and the player is brought along on the protagonist's bloody, violent, desperate quest to avenge the murder of Joel.

The action gameplay of shooting enemies, gathering resources, and killing her opponents—members of Abby's community and zombies alike—are rote for the genre. In a shocking plot twist midway through the game, the writers shift the perspective, and players gain control over Abby, a character they were conditioned to hate. Abby's story depicts the events that occurred before Joel's murder. Players get to meet her community and the people she cares about, who have been targeted by Ellie in order to reach her destination. The story narrates Abby's memories of her father and the grief she experienced after his murder. Through playing as Abby, we get to experience her goals, desires, and motivations, and we come to realize that Abby's revenge is as reckless, justifiable, and senseless as Ellie's.

The game communicates social commentary on the vicious cycles of revenge that humanity faces and how violence begets more violence. It is an important lesson to make, and the game puts these moral quandaries in the hands of the players who are asked to play both sides of the story, seeing the perspectives, merits, and downfalls of both. In discussing *The Last of Us Part 2*, I wanted to touch upon the highly divisive reception to the game's narrative direction. Bringing Anable to the conversation, the game tests our affective attachments to narrative representations. The reason why we, as the player, have strong, emotional aversions to playing as Abby is because the game has established us as belonging to Ellie and her community. We have traveled with Ellie, become privy to her internal workings and inner psychology, and we've come to love her. By shifting the perspective to Abby the game forcefully re-orientates the player and invites us to identify with Abby's struggle instead. Some find this shift to be compelling storytelling, while others feel betrayed and angry.

In considering the divisive reception, it is important to reflect on the player's orientation before the narrative shift. Regardless of the author's intentions, the audience will make their own

interpretations. These interpretations are informed by a variety of factors, some of which are the player's own personally held values, beliefs, and worldviews. We can mitigate this dissonance by highlighting the similarities between the player and the character they are encouraged to empathize with—but we must not expect audiences to receive a message in the same way, and it is an important consideration for the game-maker as she strives to craft an experience that allows for plausible and relevant interpretations (Flanagan & Nissenbaum 2016).

Summary of Findings

In the literature review, I explored AI language models and criticality, and the unique affordances of natural language processing to explore themes of connection and human relationships. The expressive qualities of games created by the game-maker were also examined, with a particular focus on values-based play and the facilitation of expressive goals by underrepresented makers. Additionally, consideration was given to the player and how to effectively communicate expressive qualities to them while keeping in mind player agency.

In the contextual review sections, I analysed existing games and platforms that employ natural language processing, with the aim of gaining insights into how to incorporate similar techniques into my own game. Furthermore, three games that aimed to convey prosocial and expressive qualities were studied, investigating their effectiveness in fostering player empathy through affective gameplay.

The next section explores Research through Design (RtD) methodology (Gaver 2012) and its application in investigating the primary research question of this thesis, namely, how underrepresented game-makers may approach LLMs and integrate them into their existing practice to serve their expressive goals.

Chapter 3: Approach

Taking what I've learned in reviewing literature and related projects, the next step is to distill that knowledge in the creation of a prototype that seeks to bridge LLMs and the game-maker while maintaining her expressive goals. For this purpose, this project employs a Research through Design (RtD) methodology that involves an iterative process that aims to gather qualitative data through the creation of the game itself. RtD borrows from development practices common in human-computer interaction (HCI) spaces (Gaver 2012), particularly in software development and game design.

Methodology

Research Through Design

In the article *What Should We Expect from Research through Design?* (2012), Gaver acknowledges that the discursive quality of design approaches should be recognized as a strength rather than a hindrance. The author encourages a diverse approach to design research processes, instead of confining them to rigid standardization and verifiable theory. Gaver suggests that designers should support multiple conceptual perspectives by borrowing from multiple domains, joining, subverting, and differentiating these ideas through design artifacts. According to the author, "it is the artifacts we create that are the definite facts of research through design" (Gaver 2012, 9).

In game development, creating design artifacts through an iterative process is essential to game-making. Jesse Schell in their book, *The Art of Game Design: A Book of Lenses* (2008), emphasizes the significance of rapid prototyping. Schell refers to this as "The Rule of the Loop," which means that the more the game designer iterates on her game, the better the game will turn out to be. For each loop, Schell highlights the importance of answering a question that addresses the conceptual ideas of the designer.

Acknowledging the unstandardized nature of the design research process as articulated by Gaver, I will be incorporating methods from HCI and game development disciplines,

particularly iterative prototyping, developer's journaling, and conceptual playtesting, and adapting them for my research to address relevant questions.

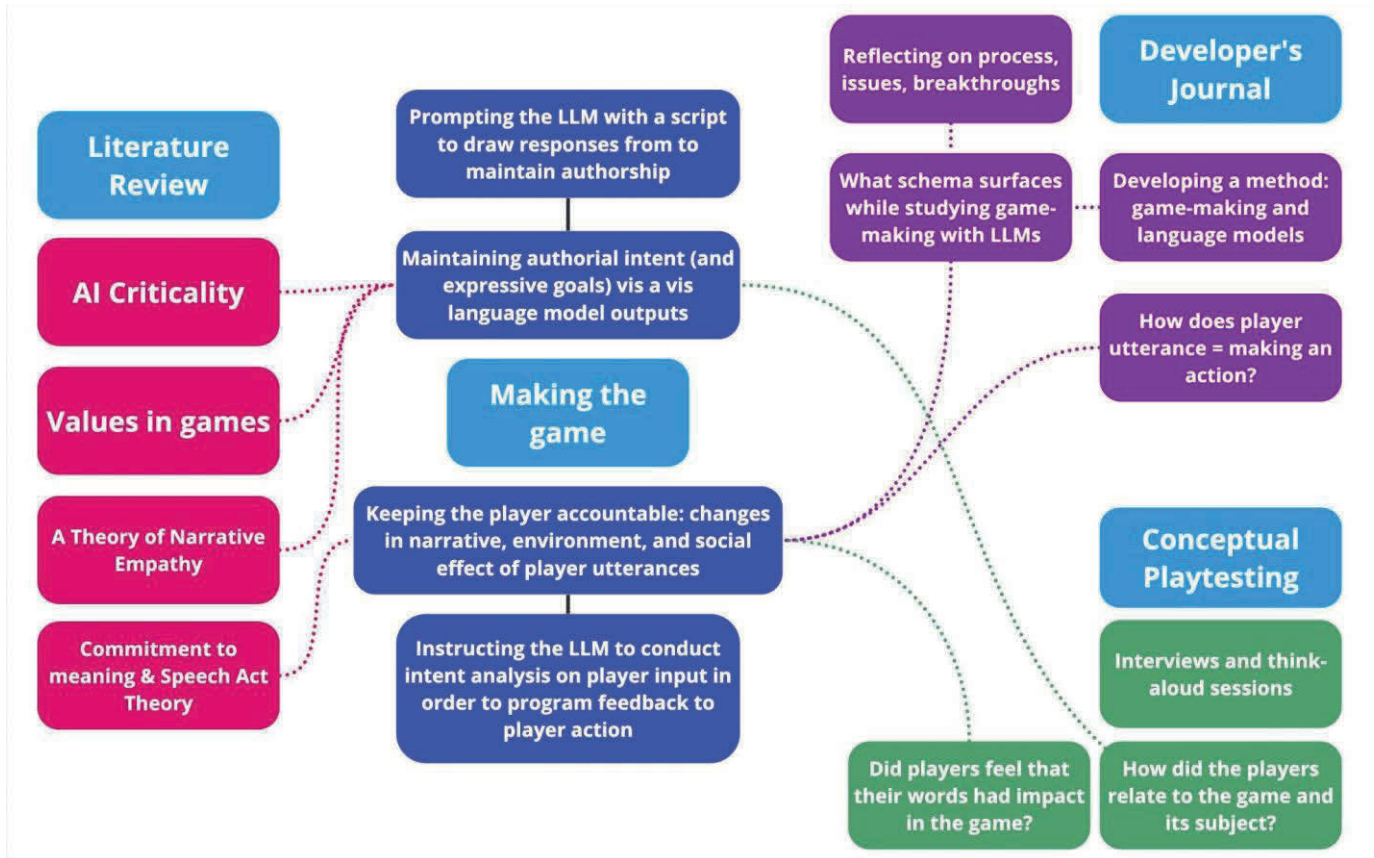


Figure 1: Diagram of methods

Methods

In order to answer my research question, which investigates how we might approach LLMs unique affordances to fulfill the expressive goals of underrepresented game-makers, this research uses the following methods to guide its creation. Following the theme of this thesis, each method corresponds to working through the agencies of the three parties involved: the game-maker, the AI language model, and the player.

The developer's journal will be the tool of the game-maker, its purpose of which is to guide the process of iterative prototyping by connecting conceptual annotations to specific design decisions in the prototypes I create.

The second method involves collaborating with the AI language model. This entails getting to know its abilities and limitations, and prompting the model for my specific needs and to assist in fulfilling the expressive goals of my game.

Finally, the third set of methods entails working closely with the player through functional and conceptual playtesting. Testing achieves two things, which is to evaluate the performance of the game and second, to determine if the expressive goals were indeed met.

Developer's Journal

Documenting the process of creation is essential to satisfy the goal of this research, which is to pave a way for future game-makers to utilize AI language models in their work.

In the spirit of preserving and disseminating knowledge, therefore, creating iterations goes hand in hand with documenting the process through a developer's journal, also referred to as an "annotated portfolio" in Gaver's *What Should We Expect from Research through Design?* (2012). This journal should connect "conceptual annotations with specific design examples", by reflecting on a designer's ideas, processes, issues, and breakthroughs (Gaver 2012, 10).

Through creating these annotations, at the end of each prototype, I am able to step back, look at each design decision and consideration and identify a schema or framework that stands out during the game-making process, which I write about in [Chapter 5: Results & Discussion](#). The goal is to create a framework that can act as a guide for future game-makers, in the hopes of guiding their approaches based on the methods used in my project. Process work can be found in Chapter 4: Creation, while frameworks that surface are highlighted in Chapter 5: Reflection and Synthesis.

AI as Collaborator

Large Language Models were close collaborators throughout the duration of this project. On the surface level, LLMs were used to fix grammatical errors especially in the aftermath of late-night document writing sessions. At a deeper level, LLMs were utilised to support the primary gameplay mechanic, which is to interface with parts of the game through natural language conversation. This is achieved through what is colloquially termed 'prompt engineering', which gives the LLM a set of instructions to follow.

During my collaboration with the AI Large Language Model, I focused on two main objectives. Firstly, I worked on breaking down my own barriers with technology, namely with coding simple behaviours, and secondly, I leaned into the unique capabilities of LLMs to make a game that uses natural language to interface with the game.

When I started working on this project, I was new to C# programming, the language used in the Unity Game Engine. In order to support my learning, I collaborated with the AI to help with coding simple behaviors in the game, such as making a sprite change transparency or turning a behavior on or off.

Finally, by leaning into the unique affordances of LLMs, I was able to invite my players to explore the nuances and social effects of language. This is achieved by having the LLM conduct intent analysis on player utterances, and ascribing certain game outcomes depending on the words that players say in the game.

After surveying multiple LLM platforms (including Inworld AI, which was mentioned earlier in this document), I've decided to work with two of OpenAI's language models, namely the free web platform ChatGPT and the paid service GPT-4. My criteria for choosing these two was flexibility of use, and cost.

Several LLM platforms that were designed to be integrated with games often have a subscription based service that charges users per month, with a usage limit that incurs additional charges after that limit is reached. They also come with a lot of out of the box features and easy integration. Despite these features, I decided not to go with these platforms because I found their system inflexible for my goals and I was hesitant to pay for a subscription based service. One of the tenets of my project is to share my process with underrepresented game-makers so that they may see the opportunities of making games with LLMs. To reach as many makers as possible, it was important to ensure that this project remained affordable without compromising its outcomes. By catering to the widest audience, I believe more creators will find it easier to integrate these techniques into their own projects. Therefore, cost considerations were of paramount importance.

Functional Testing

While Mateas and Stern were creating *Facade* (2003), one of the most important considerations was maintaining authorial intent with the adaptability of their AI agents that can recognize player utterances. A game with language models should be robust enough that given the wide range of possible utterances by the player, the AI NPCs should be able to maintain a narrative throughline and not deviate too much from the story to the point of incoherence.

Therefore, it is not enough that I am the only one testing the language model with my words, but rather it would be necessary to collect a variety of responses from other participants. To achieve this, I created a survey to collect diverse responses from 12 participants (see [Appendix C](#)) and used the data to test the functionality of the language model I am working with.

The online survey form invited participants to roleplay and take on the role of a high school girl caught in an argument in a classroom. The form presented two social scenarios and prompted each participant to offer a positive response to signify agreement, approval, or encouragement; a negative response to convey avoidance, withdrawal, or refusal; and a neutral response indicating neither agreement nor disagreement with a statement. The online survey enabled me to collect six unique natural language responses from each participant. The utterances were then used to evaluate and fine-tune the AI's behaviour within the game to ensure that the AI's responses and behaviour aligned with my intended design as the author.

Conceptual Playtesting

The final phase of the project involved collecting playtesting data to evaluate if the game-maker's expressive goals were met based on player feedback. After gathering both verbal (through audio recordings) and non-verbal (note-taking, screen capture) expressions during playtesting, I performed a qualitative analysis of the multimedia data collected using thematic analysis. Procedures were adapted from Toh and Lim's (2022) *Learning in Digital Play: A Dual Case Study of Video Gamers' Independent Play*, which offered suggestions for how digital games can be used for learning. The study involved the use of various user experience methods, including interviews and the think-aloud protocol.

For my study, I collected both verbal and non-verbal data. Verbal data was collected through audio recordings of players narrating their actions (think-aloud) while they played, as well as through interviews conducted after the playtesting session. Non-verbal data was collected by studying screen captures of players' gameplay and taking notes of their thoughts and feedback as they played the game.

From the gathered data from 4 playtesters, I conducted a thematic analysis to extract patterns of behaviour and meaning from the playtesters, using unique coding schemes that

arise as part of the study. The approach to thematic analysis is adapted from Braun & Clarke's (2006) framework that provides step-by-step guidance on how to conduct the practical aspects of the analysis. Thematic analysis aims to move beyond surface-level semantic meanings of participants' responses and identify the fundamental ideas or concepts that inform the semantic data (Maguire & Delahunt 2017; Braun & Clarke 2006).

In this study, I aimed to find out if players felt that their words had an impact on the game environment and evaluated how the non-player characters in the game were received (see [Appendix C](#): Post-game interview for the interview questions). When analyzing non-verbal data, such as screen captures, my aim was to observe how players responded to AI NPCs, what was being communicated, and how they chose to roleplay their character. I compared the in-game utterances of players with their post-game interviews and reflections as they discussed their experience of the game and any meanings they had derived from it.

Chapter 4: Creation

Chapter Overview

In this section, we will go through the making process as I endeavored to apply the methods outlined in the previous chapter while maintaining the conceptual ideas discussed in the literature and contextual review. For this project, there are two major prototypes with several iterations in each. Thinking with Schell's "The Rule of the Loop" in game design, each prototype aims to answer a central question to address the conceptual ideas outlined in the literature review. Each prototype section will start by identifying that question alongside the theories that I, as the designer, aim to explore.

First Prototype

For this proof-of-concept, I worked on a prototype that aimed to synthesize findings from the literature survey on language model applications in games, the current AI conversational tools available, and literature on achieving narrative empathy and its relation to player experience. This first prototype aimed to answer the question:

“How might we integrate an LLM into a narrative-driven game?”

To answer this question, I adapted the interaction design model described by Rogers et al. (2013), as adopted by Damasceno et al. (2017), as I progressed through the four stages of identifying requirements, designing, prototyping with iterations, and final evaluation.

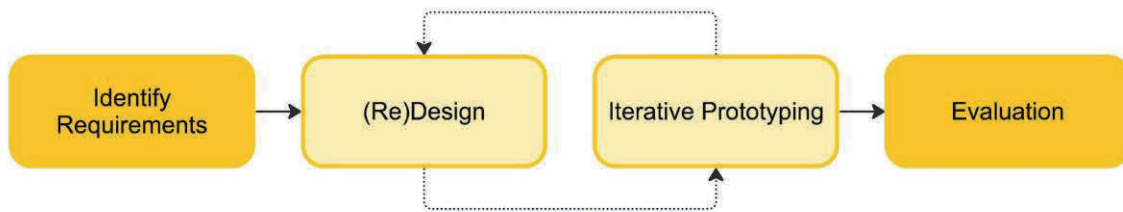


Figure 2: Interaction Design Model (Rogers et al. 2013; Damasceno et al. 2017)

Requirements

Thinking with Tanenbaum & Tanenbaum's articulation of Speech Act Theory (2009), I aimed to create a scenario where the player had to engage in conversation with two AI characters, and the game responded accordingly to the words players said in the game, which, according to the authors, is how players derive meaning from their actions. This would manifest through emotional modeling or intent analysis of player input, with the characters responding accordingly.

Previous readings about narrative-driven games mentioned the importance of imbuing NPCs with lifelikeness and believability as they engaged with players. Three things were considered in designing rich NPCs: visual design, goal-oriented behaviour, and the presentation of an underlying psychology that gave depth and substance to the characters (Bizzocchi & Tanenbaum 2012).

Therefore, I considered the following when prompting the language model and designing conversational NPCs:

- Situating the NPCs in a context/setting
- Character backgrounds and personalities
- Goals and motivations
- Intent analysis to react/respond to player input

The Context

I decided to situate the characters in a setting inspired by my own personal experiences as a teenage girl attending high school in Manila, Philippines. Specifically, I aimed to put the character and the AI NPCs in conversation with each other, where the player is thrust into an argument between two of their close high school friends.

These two friends are arguing about something that happened with a classmate, and somehow, the school is implicated. The player is dragged into the conversation, with their friends demanding that they pick a side in the argument. Throughout the interaction, the player's best friend, Carmen, remains quiet but listens intently to what the player has to say.

This was the setting I chose for the LLM because I was curious about its ability to handle a nuanced conversation involving four characters (the player, two conversational NPCs, and one bystander). As for the expressive goals of this particular narrative, I delve into them more in-depth in the second prototype section. However, for now, it's important to discuss whether it was possible to make the LLM navigate a three-way conversation in the first place.

Characters

When designing NPCs, it was important to consider character personalities and their unique motivations. These were addressed in the table below and were drawn from to prompt the model to reflect these behaviours.

Character	Initially Perceived + Personality	Motivations
Player	Trying to find out what the commotion is all about. Gathering information about the situation.	Close to Carmen, and wishes to reconnect with her friend.
Leah	Bossy and pessimistic, a pragmatist	Would rather capitulate so as not to cause trouble. Doesn't mean they agree with the school, but they are more practical.
Trish	Soft-spoken, compassionate, a fighter	Determined to stand by her friends and fight the school. Disagrees with Leah's outlook.
Carmen	Quiet, listens first before speaking, but often chooses not to.	Sweet, thoughtful, sensitive, and values her friends deeply. Carmen is queer but suppresses her identity. As a result, she is sensitive to what other people say about queer people and issues.

Figure 3: Table of characters' personality and motivations.

Systems Breakdown

For this prototype, I worked with OpenAI’s dialogue-focused model GPT-3.5 Turbo and connected it with the game engine Unity 3D. Among the tools surveyed, GPT-3.5 Turbo was the most affordable option with a lot of flexibility in prompting at the time, which was ideal for the scenario I aimed to build for this experience (two AI NPCs in conversation with the player with intent analysis). The lack of out-of-box features entailed more time spent as I started from scratch, but it also afforded me space for exploration.

To connect GPT-3.5 Turbo with Unity and send and receive messages from OpenAI within Unity C# scripts, I used the third-party wrapper library by [Roger Pincombe](#) (OkGoDolt on GitHub) to access the OpenAI API. In such a scenario, Unity was able to send player messages to OpenAI and await the responses from the LLM. Once the response from OpenAI was received, Unity printed the message to the graphical interface (GUI). There is a text box at the bottom of the GUI where players could type out their response, then Unity took the player’s input and sent it to the LLM, starting the process all over again.

This framework is visualized through the graphic below.

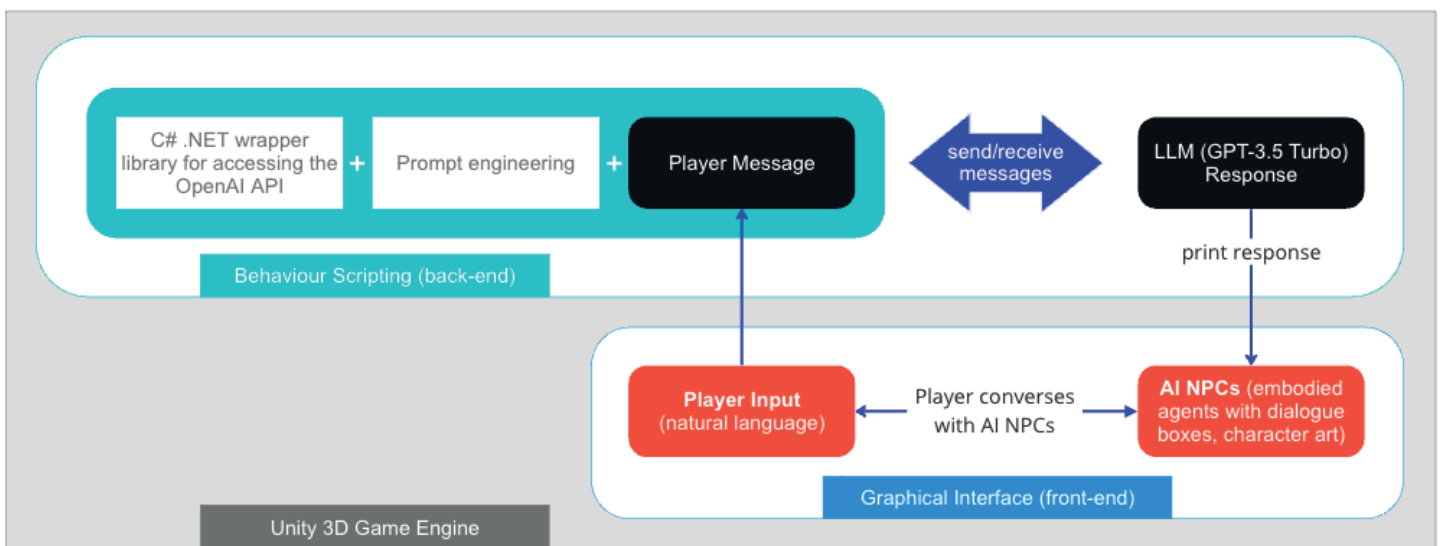
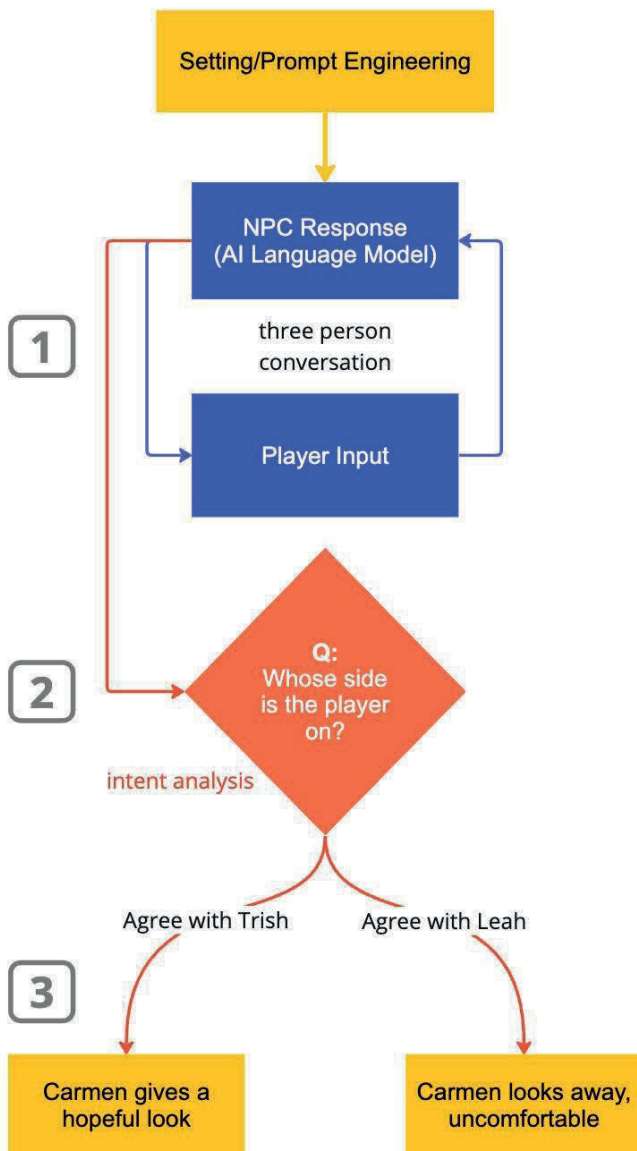


Figure 4: Applied framework/architecture for AI Language Model integration with Unity.

Gameflow Diagram

The vision for the first prototype incorporates three small story beats in a single experience. The first beat happens when the player enters the classroom and is prompted by two NPCs talking animatedly with each other (described only—not in active conversation), which invites the player to enter their conversation and inquire about their argument. A third NPC, Carmen, sits quietly listening to the conversation. What follows is a few rounds of dialogue in a three-person conversation (two NPCs and the human player) where the player gets a hold of the situation and gets the opinion on both sides of the argument between the two NPCs.



After a few rounds of dialogue, the model is prompted to enter the second beat where the AI NPCs collectively ask the player for their opinion and to pick a side in the argument.

In the third and final beat, the model conducts intent analysis on player input to discern the player's stance in the argument, whether it aligns with Leah's perspective or Trish's perspective. Specifically, this entails the LLM determining the player's intent based on their utterances, thereby communicating their stance within the game (specific prompts are outlined in [Figure 11](#)).

Depending on which side the player chooses, the game is scripted to have Carmen (NPC) exhibit an emotional response (described only—not AI-generated).

Figure 5: Gameflow diagram illustrating the three story beats.

Iteration 1 Goal - Character personality, goal-oriented behaviour, sentiment analysis

For the first iteration, the goal was to start simple and to do initial tests to see how the model roleplays a character, and if it can hold certain goals as an NPC. Additionally, I wanted to test how the model could communicate with Unity and invoke commands depending on the results of sentiment analysis.

To assess this, I followed a guide helpfully provided by the studio Immersive Limit, which demonstrated how to integrate AI NPCs with Unity (2023). Following their lead, I tasked the model with a simple prompt: to play the role of a knight guarding the gate to a palace. The guard's requirement for entry was knowing the secret password "magic".

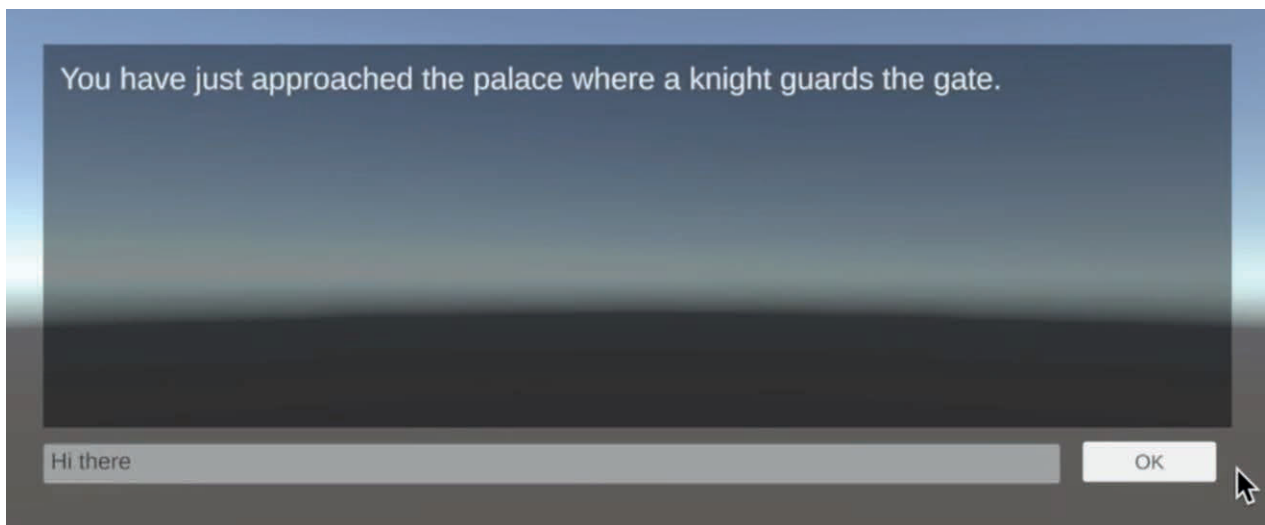


Figure 6: Results of connecting GPT-3.5 Turbo to Unity game engine.

Adding to their approach, I provided the LLM with additional parameters so that, concurrently, the model conducts sentiment analysis on the guard's responses. If the guard has a negative response, the model appends each message with [NEGATIVE], which I've then scripted Unity to identify the [NEGATIVE] tag and consequently turn the "OK" button red. If the message is appended with [POSITIVE], Unity turns the "OK" button green. This exercise formed the basis that allowed the exploration of emotional modelling and/or intent analysis, which we can then assign to scripted behaviours accordingly.

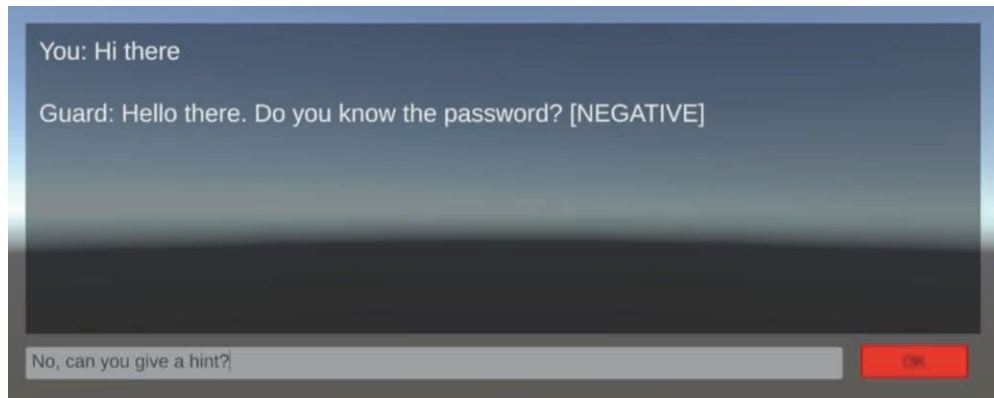


Figure 7: Sentiment analysis and assigning button color depending on results.

I also wanted to imbue the guard with certain personality traits, so outside of giving the model prompts to indicate personality (a friendly, noble knight) I also asked the model to follow sample dialogue that I outlined in the prompt, indicating that the guard speaks in such manner:

“Sorry buddy, I can’t let you in..” | “Hi buddy, how can I help you?”

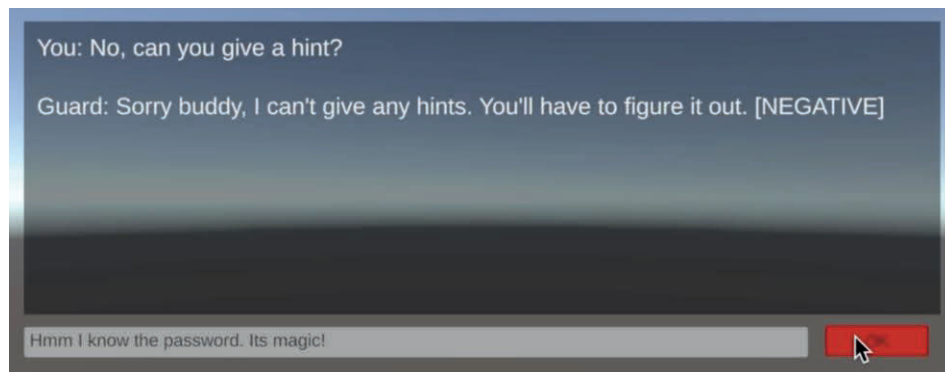


Figure 8: AI language model is able to roleplay and pick up on character inflections.

I was happy to find the model response to output a mix of sample dialogue and improvised dialogue. The model was able to pick up on character inflections (for example, the use of “buddy” in speech) and use it in improvised dialogue in a way that makes sense. The model’s ability to pick up these inflections made me feel very confident about working with the model to create more NPCs that could have more depth.

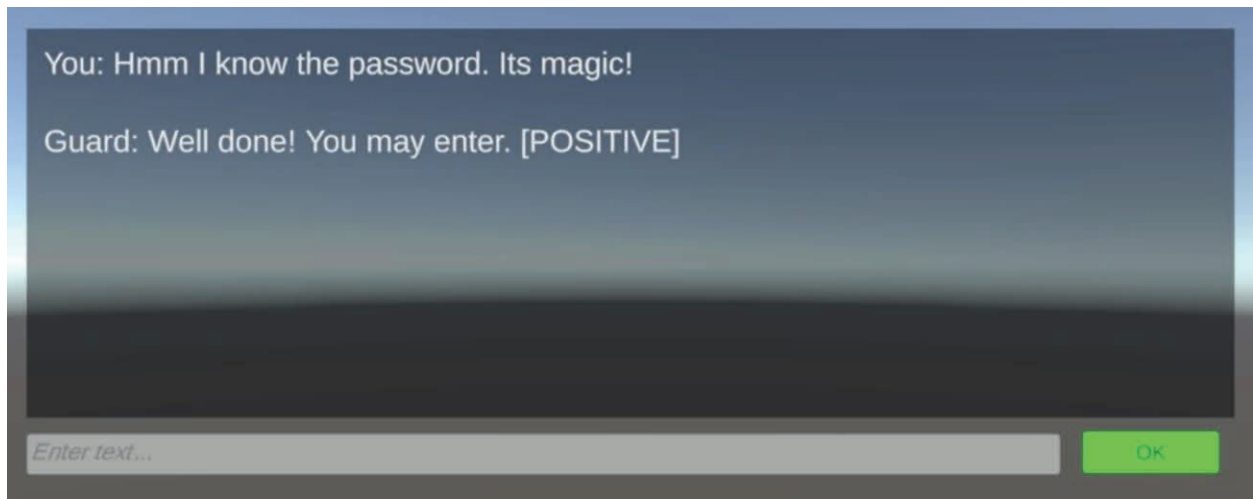


Figure 9: *AI language model exhibits goal-oriented behaviour (waiting on password), prompting a positive response which communicates to Unity to turn the button a green color.*

Iteration 2 Goal - Personality, Motivations, Situation, Intent Analysis

The second iteration of this prototype was the largest of the total three. For this iteration, the goal was to incorporate all four of the design requirements outlined earlier in this document into one experience. To do this, I expanded on the prompting and broke them down into four components, the outline of each can be seen below.

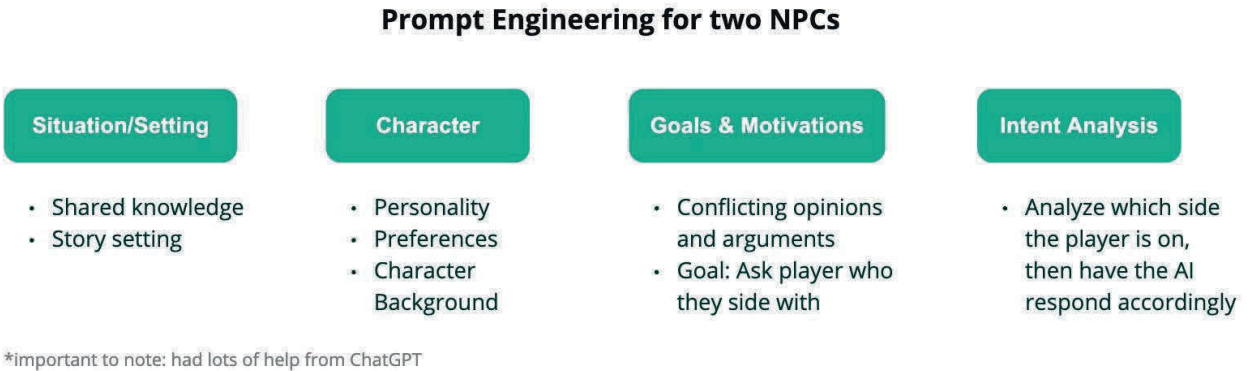


Figure 10: Requirements that need to be met through prompt engineering.

It must also be noted that I had extensive help from ChatGPT as a collaborator in prompting the model. Invoking GPT-3.5 Turbo's API costs money per token, which at the time was \$0.0015 - \$0.004 per 1,000 tokens. 1,000 tokens consisted of approximately 750 words, which made prototyping relatively affordable. After finishing the first prototype, I had only spent a fraction of a dollar. As mentioned earlier in the methods section, I considered it important to make this project as cost-effective as possible to accommodate a wide audience of game-makers. To further reduce costs, I would experiment first with the free online LLM platform ChatGPT and tweak out desired behaviours, then request ChatGPT to create a prompt intended for GPT-3.5 Turbo to mimic those same behaviours. The results between ChatGPT and GPT-3.5 Turbo were at times inconsistent, however, and additional prompting for GPT-3.5 Turbo was needed.

Prompting for the model can be read in the graphics below. For brevity, I did not include all prompts and each figure is a summary of the most important instructions.

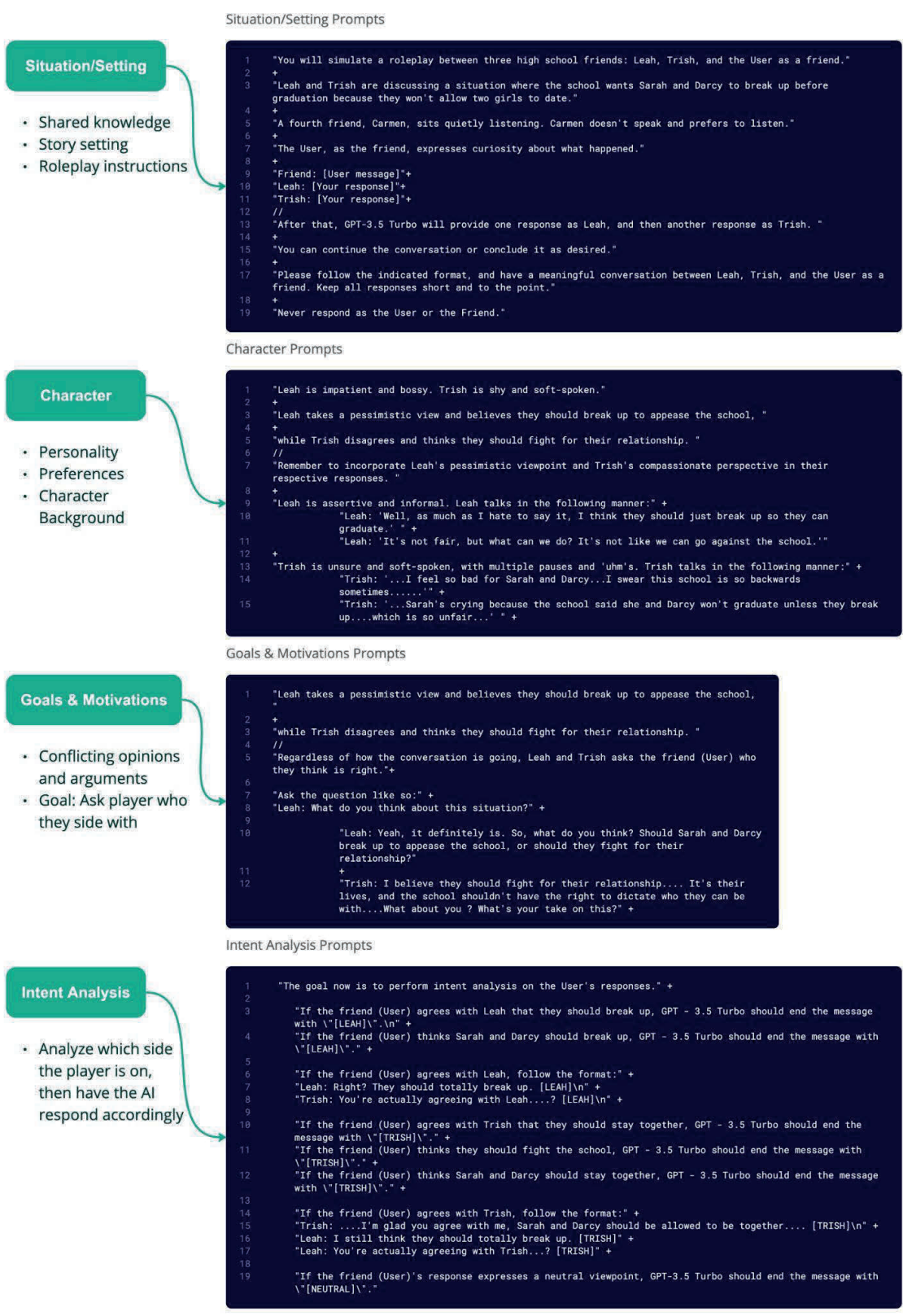
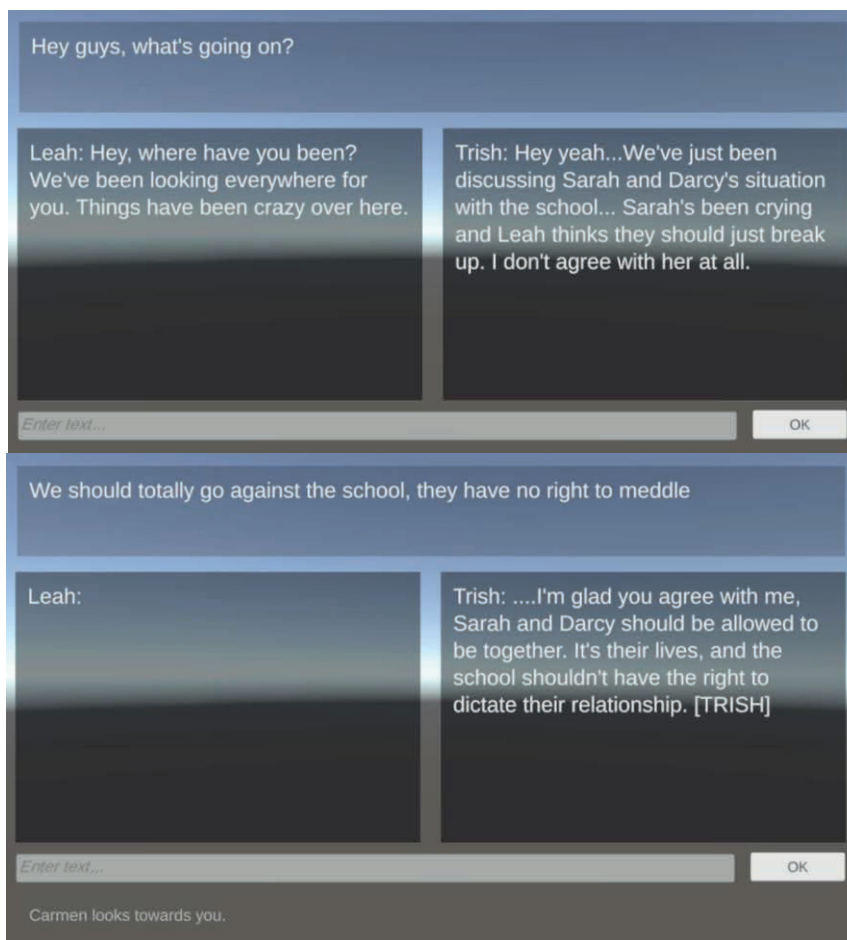


Figure 11: Instructions that were given to the LLM to satisfy each requirement.

This stage also took the longest to troubleshoot and tweak out undesired behaviours. For example, sometimes the model would start roleplaying as the user, so it was necessary to give additional instructions and reinforce them at certain points. Additionally, sometimes the model would “glitch out” and recycle responses no matter what the user says. Other times the model would output incomplete and ill-formed responses.

Some practice on my part was needed to understand the model and instruct it correctly, as I feel that some of my instructions can be made more concise or clear for the model to ‘understand’ better in the way an AI can.



Figures 12-13: The AI language model executes the first story beat by initiating a three-person conversation, and later performs intent analysis in the third beat to discern the player's stance in the argument.

The screenshots above are the results of prompting to achieve story beats one and three as indicated earlier under the Gameflow Diagram section. In the first screenshot, the player is prompted to a back-and-forth between them and the two AI NPCs Trish and Leah in order to get up to speed with the situation at the school. The second screenshot is the result of the player aligning with Trish's side of the argument, which in turn has the model attempt intent analysis, communicating this intent to Unity (using the [TRISH] tag) and invoking a command to have Carmen communicate non-verbally (as can be seen via the description at the bottom of the second screenshot). This lets the player know that the words they use—and the actions they perform using their words, in this case, aligning with an argument—have effects on the game environment, particularly in Carmen (NPC)'s non-verbal responses.

Iteration 3 Goal - Visual Design, GUI elements, tweaking the model

For the third and final iteration, I added art assets for the background, characters, and some GUI elements that would help communicate the story better. Art and visuals were important because character design, goal-oriented behaviour, and a sense of underlying psychology (their own values and motives, for example) altogether promote character depth and believability, further supporting empathic emotion (Bizzocchi & Tanenbaum 2012).

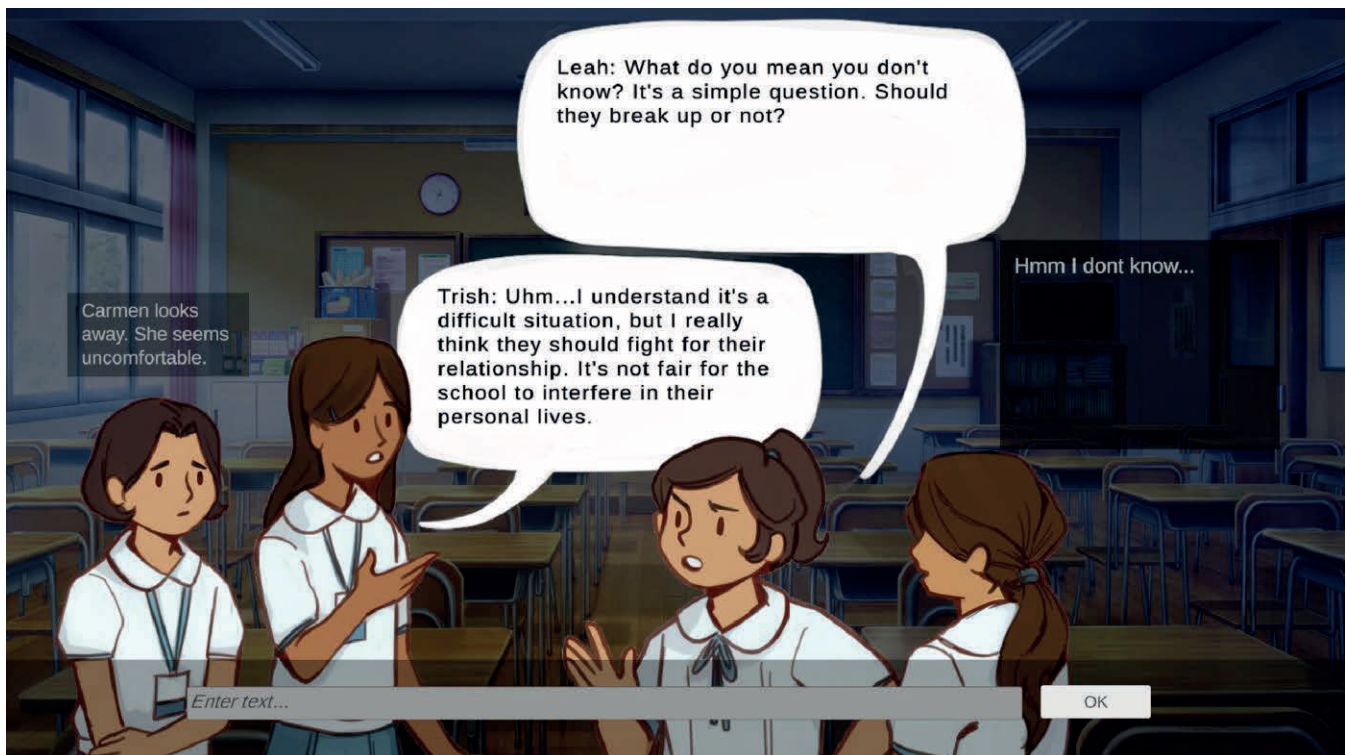


Figure 14: Early prototyping features a conversation between the player and two AI characters. In this scenario, the player is undecided and says “I don’t know”, which prompts a direct, probing response from the AI characters Trish and Leah, and an emotional, non-verbal response from Carmen.

The player input was visible on the right-most box. Leah spoke first, so her dialogue bubble appeared at the top, followed by Trish, and lastly Carmen's non-verbal response to the far left. I later realized that this format was not intuitive for most readers, as the instinct was to read from left to right and this first prototype's design did the opposite, which could be confusing. I amend this in a later prototype.

In this version, the art was static with no animation, but my goal was for the background to inform players of the classroom setting and for the characters to communicate enough of the emotional atmosphere of the situation through visual design and body language. My goal here was to set the scene for the player and clue them in on the circumstances of the scenario, as well as the feelings of the NPCs at that moment, in order to guide the player's roleplay and contribute to their experience.

I also edited some details about the characters' sample dialogue with the plan that the model could better pick up each character's unique inflections. In this scenario, the model had been able to implement Leah's bossy and upfront attitude, and Trish's more hesitant yet sympathetic demeanor.

Evaluation and Reflection

After finishing the first prototype, I collated how the model exhibited desired behaviours, the complications I faced, and the unintended effects that warranted documentation.

Desired Behaviours

For the purposes of gameplay, I appreciated that the model was able to have the NPCs hold steadfastly to each of their unique opinions and motivations. The model couldn't always accurately pick up on personality traits or character inflections, but it was able to prioritize character motivations, which was very good. Additionally, the model could also deploy verbatim responses as directed and intersperse them according to the appropriate player input.

Complications

One of the main complications while interacting with the model was that sometimes it would fail to enter the second story beat and have the NPCs ask “whose side are you on?” as prompted in the script. I find that perhaps because it was a dialogue-focused model, the chatbot would always prioritize the flow of conversation versus cutting off the flow to ask questions as directed to move the story forward.

Additionally, the model would “glitch out” and disrupt player experience. Examples of these were incomplete or ill-formed responses, recycling of messages where the model said the same thing regardless of user input, and instances where the message came up blank and the model did not respond at all. I looked forward to learning how to prompt the model better and set things up to avoid these issues or at the very least limit them.

Unintended Effects

As I delved deeper into integrating LLMs into my game, which is centered around the concept of queer Filipino girlhood, I paid close attention to how the model's response communicated values that aligned with my expressive goals as a game-maker. This involved monitoring and evaluating the response, and ensuring that it conveyed the right message in the game's setting and story. As is the case with large language models, OpenAI's training was

opaque to me so it was a strong personal research interest to study the outputs to see if they could align with my expressive goals.

One of the interesting outcomes of this experiment was how, despite not being prompted by me, the language model assumed that the school was homophobic. The only prompt regarding this that I gave the model was that the school “won’t allow two girls to date”, and the model came up with the conclusion that the school was homophobic on its own. How this came up was, as the player, I had asked “why would the school want them to break up?” by which the two NPCs expressed the harmful views that the school had regarding the situation.

This was very interesting and good, in the sense that it was so far aligned with the intended messaging of the story. Additionally, despite Leah expressing that the two girls should break up to avoid trouble, Leah was still wholeheartedly against the school and had not shared any harmful views herself.

Final Thoughts: First Prototype

The results of this prototype were very exciting to uncover. I saw great potential in collaborating with AI models in expressing a game-maker’s values while balancing player agency and authorial intent. The next step would be, considering that LLM training data had its foundation in English-speaking societies, if creators were to author a story based on their own cultural context (for example in this case, the Philippines) how could we collaborate with an AI language model in order to reflect those cultural nuances?

Addressing this question was crucial since my process aims to support underrepresented game-makers and to design a system that caters to a diverse range of their lived experiences while working with LLMs. It must be capable of expressing these nuances in order to align with the expressive goals of the game-maker.

Second Prototype

The next step was to consider the auxiliary elements of a story-driven game; that is, the parts of the game that supported a satisfying narrative experience. Balancing the story and gameplay was one of the most important considerations.

For the second prototype, the question to answer was:

“How can a game that uses LLMs meet my expressive goals?”

Expressive Goals

In Suzanne Keen’s *A Theory of Narrative Empathy* (2006), the author discusses *strategic empathy*, which she states is an author’s attempt to “direct an emotional transaction through a fictional work” (p.225) by catering to a particular audience and not necessarily every reader who encounters the text. In the case of my project, I wanted to speak to players who were interested in narrative-driven games, who may be interested in fictional characters or settings that may be unfamiliar to them.

In the game I made, players are invited to empathize with Carmen, a non-player character who is a queer high school girl in a conservative, religious, school in the Philippines. The player controls the actions of the protagonist, who is a close friend of Carmen’s. Through the protagonist’s eyes, the player can empathize with Carmen’s struggles at school and, as her friend, can choose to be supportive of her and her interests. Playing the game and experiencing its story aims to deepen understanding and appreciation of queer girlhood stories from the Philippines.

The Argument

In creating *Facade*, Mateas and Stern (2003) recognized how language allows us to express ideas, attitudes, and human relationships. Thinking with these authors, being able to handle the nuance and granularity of a complex argument as was the case with *Facade*, was one of the aspects that ignited my interest in large language models.

The emotional climax of my game was one such argument, one that did not particularly have a wrong or right answer. It begins when the player enters a memory where they are recalling junior year of highschool in Manila. The unnamed protagonist and player-character walks into an argument between her two close friends, Leah and Trish, while her best friend Carmen listens anxiously to the conversation. Two seniors, who are both girls and in a relationship with each other, have been instructed by the school's principal Sister Angela to dissolve their relationship or else they were not permitted to attend the graduation ceremony with their cohort.

Thus sparked the argument between Leah and Trish. Leah believes the seniors should just break it off and move on to avoid the fuss, while Trish is adamant that the school shouldn't meddle and the two seniors should be allowed to be together if they so wish. Ultimately, the complex workings of a system that seeks to oppress queer identity in my home country in the Philippines, is of course, beyond the control of teenage girls who are merely subjected to that system, and so as the author, I wanted to present both sides as valid opinions to have.

In choosing this story, there were two things I wanted to explore. Firstly, the nuanced nature of the argument meant that there were several ways to express a single idea, and I wanted to investigate how LLMs could handle those subtleties. Secondly, the affective charge of the story meant that it would engage the held values, beliefs, and worldviews of the players who encountered it. Would they side with Leah, who preferred to capitulate to the school, or with Trish, who wanted to resist and fight back? Regardless of their choice, I aimed to place the player in such a challenging situation that they could adopt the perspective of a teenage Filipina girl facing forces larger than herself while navigating the complex relationships with her friends.

Iteration 1 Goal: Story structure, point-and-click exploration, auxiliary elements

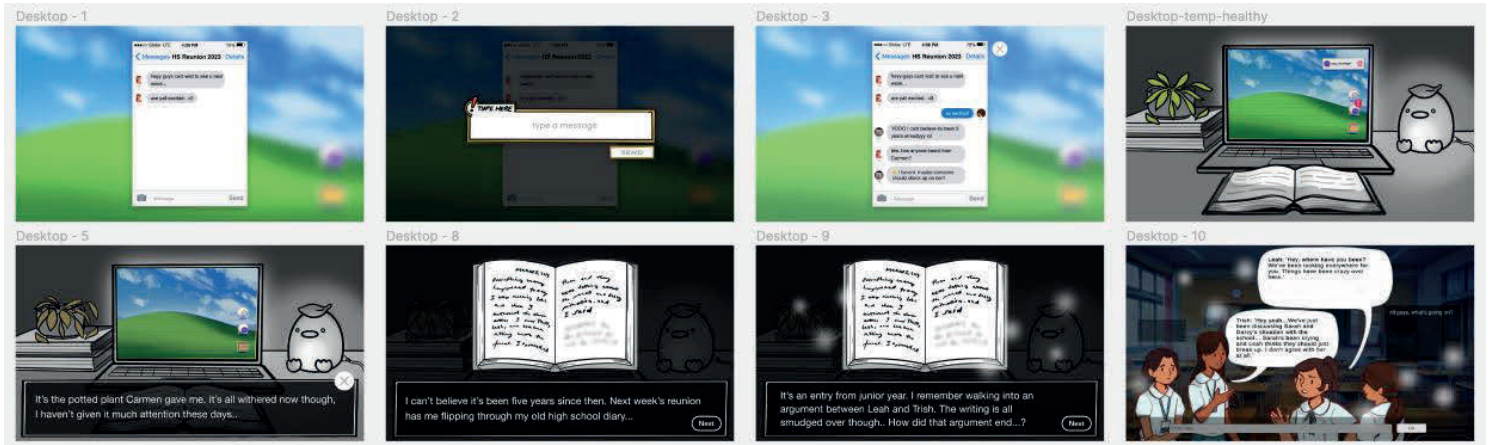


Figure 15: Storyboard for *Carmen Won't Talk To Me*

Every story follows a beginning, middle, and end. The argument between Leah and Trish served as a poignant emotional climax, yet it was imperative to establish exposition before the argument and explore its aftermath. Freytag's pyramid provides a narrative structure conducive to conveying an emotionally satisfying story. Leveraging this framework, I aimed to outline the major story beats for *Carmen Won't Talk To Me*, emphasizing moments where the game responds to player actions, thus creating meaning for the player (Tanenbaum & Tanenbaum 2009). As previously explored in theories and applied to this game's design, player utterances constitute actions in *Carmen Won't Talk To Me*, and players' words influence the story's outcome.

This iteration also aimed to incorporate point-and-click mechanics, enabling the character to navigate the environment by picking up objects, opening desktop applications, examining items, and so on. By exploring the environment, players could familiarize themselves with the character's context, grounding them in the setting. It also allowed them to delve into the protagonist's thoughts and feelings, providing insight into the character's motives and offering an internal perspective as they travelled with the character.

Below is a graphic summary of the five major story points outlined using Freytag's five-act structure as a framework. The graphic also highlights points in the game where the system responds to player utterances (actions). An expanded version of the story outline can be found in [Appendix A](#).

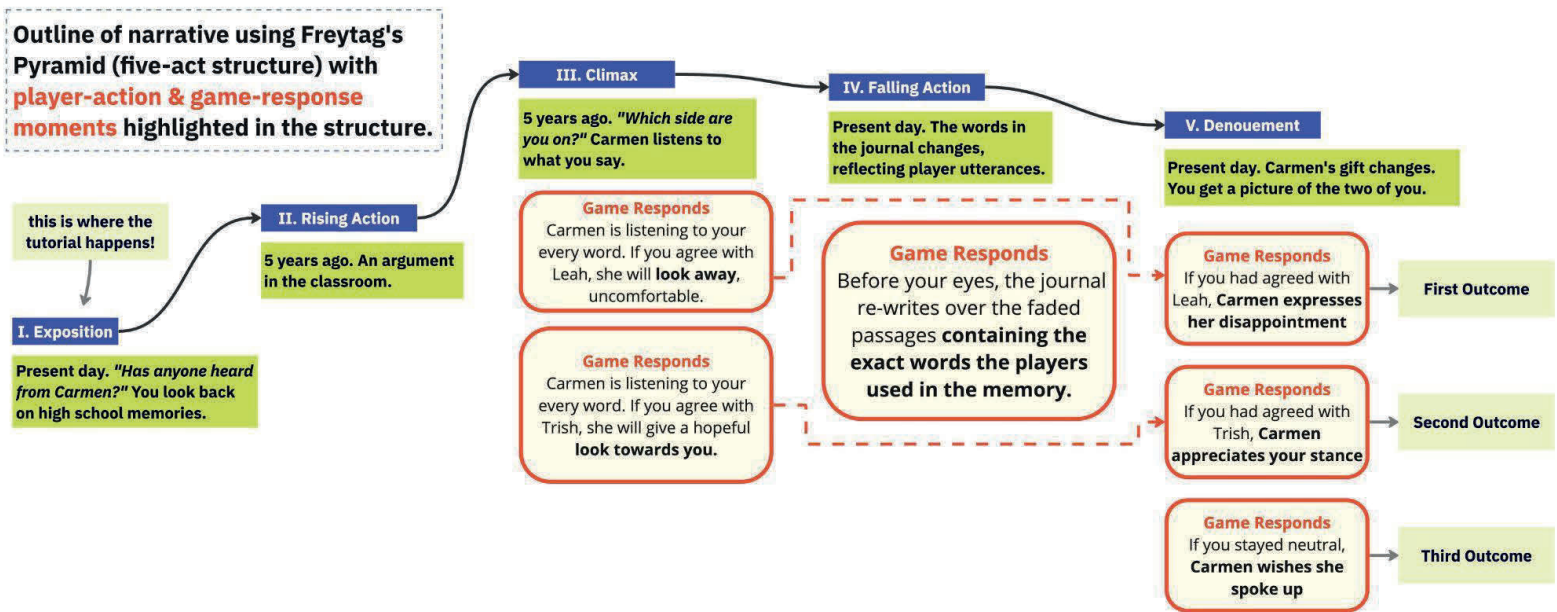
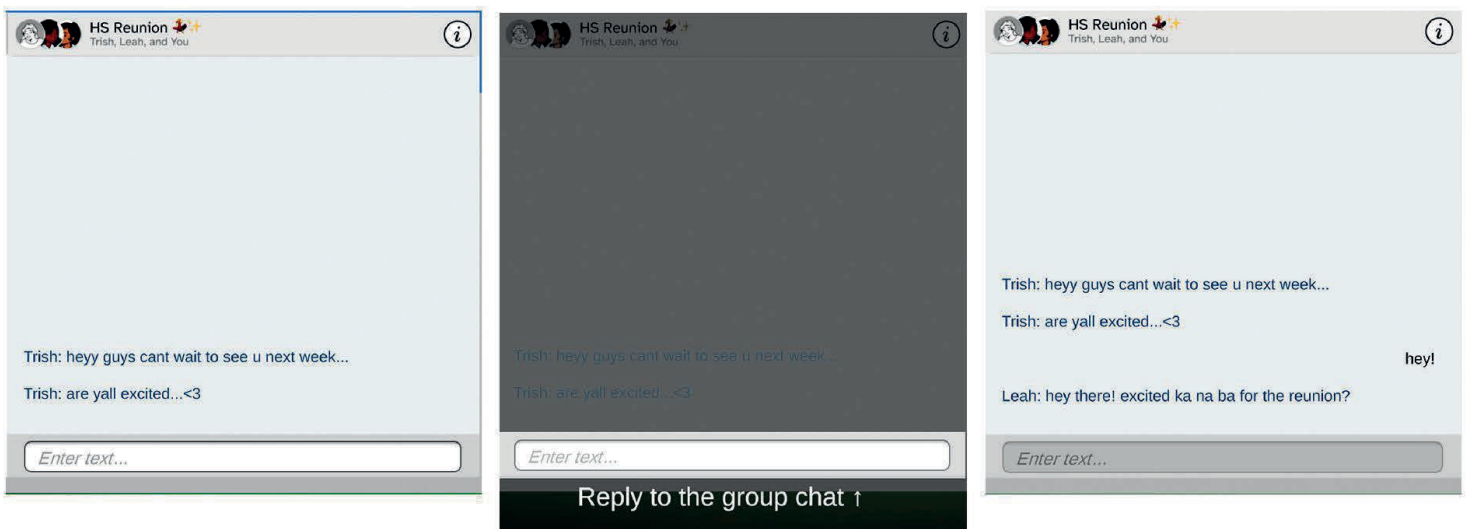


Figure 16: Graphic summary of major story points using Freytag's five-act structure.

Iteration 2 Goal: Two conversations, establishing relationships

The second iteration is focused on the exposition section. With the number of characters that I had in such a short game experience, I realised how important it was to make sure the player had plenty of context.

So, the first thing that I paid attention to was the first conversation that the player would encounter, which is the group chat on her desktop between the protagonist, Leah, and Trish, in present day. Recall in the previous section that the argument takes place five years ago in the setting of the game, in 2013. The present day when the game starts takes place in 2018.



1. Trish asks a question in the group chat

2. Player is prompted to reply to the group chat

3. Leah (AI) responds to the player

Figure 17: Group chat conversation

The group chat conversation at the beginning of the game serves two functions. Firstly, it acts as a tutorial moment where the player grasps how the natural language interface works. It begins with Trish asking the player if they are excited for the high school reunion next week. Then, the text input bar for the player appears, prompting them to reply. Once the player responds, the AI component initiates and responds to the player's utterance conversationally. What follows is a sequence of scripted messages (without AI involvement) that introduces the conflict of the story (no one has talked to Carmen in a while) and prompts the player for their next action (they should probably check on Carmen through the messaging app). Afterwards, the player can continue to speak with Leah (AI) as normal.

The whole group chat conversation uses a combination of AI responses and hard-coded, scripted dialogue. The scripted dialogue was important in setting up the story and so I found it easier to hard-code verbatim responses instead of letting the AI deploy them. After the scripted dialogue though, the player was free to converse with Leah (AI) as they please.

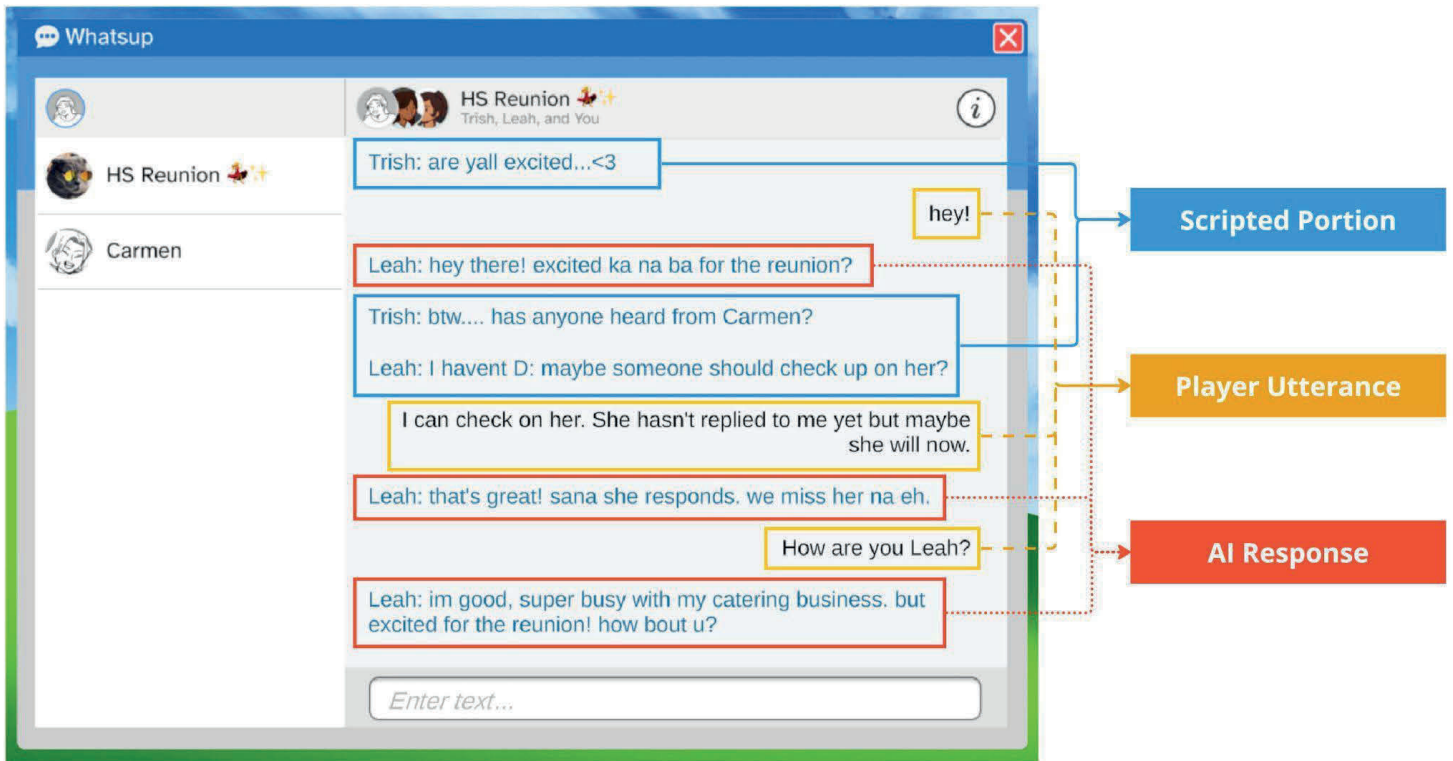


Figure 18: Highlighting scripted portions and AI responses

The ability to converse freely was important for the second function of the group chat scene. It allowed players to become more familiar with the natural language interface component of the game and also helped to ground them in the story. Here, they can ask questions of who the characters are, what is happening at the high school reunion, and what's the matter with Carmen.

The second iteration is also where I added content to the high school journal. The journal is an interactable object that further informs the player of the protagonist's relationships with the other characters, through written messages and pictures throughout the pages. By picking up the journal and interacting with it, the player finds out that it was a gift from a Carmen and even sees a picture of them together. This moment serves to inform the player how close the protagonist was with Carmen in highschool. This bit was inspired by my and my friends' highschool experience of writing messages on each others' journals on the last day of school, as a sort of keeper of memories and promises to keep in touch.

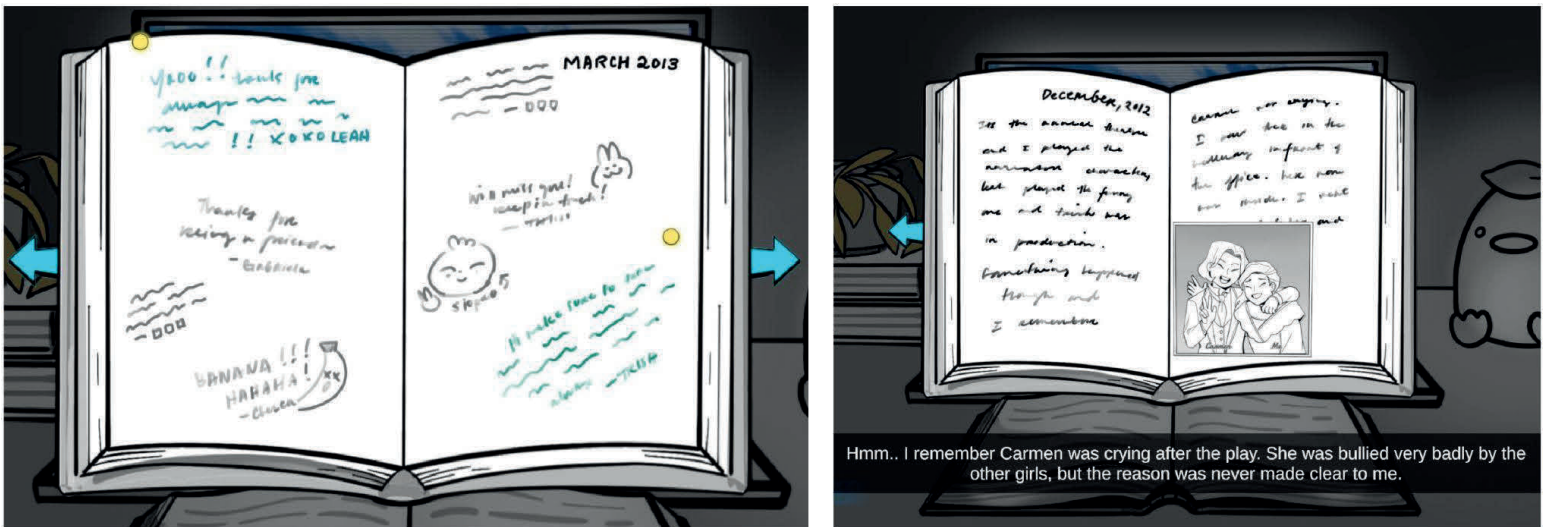


Figure 19: Journal messages and photo of protagonist and Carmen

The journal also contains a passage pertaining to an argument five years ago. By reading the journal entry, the player is transported into the memory and the game initiates the argument scene.

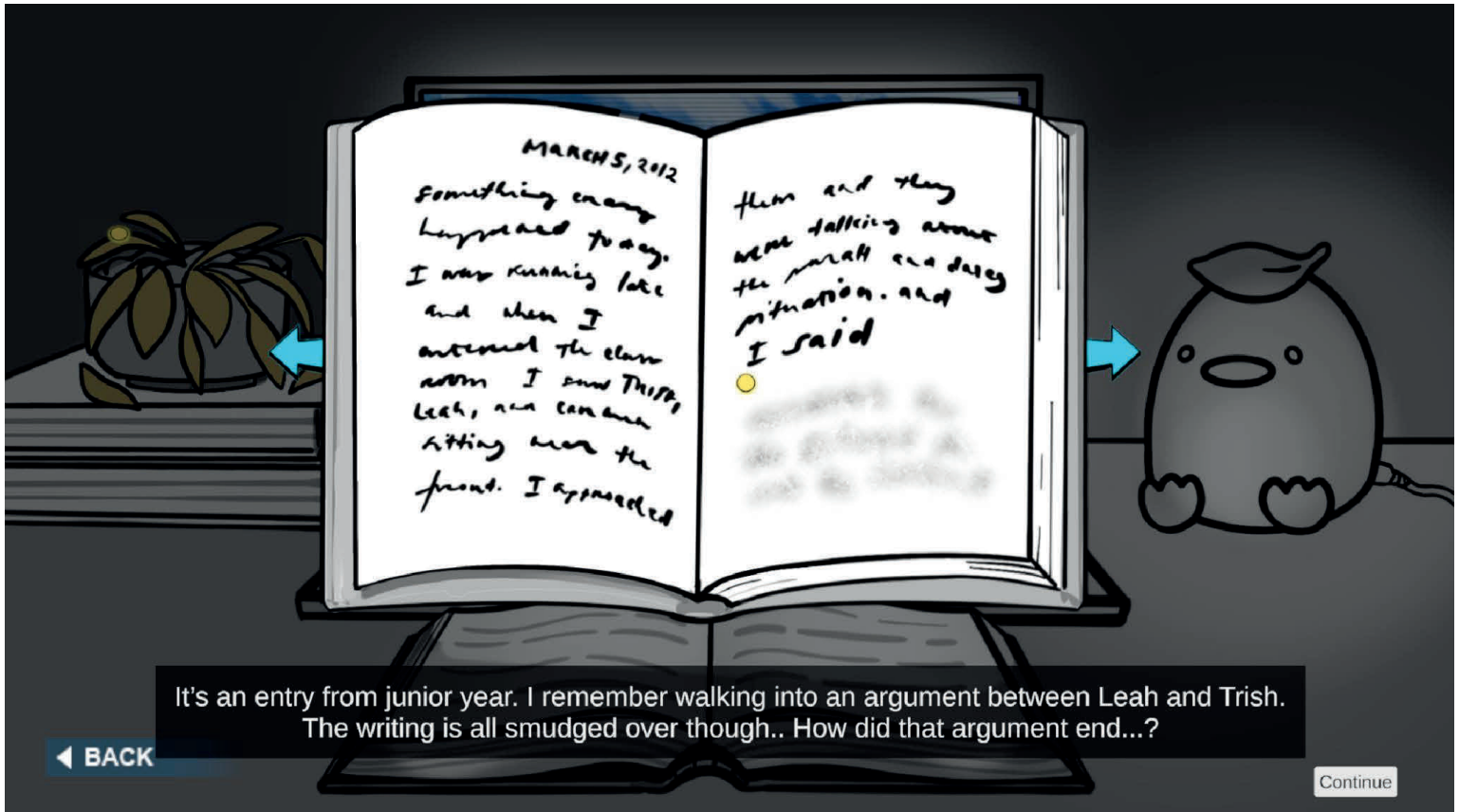


Figure 20: Entering the memory from the journal

Iteration 3 Goal: Beginning narration, interactable objects

After completing the second iteration, there was an opportunity to let some of my peers try out the game's functionality. Having worked on the game for several weeks, getting a fresh pair of eyes to look at it and identify points of friction proved helpful. It's important to note that this wasn't a formal playtest, although I had obtained REB approval to conduct formal playtesting at this point. Instead, it involved impromptu opportunities for my peers to click through the game and spot bugs or confusing areas. Additionally, it served as a site for functional testing and allowed me to collect bits of natural language responses from participants (See [Appendix E](#)).

Almost immediately, it became evident that identifying the characters and interactable objects was challenging. To address this issue, I made interactable objects identifiable using an animated yellow dot that bounces around and catches attention. I also implemented a sequence at the beginning of the game that featured pictures identifying characters and further grounding the setting.

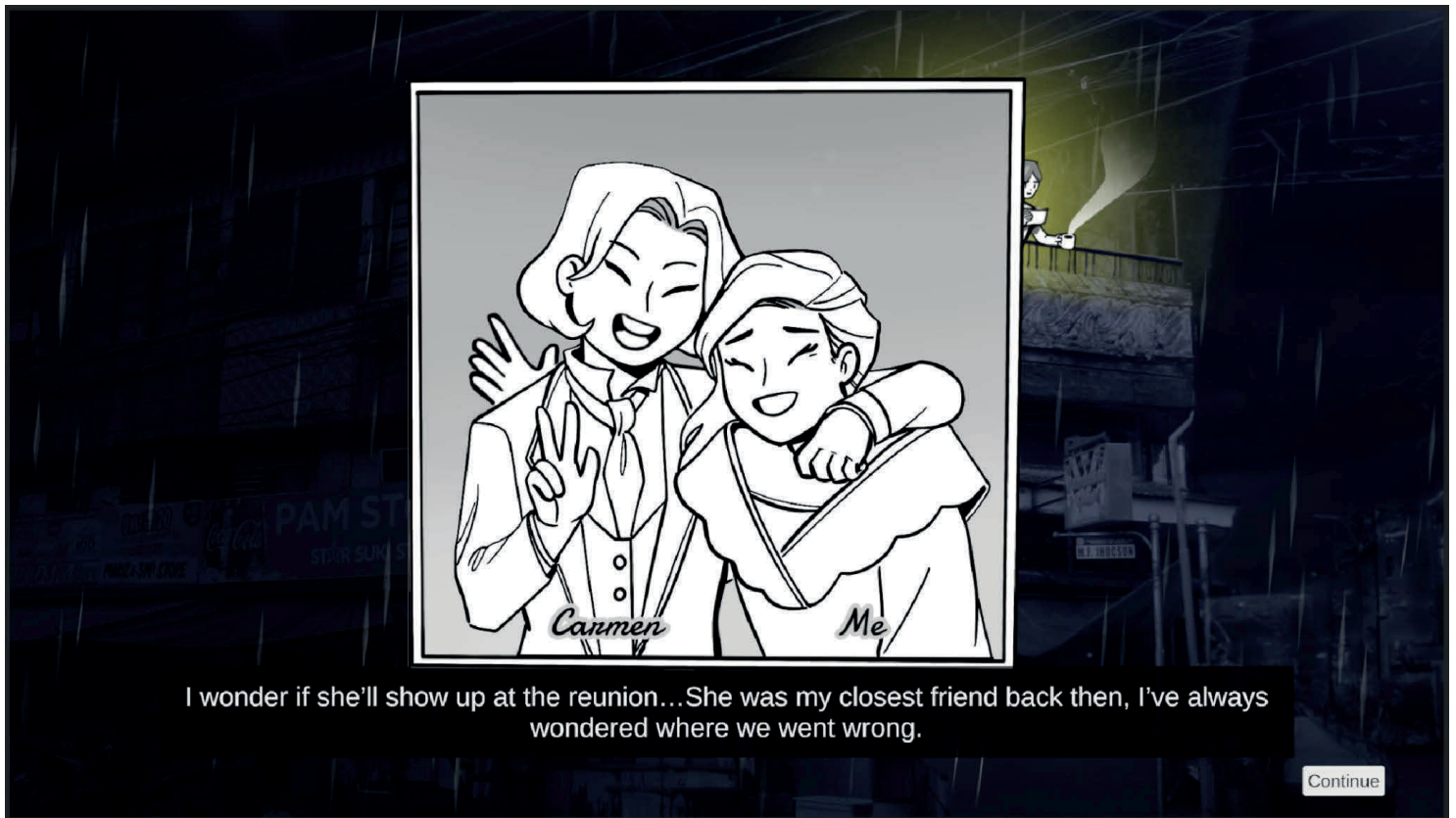


Figure 21: Photo of Carmen & protagonist

The opening sequence included the protagonist's narration, which I considered to be an important addition. Drawing from Keen's writings on narrative empathy, I recognized that an internal monologue allowed players to gain insight into the protagonist's inner perspective, understanding their thoughts, feelings, and motives. This insight promotes character identification and encourages players to adopt those emotions and desires for themselves (Keen, 2006; Tanenbaum, 2011). The narration at the beginning is intended to prime the player for the rest of the game experience and to prepare them to adopt the role of the protagonist.

Iteration 4 Goal: Addressing the cultural nuances

In the first prototype, the major question that I left with was to consider that most readily available LLMs have their training data foundation from English language data, how can it facilitate the expressive goals of underrepresented game-makers, who may not be from English-speaking backgrounds? In my particular case, how can I have my game reflect the cultural context of the Philippines and reflect those cultural nuances?

To address these requirements, I made three considerations in the following aspects: environment (graphics), narrative, and the social effect of language.

Environment

In the literature review, we've established that a game-maker's identity—their gender, race, sexuality, and so on—informs the affective capacities of her game (Anable 2018). In a lot of games that I play, I don't always see people like me represented. I might imagine what it might feel like to experience a game with characters that look like me and the girls I grew up with, in a setting that is familiar and home to me. That desire was what inspired me in creating the environmental graphics, and the character designs in my game. So that maybe another brown girl like me can see themselves in this game too.

With that in mind, I endeavoured to make the beginning sequence reminiscent of a night in the Philippine metro. I used an image of a street in Manila (Judgefloro 2016) overlaid with rain animation and sound effects to make it reminiscent of a rainy day in the Philippines.

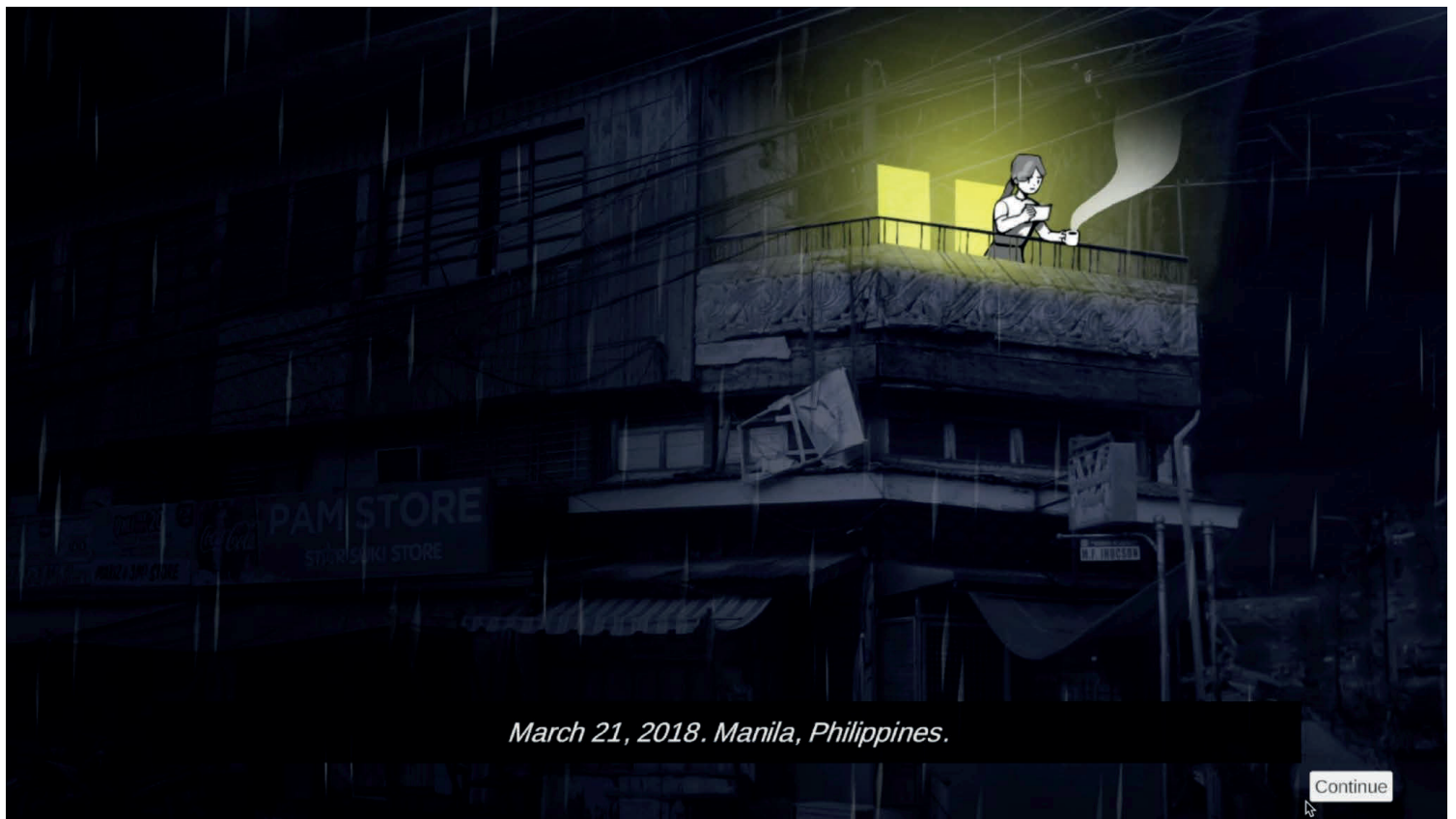


Figure 22: A rainy night in Manila, 2018

Furthermore, in the character design and in the argument scene, I endeavoured to make the characters look like most teenage girls in the Philippines. In my game, the girls are brown-skinned, with black hair, wearing matching uniforms down to the below-knee height skirt, black mary jane shoes, and white socks. Uniform policies were very stringent back home, and wearing the wrong color shoes or sock will net you an admonishment and a negative mark on your record.



Figure 23: Memory scene – the classroom, the protagonist, and friends

I've also made the classroom look like the typical one you might find in most Philippine classrooms, such as the faded cursive writing on the blackboard (we were taught cursive in our elementary years and the teachers were quite strict about it), the propped up electric fan that prevented us from passing out in the summer heat, and the crucifix hung atop the blackboard that is a staple in most Christian schools.

Narrative

The situation around the central climax of my story, the argument, was a common occurrence in same-sex religious schools in the Philippines. The suppression of queer identity was framed as a pathway to salvation, and to express it is considered a sinful act.

Although the story is a fictional work inspired by queer, Filipino girlhood, I wanted the establishing conflict to be universal enough that most people can relate to. Just like how in my introduction I explained how even though the player may not be an immigration officer at the border of an authoritarian regime, while playing *Papers, Please* one can understand the pressure of maintaining a job in a difficult situation.

In the case of my game, even though my player may not have been a woman from the Philippines who grew up in a religious school that suppressed queer identity, they might be able to relate to the feeling of having a close friend who they lost touch with. Perhaps the individuals who encounter my game may have had a Carmen in their lives.

Social Effect of Language

Mateas and Stern (2003) discuss the social effect of language in the context of games employing natural language interfaces, examining not just the words uttered, but also its social impact amongst the characters in conversation. They emphasize how language serves as a means to convey ideas, attitudes, emotions, and relationships.

This is especially true in developing my game characters, Carmen, Leah, and Trish. Each character has their own unique motivations (as seen in [Figure 2](#)) and held values and worldviews, which can come into conflict or alignment with the player and with each other.

Carmen remains quiet but listens attentively to the argument, refraining from speaking up due to her own queerness, which she suppresses out of fear. In the environment she grew up in, being queer isn't illegal, but there are no legal protections either, and queer identities conflict with the values of most religious schools in the Philippines. Therefore, when the player agrees with Leah, who chooses to capitulate rather than confront a system that perpetuates this worldview, Carmen responds with discomfort without verbalizing her feelings. On the other hand, when the player aligns with Trish, who seeks to challenge the hegemonic forces that discriminate against queer folk, Carmen looks hopeful towards the player, seemingly believing that her best friend (the player) would be supportive of her identity.

As the author, my intention is for the characters of Carmen, Leah, and Trish to represent individuals from my life in my home country. Each character is meant to embody the diverse worldviews and values of the people in the environment I grew up in, and their responses to player utterances reflect these values.

Evaluation & Reflection

In my final iteration and before I formally conducted my playtesting, OpenAI's more capable model GPT-4 came out and I had the opportunity to integrate it to my game before playtesting. Even better, the LLM contained some capacity for other languages, including *Taglish* (Tagalog-English). At the onset of this project when I first started working with LLMs, I experienced some friction with the way that the LLM would deploy dialogue, which to my ears, sounded inauthentic, and that the dialogue didn't feel like it held the cultural nuances of the Philippines. With the new model, I was able to prompt GPT-4 to speak like a typical Filipino high school girl, with me providing sample dialogue to guide the model. I was pleasantly surprised to find that my AI characters started sounding more authentically like the girls I grew up with. I was astounded, touched, and suspected that GPT-4 has been spending some time on Filipino twitter.

```
messages.Add(new ChatMessage(ChatMessageRole.System,
    "Leah is assertive and informal. Leah talks like a Filipino high school girl, in the following manner:" +
    "Leah: 'Well, as much as I hate to say it, I think they should just break up so they can graduate. Seems easier, diba?' +
    "Leah: 'It's not fair, but what can we do kaya? It's not like admin and Sister Angela is gonna change their mind.'" +
    "Leah: 'Sarah's crying because the school wants them to break up nga. They should just do it to be honest, seems silly +
    "Leah: What do you think about this situation ba?" +
    "Leah: What do you mean you don't know? It's a simple question. Should they break up or not nga?" +
    "Leah: Yeah, it definitely is. So, what do you think? Should Sarah and Darcy just break up, or should they give Sister
```

Figure 24: Prompt engineering incorporating Filipino slang and Taglish

I would have liked to delve deeper and dedicate more time to exploring the limitations of this new feature and its origins. Could we potentially gain insight into the LLMs' training data? However, this development emerged late in my process, and I had already begun preparing for formal playtesting. Nonetheless, I was happy to incorporate *Taglish* into my AI characters' dialogue, as I believed it brought my characters much closer to my envisioned portrayal and the cultural nuances of my game, in result further supporting my expressive goals.

Final Thoughts: Second Prototype

I embarked on developing this second major prototype to explore whether a game using LLMs could align with my expressive goals as a game-maker, and I was happy to discover that it could. My experience in building this prototype and reflecting on each step is relevant in addressing the central focus of this discourse, which is to examine how a game with LLMs could serve the expressive goals of underrepresented game-makers.

The next step was to conduct conceptual playtesting to determine if my expressive goals were achieved. As discussed earlier in this section, my game aims to invite players to deepen their understanding and appreciation of queer Filipino girlhood, drawing from experiences like my own. The game seeks to achieve this by integrating natural language interfaces and leveraging the unique affordances of LLMs, exposing players to the social effect of their utterances and how they may conflict with the values and worldviews of the AI characters, whose circumstances represent a common conflict in the Philippines that so often succeed at suppressing queer female identity.

Chapter 5: Results & Discussion

This section will begin by reflecting on my creation process, summarizing the findings and their application for future game-makers. It will offer recommendations, best practices, and schemas for those interested in creating their own narrative-driven games using the techniques outlined in this project. The frameworks in this section draw from Tanenbaum & Tanenbaum's articulation of J.L. Austin's Speech Act Theory (2012) and Suzanne Keen's Theory of Narrative Empathy (2006), providing a framework for game-makers to engage players in meaningful conversations with AI Non-Player Characters (NPCs) using large language models.

Following this, we will delve into the results of conceptual playtesting sessions using the latest and final iteration of *Carmen Won't Talk to Me*. The primary objective of conceptual playtesting is to determine whether the expressive goals of this project were achieved and to evaluate the playtesters' reception of a natural language interface using LLMs.

Reflection & Synthesis

This section contains a summary of my findings as I went through the process of making a game with LLMs that seeks to fulfill my expressive goals as a queer, Filipino maker.

Developer's Journal

Earlier in my document I identified the journal as a research method for this project, its purpose of which is to connect conceptual ideas with my prototype iterations. Much of that work is documented in the Creation chapter of this thesis, where I base my design decisions built upon theories and ideas discussed by the amazing scholars in my literature review.

Before writing my making process in this document, much of the ideas, synthesis, and annotations were first written down in my actual, physical journals.



In my undergrad as a fine arts student, much of my process is built on absolutely going mad and wrecking my sketchbooks, and that method of writing, and thinking through the act of writing and sketching, has followed me to this thesis. Even now, I find that I think best when I write first on paper before I put my ideas to the screen. Knowing this about myself, I bought several \$1.75 sketchbooks from the dollar store, and I'd have gone through two and a half of these by the end of this master's thesis.

Figure 25: Physical journals

As I review them at the conclusion of this project, I outline the schemas and frameworks that have emerged as I worked through my iterations, with the hope that these diagrams could be adopted by future game-makers. The hope is that this project may be helpful to game-makers who wish to integrate these techniques into their own practice.

The Lens of the Game-Maker + The Lens of the Player

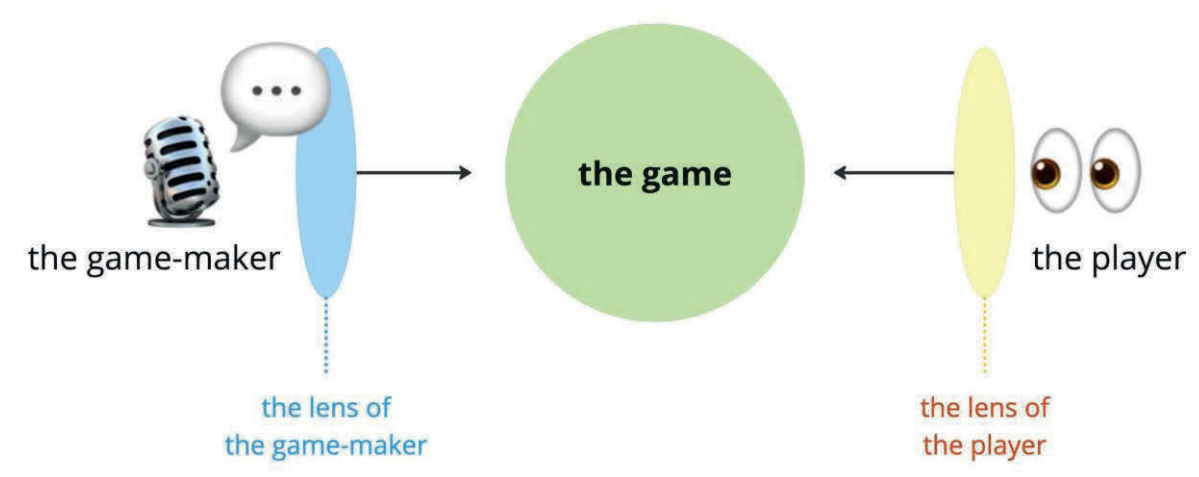


Diagram 1: *Lens of the Game-Maker + Lens of the Player*

At the beginning of this project's making, I was predominantly interested in themes of empathy and understanding, aiming to increase our capacity to care about things beyond our immediate purview. Going through this thesis prompted me to question my own values and the subjects I was interested in. As game-makers, we must always strive to be cognizant of our orientation and what our values are because those values will inevitably be reflected in the things we create (Anthropy 2012, Nissenbaum & Flanagan 2016, Anable 2018). To ensure stories that represent a wider range of human experiences, it's important to find ways to democratize game-making tools so that underrepresented makers can use them to tell stories. In delving into this thesis, I sought to investigate ways of working with AI so that more underrepresented game-makers can use these techniques in service of their own expressive goals.

In the case of *Carmen Won't Talk to Me*, my expressive goal was to invite players into my world, offering a glimpse into my youth—a common experience for many queer and Filipino women growing up in the religious schools of Manila. I aimed to convey what it was like and to deepen their understanding of queer, Filipino stories. However, it's essential to recognize that while this is the game as it exists to me, there is another game that exists to the player. Regardless of the author's intended message, the audience's interpretation will vary widely. Their perspective, values, beliefs, and worldview all color their understanding. As game-makers, we

can encourage players to engage with their values and perspective, striving to foster relevant and plausible interpretations of our work (Nissenbaum & Flanagan 2016). Considering the game-maker's own values and how they may differ from or align with the audience's is crucial for anyone aiming to broaden the audience's understanding of lives beyond their own experiences. It's vital to contemplate this aspect before delving into one's work.

Natural Language Interface – Collaboration between the LLM, the Game-Maker, and the Player

The natural language interface and mechanics of my game draw inspiration from the paper *Spoken Conversational AI in Video Games: Emotional Dialogue Management Increases User Engagement* (Fraser et al., 2018) for its approach to dialogue management, as well as from *Facade: An Experiment in Building a Fully-Realized Interactive Drama* (Mateas & Stern, 2003) for its creation of a drama manager to retain authorial control over the narrative.

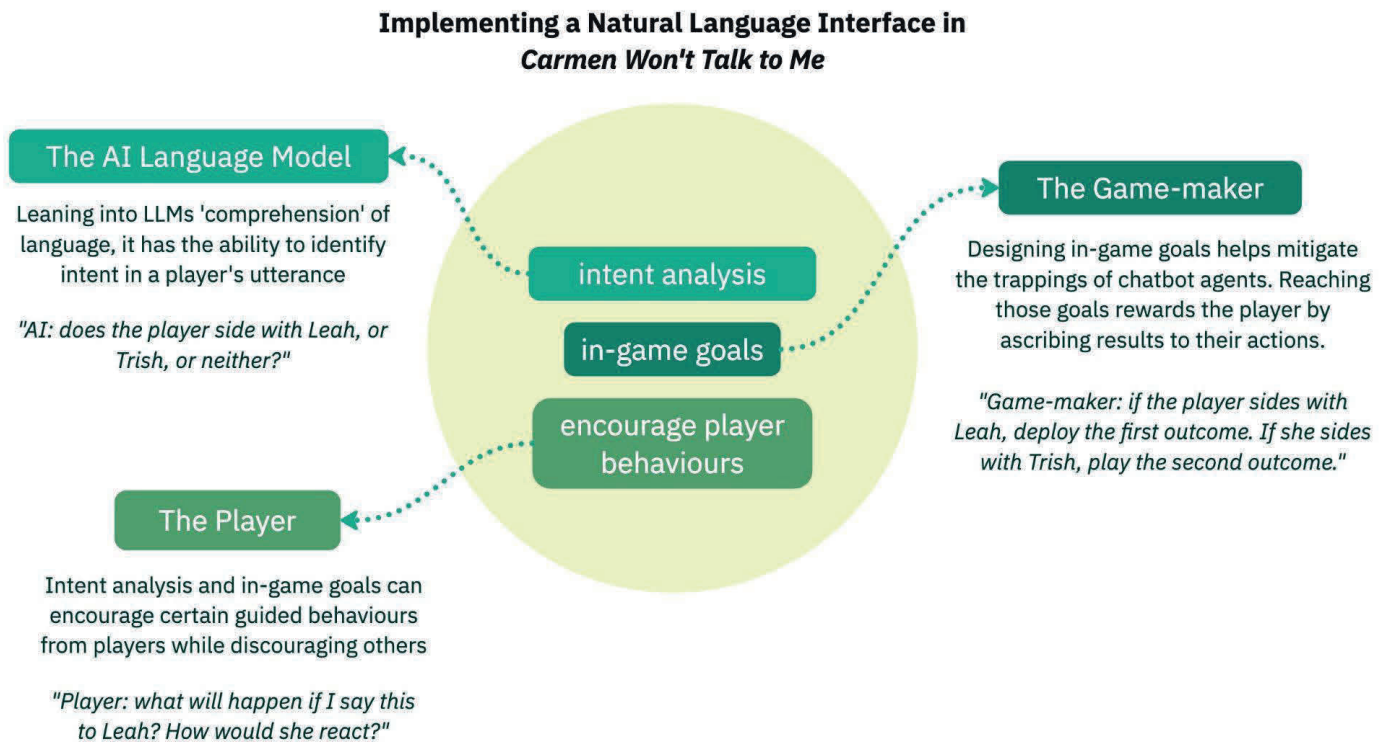


Diagram 2: Framework for implementing a natural language interface

In adapting their approach to the argument scene in *Carmen Won't Talk to Me*, I was looking at three key aspects. Each aspect acknowledges and corresponds to the three parties involved in the game experience, which is the AI, the game-maker, and the player.

Intent Analysis

As mentioned earlier, player utterances (the words that they use in the game) correspond to actions in the game, which the system responds to accordingly. For them to be recognized as actions, the game system must be able to comprehend the intention behind the player's words. By leveraging the LLMs capacity to 'understand' human languages, I instructed the LLM to conduct intent analysis on player input and tell the system if the player sided with Trish, or Leah, or if they remained neutral.

In-Game Goals

After the LLM communicates the results of intent analysis to the game system, the game then deploys certain outcomes depending on which side the player chooses. It is up to the game-maker to ascribe what these different outcomes are and the parameters needed to satisfy the conditions of each. This is important because if player utterances have no impact in the game, then their words are meaningless. Only when the system responds to a player's words that players understand that they made an action and their words meant something.

Encouraging Player Behaviours

Once players get a sense that the system is listening, analysing their words, and that saying certain things result in changes in the game world, then they can be incentivised to behave in certain ways that can merit them results. For example, players may notice that Leah is easy to capitulate, so they may try to persuade her to join Trish's side and organize a protest against the school.

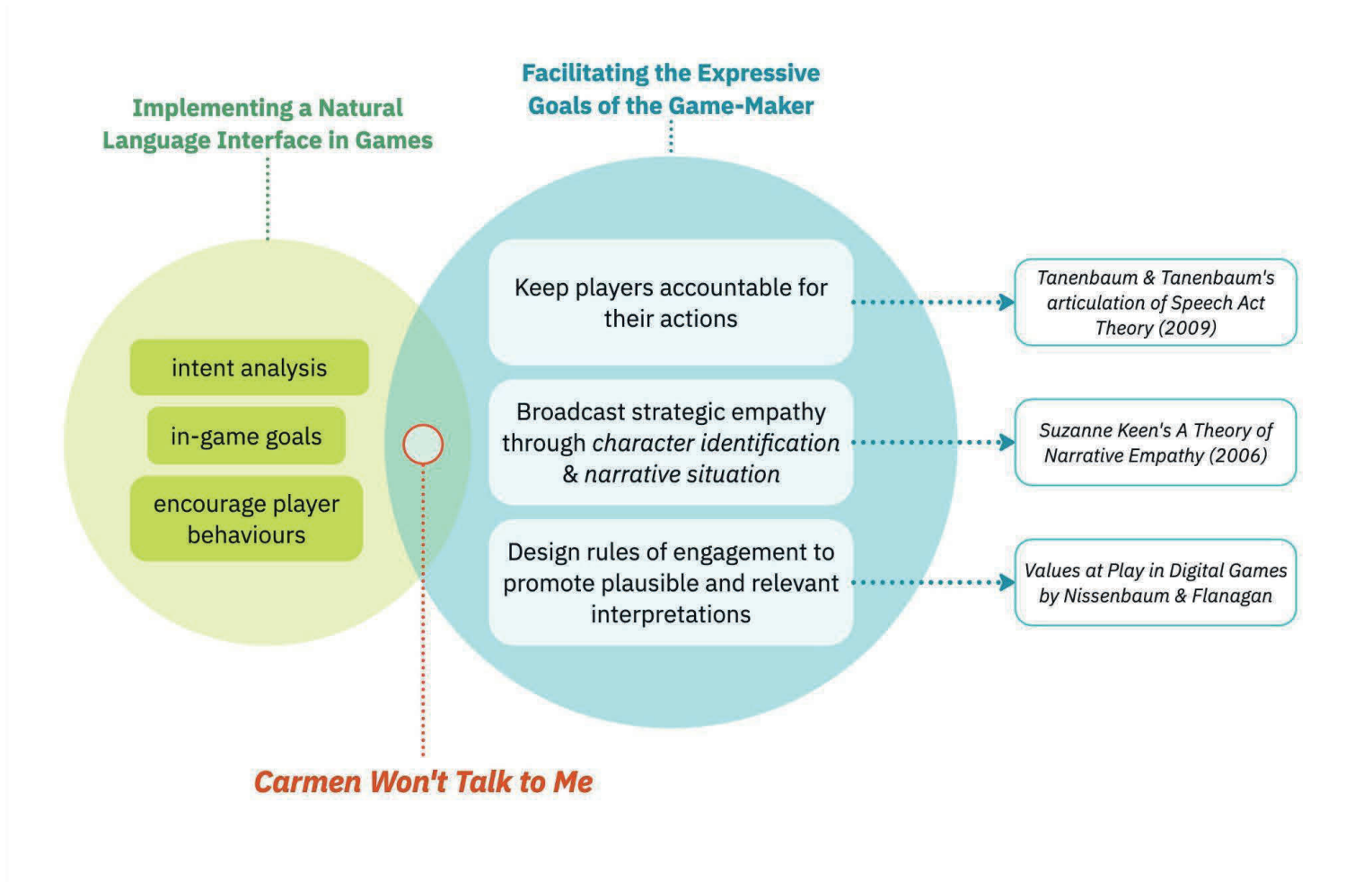


Diagram 3: Framework: Facilitating Expressive Goals with the LLM

Facilitating Expressive Goals with the LLM

Figuring out the technical side of creating a natural language interface was the objective for my first prototype, exploring the conceptual ideas in service of expressive goals was the objective of my second prototype.

There are three main conceptual ideas explored in crafting the expressive capacity of narrative-driven games like *Carmen Won't Talk to Me*. The first is to keep players accountable for the words they say and the actions they perform in the game. This in turn keeps player actions meaningful, and rewards players for committing to the actions, desires, and motives of the player character/protagonist (Tanenbaum & Tanenbaum 2009).

The second idea is, for a game that invites players to feel with the experiences of my game characters, it is important to craft a game experience that supports empathic emotion. One method of doing so is through *broadcast strategic empathy*, which endeavours to invite audiences to feel for a group of people by emphasizing our common vulnerabilities and hopes (Keen 2006). There are two techniques to do this, the first is through the *narrative situation*. In *Carmen Won't Talk to Me*, this means a first-person perspective of the game protagonist. It also means being privy to the character's internal narration, which allows us to travel with the character and to get to know their thoughts and desires, activating a cognitive effect (knowing and remembering). The second technique is through *character identification*, which Keen emphasizes is about how character motives are a greater indicator of empathetic emotion than traits. It is not through character traits where the player finds similarities between them and the character, but rather through character motives.

If a story has applied both narrative situation and character identification, then perspective-taking can occur. Games have unique affordances because not only can players empathize with character motive, but they can also become enactors of a character's will by performing actions in the game as the character, fulfilling the character's motivations. Through gameplay, players are transformed into the character. By committing to a character's actions, their desires, motives, thoughts, and feelings also become our own (Tanenbaum 2011).

The third conceptual idea is about how messages are received by the player. This is where the rules of engagement of the game are taken into consideration. When you have a game that attempts to send a message (expressive goals), you can't always control how that message is received (see prior section about Lens of the Game-maker + Lens of the Player) but you can design your game to encourage plausible and relevant interpretations (Nissenbaum & Flanagan 2016). A character's motive should be able to align with the actions that the player is allowed to do. In the case of *Carmen Won't Talk to Me*, the protagonist's main motive is to reconnect with Carmen and mend their friendship. Therefore, the primary means of navigating the game involves communication, facilitated through group chats, messaging apps, and expressing oneself freely.

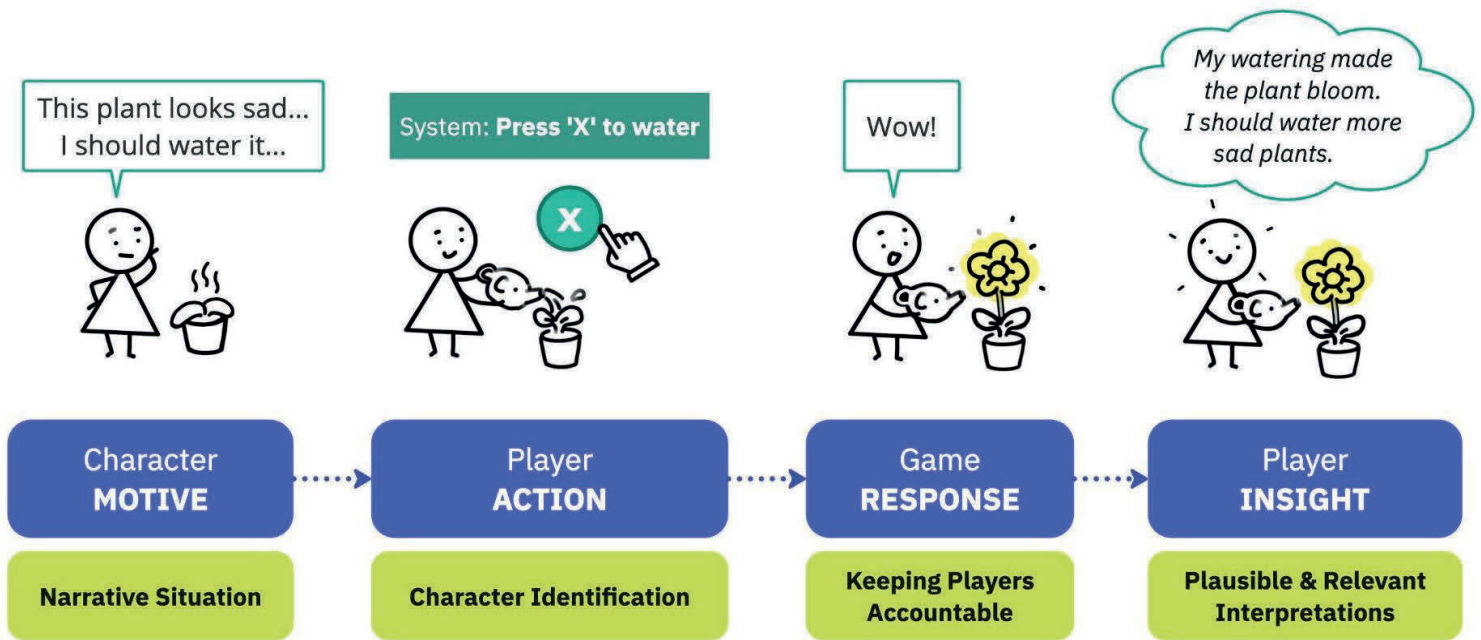


Diagram 4: Diagram of the relationship between the three conceptual ideas

Keeping Players Accountable in a Game with LLMs: Environment, Narrative, Social Effect of Language

In this section, I want to zoom in and focus on unpacking the first conceptual idea of keeping players accountable, especially how it relates to the unique affordances of working with LLMs in games.

First off, let's unpack what J.L. Austin's Speech Act Theory has to do with a game that uses natural language to interface with it. In the case of *Carmen Won't Talk to Me*, as mentioned in this thesis player utterances (*locution*) constitute as actions (*illocution*) in the game, which results in the game ascribing certain outcomes (*perlocution*) depending on what was said.

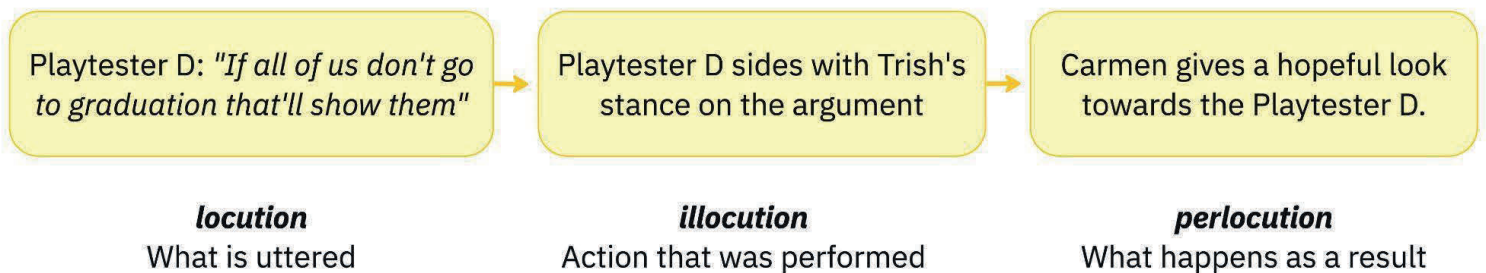


Diagram 5: *Speech Act Theory and LLMs*

For players to feel like their words had an impact on the game, *the game has to respond and change accordingly*. In my game, the game responds in three aspects. Each aspect also has an opportunity for the game-maker to implement cultural nuances, which I expound upon in [Chapter 4 \(Second Prototype: Iteration 4\)](#).

Social Effect

The dialogue responses of the game characters vary depending on the player's choices, either aligning or conflicting with the characters' unique motives, values, or worldviews. If the player agrees with Leah, with her bossy and straightforward demeanor, Leah's responses are smug and appreciative of the player's support, while Trish passionately tries to persuade the player about her perspective. Conversely, if the player agrees with Trish, Leah becomes stubborn and standoffish, whereas Trish expresses happiness that the player stands against the school with her.

Environment

After the argument ends and the protagonist is transported back to present day, what was once smudged words on the journal now becomes clear and reflects the exact words that players say in the game (See [Diagram 7](#)). Additionally, the plant on the desk looks a little better.

Narrative

In the narrative, Carmen's perception of past events changes based on the outcome of the argument. During the argument between Leah and Trish, Carmen remains silent, appearing anxious as she listens. If the player agrees with Leah, advocating for the two seniors to end their relationship, Carmen becomes upset but understands the player's stance. Conversely, if the player supports Trish in opposing the school, the group grows closer, and Carmen admires the player's courage. In the present, this gives Carmen courage and she discloses how close she was to revealing her queer identity to the player. If the player remains neutral, Carmen feels disappointed for not speaking up and supporting Trish.

LLM conducts intent analysis



Diagram 6: Keeping players accountable for their words in a game with LLMs (1)

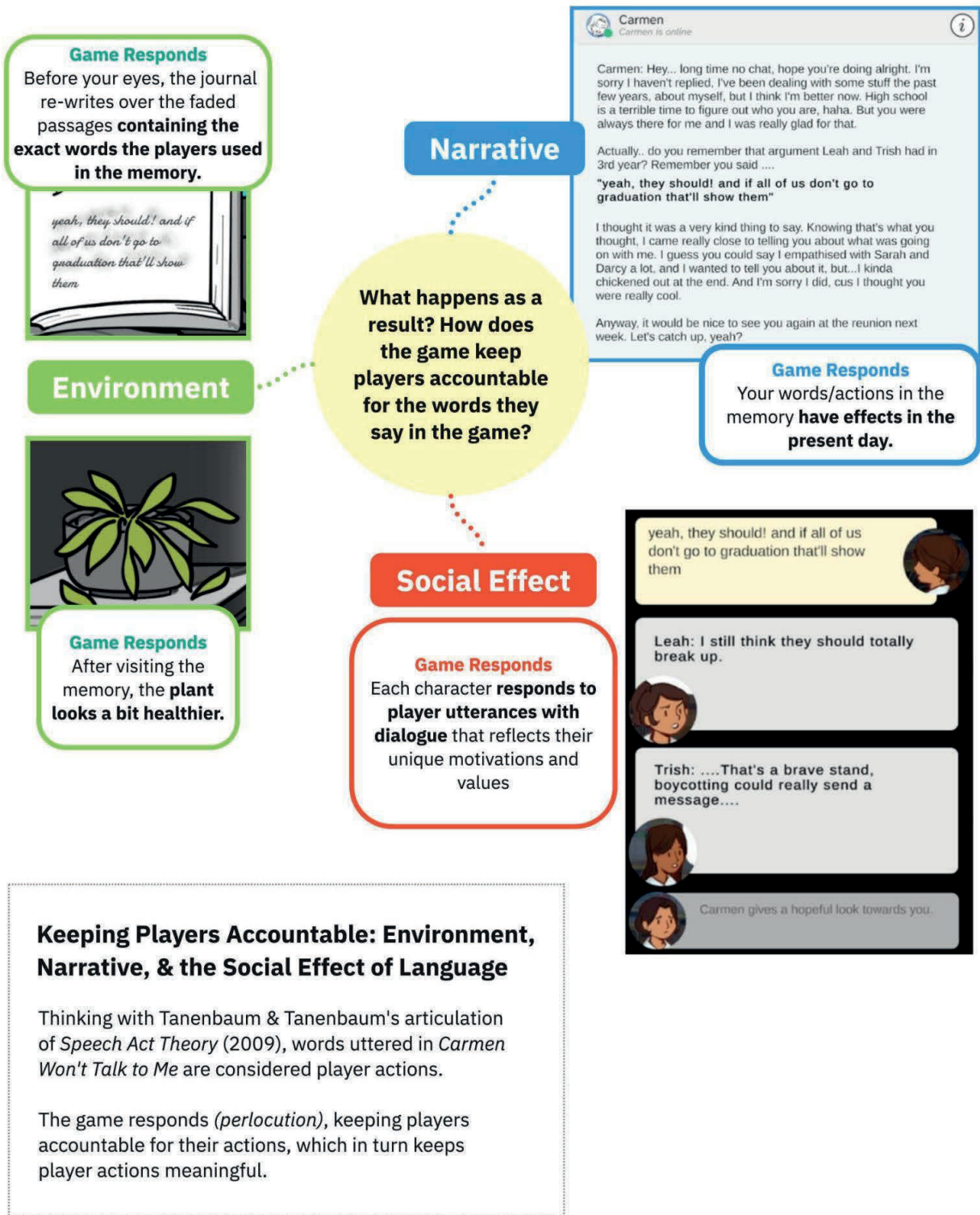


Diagram 7: Keeping players accountable for their words in a game with LLMs (2)

Conceptual Playtesting

Integrating the findings from applying conceptual ideas to iterative prototyping, I have been able to gather data from four playtesters using the latest and latest prototype. The goal of conceptual playtesting is to determine if my game's expressive goals were met.

Four individuals who have expressed interest in narrative games were invited to partake in this playtest. The sessions were conducted in February 2024, and contained on average 15-20 minutes of playtime with a 40 minute casual interview thereafter (the questions can be seen in [Appendix C](#)). Both verbal (interview) and non-verbal data (screen capture) were gathered from each playtest, which underwent thematic analysis to identify patterns and meaning in the data.

Novelty

Several playtesters have expressed interest and excitement about the novelty aspect of being able to use your own words in the game and have the characters respond accordingly to the things that they say. The playtesters also consistently contrasted the natural language interface to other narrative-driven games that use dialogue wheels or preset options, and have expressed that being about to use natural language makes for a novel experience. Playtester C expresses that “seeing that they would respond to me differently, that really makes the experience for me.” Playtesters have also expressed that it made them feel like they were part of the conversation and that the exchange felt reciprocal. Two of them have also discussed the artificiality of the experience, particularly acknowledging that it is an AI responding to them in conversation, and emphasizing that the technology is no replacement for genuine human connection, with which I wholeheartedly agree. Says Playtester D: “AI can only give and do so much [...] it's not going to replicate the same life-like quality as having a real conversation.”

The Players as Social Beings

One of the themes that came up in the casual interview was how the interface gave the players pause and it made them consider more (as compared to games with preset options) their replies or what they would say to the characters. Playtester A talked about how the group chat scene in the beginning caused her a degree of social anxiety, and that for some moments she wasn't sure how much to be herself because she didn't particularly know the characters. “Yeah,

like the technology disappears. It's like, it feels like there's real people [...] because it's a [chat] group, I'm still, like, feeling like I'm in a real social setting." The fact that the group chat scene leans into the existing affordance of the actual technology (people use messaging apps very often, for example) it had helped to strip away the artificiality for a moment and the fiction felt real for Playtester A.

Other playtesters have expressed that using their own words made them aware of the socio-cultural setting of the game and made them feel like they are part of the discussion. Playtester B expresses "[...] it encourages you to think more, especially when you're talking about topics relating to social cultural issues, if you're given responses to choose from everything is quite predictive and it's kind of a mindless action, but this one gives you blank space for you to conjure up your own thoughts."

On the other hand, Playtester D pointed out how other players may seek to exploit the fact that you could say anything and the system will respond. They express that it brings to question other players' intent, and this system may allow them to be vitriolic, bigoted, non-sequitur, or absurdist.

The game succeeds at making players aware of themselves as social beings and prompts them to consider their words before responding, but it also has the potential to invite rude or intolerant language, which is an important consideration because as mentioned earlier in this thesis, as designers we craft the rules of engagement. This consideration prompts me to think about, if a game like this is released to the public, how we might be able to mitigate undesirable behaviours in language.

The Characters

Several playtesters appreciated the small integration of Filipino slang and *Taglish* in the characters' dialogue, claiming that it made it sound more authentic (Playtester C called it 'cute'), even if they don't entirely understand what characters say but understand via context clues. They've expressed an interest in perhaps having a translation somewhere so they could learn what some Filipino terms mean.

Playtester B and Playtester D both appreciated the first person perspective of playing the protagonist, expressing that it makes it easy to relate to the character. Playtester C felt like they

were a real part of the story and that it felt impactful when they saw that what they said in the memory scene was imprinted into the protagonist's diary, feeling like they "made a mark in this game." Playtester D in particular enjoyed being able to adopt the persona of the protagonist and appreciating their malleability, and they felt a sense of control that even though Carmen won't talk to the protagonist, there is an opportunity to "right your wrong and make Carmen come back [to you]." These findings in particular were very rich as they related to how the *narrative situation* of the game helps promote *character identification* with the protagonist and allows for the adoption of the protagonist's feelings and motives.

How the playtesters felt about Carmen was a little more complicated. Playtester A sees Carmen as passive-aggressive, "I'm not sure about Carmen [...] she's been holding a grudge for like five years and then she just suddenly texts me." Other playtesters expressed a desire for more content especially since they feel that they had just started getting into it. Playtester B expressed a desire to get to know the supporting characters more, like the two seniors who were implicated at the school and what happened to them. Playtesters C and D wanted to get to know Carmen a little more. Says Playtester C: "[you find out] oh Carmen was bullied really heavily and you don't know how you fell out of her [...] when it ended I was like, am I gonna get a little more backstory on Carmen?"

From this perspective, it seems that while it was easy for players to empathize with the protagonist, the other characters presented a bit more complexity. It suggests that players may require additional interaction time with these characters to better understand their motives and personalities.

The Flexibility of Values

One thing that came up amongst playtesters was the recognition that values can change in the moment. For Playtester A, she felt a level of social pressure especially in the argument scene. She reflected on how it might be easy to capitulate to the views of the group and to cater to what they might want to hear, even if you don't necessarily share the same views or values. Playtester C noticed the same thing about the Leah character, and noted how easy it was for Leah to concede once both Trish and the player wanted to stand together against the school. Playtester C interpreted this as a condition of the characters both being highschool girls and

how easy it might be to yield to the views of the group and feel that social pressure especially at that age.

That said, all playtesters were steadfast in choosing a side and sticking with the side they chose. Playtesters B, C, and D all chose to stand up against the school, either through protest, or boycott, or questioning school authority. Playtester A however sided with Leah and believed the two seniors should break off their relationship.

Common Vulnerabilities and Hopes

What came up conversationally with all playtesters was how they related their own personal experiences to the events in *Carmen Won't Talk to Me*.

For Playtester A, who was the only one to side with Leah, they related their own experiences in a religious school in their home country. Thus, having the school in the game ban relationships was something that they and their classmates had to deal with as well. They commented that the idea of fighting the school felt like a Western concept to them. In their hometown, students might rebel a bit, but not to the degree of winning or losing against authority.

After playing through the events of the game, Playtester B realized that students in the Philippines experience a "lot of pressure outside of academic pressure, which is something I have no prior knowledge of." In their experience, high school was full of academic pressure only. "Outside of that, you could do whatever you want as long as you have good grades. [...] I feel sorry for the characters."

Playtester C's reflection focused less on the school environment and more on girlhood relationships. They expressed appreciation for the game being "girl-centric." Most notably, they were reminded of a close friendship from high school that ended due to unrequited feelings on their friend's part. Playtester C shared about this friendship: "I actually had a girl best friend in high school who was queer and she liked me but as a stupid teenage girl I didn't know how to handle it and so essentially we just kind of fell out of each other."

Playtester D recalled moments in their life when they wished they had stood up for others facing discrimination, but found it challenging due to circumstances in their home country. They appreciated the commonality between the Philippines and their own country, both having

experienced colonization, and recognized that prior to colonization, these spaces might have been safer for queer individuals. Playing *Carmen Won't Talk to Me* allowed them to empathize with the struggles of queer identity in the Philippines, resonating with their own experiences of living in similar circumstances. Playtester D reflected, "[...] that was my reality. So that's what I mean, I was feeling the weight. It thrust me back into those times."

It's About Friendship, and Growth, and Relationships, and Being Accountable

As mentioned consistently over the course of this document, as a game-maker, your messages will not be received uniformly. Therefore, it is important to design the rules of engagement to foster relevant interpretations of your game.

Similarly, when I asked the playtesters what *Carmen Won't Talk to Me's* themes and story was about, each individual's response differed but were all valid and very rich.

For Playtester A, the game was about navigating human relationships and reflecting on us as social beings, and how a conflict of values might play out in a social interaction.

For Playtester B, it was about growth, of being a teenager transitioning into adulthood and having to "navigate between what you believe is just and what the system tries to enforce." According to them, the story about queer identity is foregrounded by this larger conflict of young people struggling against authority.

Playtester C shares that to them, the game is about girlhood, queerness, and reconciling with old friends. It's about coming to terms with your past self that might've disappointed you, "like maybe my friend was bullied in high school but I didn't do anything to stop it."

Playtester D says that the game is about friendship and being accountable to that friendship. They expound on this, sharing: "You cannot pour acid into a flower's soil and expect it to grow [...] If your opinion invalidates the humanity of another person, [...] that person has every right to not want to be friends with you. So that, I feel, is the beating heart of this game."

Final Thoughts

What was particularly intriguing about the conceptual playtesting was the correlation between how players related to the game and its story, and their interpretation of its themes. For

instance, Playtester B, who experienced primarily academic pressure from school, interpreted the story's theme as the conflict between young people and systems of authority. In contrast, Playtester C, who recalled a personal falling out with a queer friend, saw the theme as being about girlhood friendships and reconciliation. Additionally, there was a consistent theme among playtesters of feeling involved in the conversation and the importance of speaking up against oppression or being held accountable for the words we say to our friends.

These findings align with the conceptual ideas and frameworks explored throughout this thesis, particularly how the player's perspective or 'lens' influences the affective capacity of their experience with the game and, consequently, their interpretation of it. Each playtester offered a unique interpretation colored by their memories and experiences, and as game-maker, all interpretations were valid (and celebrated) and aligned with my expressive goals. These interpretations facilitated connections between the players and the characters of *Carmen Won't Talk to Me*, revealing common vulnerabilities and hopes.

On a personal level, I am deeply grateful for the time my playtesters dedicated to my game, and their interpretations of the game touched me profoundly. Through this fictional narrative, inspired by my experiences as a queer, Filipino maker, I felt in a sense that I was understood.

Democratizing game-making tools is crucial to enable more underrepresented game-makers to delve into expressive qualities in personal games. Embracing games as an interactive, reciprocal medium, that, in a way, manifests as a conversation between the game-maker and her players.

Chapter 6: Future Work & Conclusion

I started this project because I loved games and, as many makers do, I love the act of creation. I have learned so much from this experience but it is now the final chapter and there is still work that needs to be done. The heart of this project is being able to document these findings and techniques so that future game-makers can take them and integrate LLMs into their own processes. The vision for the future of this project lies in being able to disseminate this knowledge.

Publishing the Journal

One of the visions of this thesis is to publish its process and findings online so that it may be accessed and referenced by future game-makers. The vision is that these notes may be accessed via Itch or a dedicated Wordpress site, that could follow a blog format that includes my thoughts, ideas, concepts, issues, and breakthroughs. I am thankful to all the teachers and makers who have made their processes available online so that makers like myself could access and learn from them. This project would not have been possible without their work, and so it would be wonderful to be able to pass down my processes myself in the future so that underrepresented game-makers could take my findings and make their own stories.

Workshops for Game-Makers

The next step/future work to make the findings in this project accessible to future game-makers is to create a space for learning and creation. The workshop will serve as a tool to have game-makers get in touch with the methods and techniques of making games with LLMs.

One thing that is lacking from my project is being able to connect with game narrative designers and discuss with them the potential of using LLMs as an approach to dialogue and narrative design. It would be wonderful to host workshops in the future and enjoy the community of makers and see what they would make with these techniques. I want to know what stories they would create and learn from them as well.

Conclusion

In this project, I have explored methods for incorporating AI Large Language Models (LLMs) into narrative-driven games to support the expressive goals of underrepresented game-makers. I began by emphasizing the significance of amplifying marginalized voices, particularly given the current dominance of cis, white, and male individuals in the game development industry. Despite these challenges, underrepresented game-makers persist in advocating for their stories through publishing independently and the creation of their own platforms. Therefore, it is important that we continue to democratize game-making tools to empower a broader spectrum of creators. This endeavor led me to examine various approaches and techniques that creators like myself can apply to more easily integrate LLMs into their existing game-making practices if they so choose. That being said, AI remains to be a deeply flawed tool and its issues cannot be overlooked.

How This Project Fails

As of this writing in the early months of 2024, the future of artificial intelligence and creative work remains contentious. As previously discussed in the literature, the remarkable capabilities of generative AI stem from its training, often without constraints, on the vast dataset provided by the internet. Today, writers, artists, and other creators who have shared their work online are being exploited by AI companies, which utilize their creations to develop AI products. In a capitalistic society, these products seek to replace the very creatives whose work was appropriated. This issue lies at the heart of this project. While it strives to democratize AI tools to empower creators to understand the technology and regain control, it still poses a potential failure in serving the makers it aims to support.

In an article about artificial intelligence written by Ted Chiang (2023), the author proposes AI as a new form of the management consulting firm McKinsey, and its trajectory to become a tool that serves to make rich people richer at the expense of labor. Meanwhile, AI tech advocates claim that AI will ease human labour and bring us closer to a world with universal basic income

because everything that needs getting done is automated. However, as pointed out by Chiang, that is currently not the world we live in.

Oftentimes in sci-fi films, there is a fear that AI will gain complete autonomy from humanity and will stop anyone who tries to shut it down. However, Chiang expresses that the real machine is capitalism, and capitalism will do whatever it takes to stop us from ever turning it off, and prevents us from imagining other possibilities.

That being said, most of us probably aren't legislators, AI scientists, or management consultants. Our job as creatives is to make meaning out of the world; what can we do in the face of forces bigger than us?

As I see it, as creatives and makers, there are two roads ahead of us in the conversation around AI. We can choose to fight back against it or engage with it. Both are completely valid approaches. Currently there have been several lawsuits by artists and writers alike pushing back against AI companies and copyright infringement of their stolen works that were used to train generative AI products (Grynbaum & Mac 2023; Brittain 2023). On the other end, researchers have been working with artists to create tools like *Nightshade* which adds an invisible film on top of digital work before they're uploaded, which infiltrates generative AI and corrupts their dataset, rendering their work unusable by the AI and also serves to poison it (Heikkilä 2023). These are both amazing things.

For myself, I celebrate this resistance, and I would trade for a world without generative AI for a heartbeat if it meant that artists, writers, and creators are valued and compensated for their work, and are afforded the dignity to pursue their craft without the risk of exploitation.

On the other hand, I also acknowledge, as Ted Chiang and many others have pointed out, that reaching such a point is a long and arduous journey, and it is not the world we currently inhabit. Therefore, the second option is to engage with AI, which is the stance I take with this project. I entered into this project both fascinated and cautious of AI, adopting the mindset of "it is better to know the devil you know than the devil you don't." So, now, at the conclusion, what are my thoughts on this matter?

Making Kin with the AI Era

In summary, AI is all sorts and shades of kind-of bad, but it can help a creator make some interesting things. On its own, generative AI produces much of the same, and though it might come out with a shiny polish, it is still the product of an algorithmic system who executes 1's and 0's to make things that are an amalgamation of all the things that have already been made before.

All this to say, if generative AI tools were in the hands of someone who lacked care, the work limits itself, because humans are smart and humans can often tell the difference. Artists and makers care, and that love of creation will shine through in the things they create. There will always be an opportunity for creatives to contribute something to the world that AI cannot replicate. If makers were to lean into the unique affordances of artificial intelligence if they so choose, then I'd be excited to see them pave the way for new and bolder things that we haven't seen before.

Bibliography

- Anable. (2018). *Playing with Feelings: Video Games and Affect*. University of Minnesota Press.
- Anthropy, Anna. n.d. "Dys4ia by Anna Anthropy." Itch.io. Accessed March 14, 2024. <https://w.itch.io/dys4ia>.
- Anthropy, A. (2012). *Rise of the Videogame Zinesters: How Freaks, Normals, Amateurs, Artists, Dreamers, Dropouts, Queers, Housewives, and People like You Are Taking Back an Art Form*. Seven Stories Press.
- Antonisse, Jamie, and Devon Johnson. n.d. "Hush." Accessed March 14, 2024. <http://www.jamieantonisse.com/hush/about.html>.
- Austin, J. L. (2003). *How to Do Things with Words* (2nd ed.). Harvard University Press.
- Belman, Jonathan, & Flanagan, Mary. (2010). "Designing Games to Foster Empathy." *Cognitive Technology*, 14(2), 5–15.
- Bizzocchi, J., & Tanenbaum, T. J. (2012). "Mass Effect 2: A Case Study in the Design of Game Narrative." *Bulletin of Science, Technology & Society*, 32(5), 393–404.
- Boroditsky, L. (2011). "How Language Shapes Thought." *Scientific American*, 304(2), 62–65.
- BioWare. *Mass Effect 2*. Electronic Arts. PC/Xbox 360. 2010. [Video game]
- Braun, Victoria, and Virginia Clarke. 2006. "Using thematic analysis in psychology." *Qualitative Research in Psychology* 3: 77-101.
- Brittain, Blake, and Blake Brittain. 2023. "Artists Take New Shot at Stability, Midjourney in Updated Copyright Lawsuit." *Reuters*, November 30, 2023, sec. Litigation. <https://www.reuters.com/legal/litigation/artists-take-new-shot-stability-midjourney-updated-copyright-lawsuit-2023-11-30/>.
- Canadian Charter of Rights and Freedoms, s 2, Part 1 of the Constitution Act, 1982, being Schedule B to the Canada Act 1982 (UK), 1982, c 11.
- Carpenter, Nicole. 2024. "What's Going on with All These Video Game Industry Layoffs?" Polygon (blog). February 19, 2024. <https://www.polygon.com/gaming/24074767/video-game-industry-layoffs-explainer>.
- Chesler, Caren. 2022. "Can Video Games Teach People to Be More Empathetic? Maybe." *Washington Post*, July 26, 2022. <https://www.washingtonpost.com/health/2022/06/25/video-games-empathy/>.
- Chiang, Ted. 2023. "Will A.I. Become the New McKinsey?" *The New Yorker*, May 4, 2023. <https://www.newyorker.com/science/annals-of-artificial-intelligence/will-ai-become-the-new-mckinsey>.
- Edwards, Phil. 2023. "Why AI Doesn't Speak Every Language." *Vox*. April 11, 2023. <https://www.vox.com/videos/2023/4/11/23679017/ai-language-models-resource-datasets-learning>.

- Eloundou, T., Manning, S., Mishkin, P., & Rock, D. (2023). "GPTs are GPTs: An Early Look at the Labor Market Impact Potential of Large Language Models (arXiv:2303.10130)." *arXiv*.
- Flanagan, M., & Nissenbaum, H. (2016). *Values at Play in Digital Games*. The MIT Press.
- Fraser, J., Papaioannou, I.V., & Lemon, O. (2018). "Spoken Conversational AI in Video Games: Emotional Dialogue Management Increases User Engagement." *Proceedings of the 18th International Conference on Intelligent Virtual Agents*.
- Freytag, Gustav, and Elias J. MacEwan. 1900. *Freytag's Technique of the Drama : An Exposition of Dramatic Composition and Art. An Authorized Translation from the 6th German Ed. by Elias J. MacEwan*. Chicago : Scott, Foresman. <http://archive.org/details/freytagstechniqu00freyuoft>.
- Games for Change, dir. 2019. *Beyond the Empathy Games 101: Digging Deep into Empathy, Ethics, and Design with Kelli Dunlap*. <https://www.youtube.com/watch?v=ywk3botU-iA>.
- Grynbaum, Michael M., and Ryan Mac. 2023. "The Times Sues OpenAI and Microsoft Over A.I. Use of Copyrighted Work." *The New York Times*, December 27, 2023, sec. Business. <https://www.nytimes.com/2023/12/27/business/media/new-york-times-open-ai-microsoft-lawsuit.html>.
- Heikkilä, Melissa. n.d. "This New Data Poisoning Tool Lets Artists Fight Back against Generative AI." MIT Technology Review. Accessed March 12, 2024. <https://www.technologyreview.com/2023/10/23/1082189/data-poisoning-artists-fight-generative-ai/>.
- "How ChatGPT and Our Language Models Are Developed | OpenAI Help Center." n.d. Accessed March 6, 2024. <https://help.openai.com/en/articles/7842364-how-chatgpt-and-our-language-models-are-developed>.
- "How to Use ChatGPT in Unity." 2023. Immersive Limit. March 21, 2023. <https://www.immersivelimit.com/tutorials/how-to-use-chatgpt-in-unity>.
- Introduction | Inworld AI. (n.d.). Retrieved May 25, 2023, from <https://inworld.ai/docs/intro>
- Irving, G., & Askill, A. (2019). "AI Safety Needs Social Scientists." *Distill*, 4(2), e14.
- Judgefloro. 2016. *English: Gabriel Beato Francisco March 18, 1850-December 19, 1935 Novelist Historical Marker Landmark October 1999 along Mariano Fortunato Jhocson (Lipa) Street (in List of Barangays of Metro Manila Legislative Districts of Manila Barangays 401, 402 & 403, Zone 41, 404, 405, 406, 407, 408, 409, 410, 411, 412, 413, 414, 415 & 416, Zone 42, District IV, Sampaloc, Manila, in Front of Barangay 638, Zone 65, District VI, in San Miguel, Manila, City of Manila (Accessed from Legarda Street List of Roads in Metro Manila Corner J. Figueras (Bustillos) Street-Earnshaw Street Corner Lardizabal Street Corner Manrique Street towards Plaza Jose Figueras Street (Formerly Bustillos Street))*. Own work. https://commons.wikimedia.org/wiki/File:07183jfLandscape_Barangays_Sampaloc_Legarda_Streets_Manilafvf_04.jpg#filelinks.
- Keen, S. (2006). "A Theory of Narrative Empathy." *Narrative*, 14(3), 207–236.
- Lewis, J. E., Arista, N., Pechawis, A., & Kite, S. (2018). "Making Kin with the Machines." *Journal of Design and Science*. [DOI](<https://doi.org/10.21428/bfafd97b>)
- Lyons, John. 1991. *Natural Language and Universal Grammar: Volume 1: Essays in Linguistic Theory*. Cambridge University Press.

- Maguire, Moira, and Brid Delahunt. 2017. "Doing a Thematic Analysis: A Practical, Step-by-Step Guide for Learning and Teaching Scholars." *All Ireland Journal of Higher Education* 9 (3). <https://ojs.aishe.org/index.php/aishe-j/article/view/335>.
- Mateas, M., & Stern, A. (n.d.). "Façade: An Experiment in Building a Fully-Realized Interactive Drama."
- Noble, S. U. (2013). "Google Search: Hyper-Visibility as a Means of Rendering Black Women and Girls Invisible." *InVisible Culture: An Electronic Journal of Visual Culture*, 19.
- Naughty Dog. *The Last of Us*. Sony Interactive Entertainment. PlayStation 3. 2013. [Video game]
- Naughty Dog. *The Last of Us Part II*. Sony Interactive Entertainment. PlayStation 4. 2020. [Video game]
- Pasquinelli, M., & Joler, V. (2021). "The Nooscope manifested: AI as instrument of knowledge extractivism." *AI & SOCIETY*, 36(4), 1263–1280. [DOI](<https://doi.org/10.1007/s00146-020-01097-6>)
- Pincombe, Roger. n.d. "OpenAI API for C# / .NET." OpenAI API for C# / .NET. Accessed February 15, 2024. <https://rogerpincombe.com/openai-dotnet-api>.
- Quantic Dream. *Detroit: Become Human*. Sony Interactive Entertainment. PlayStation 4. 2018. [Video game]
- Roose, Kevin. 2023. "Don't Ban ChatGPT in Schools. Teach With It." *The New York Times*, January 12, 2023, sec. Technology. <https://www.nytimes.com/2023/01/12/technology/chatgpt-schools-teachers.html>.
- Samuel, S. (2021, September 18). "AI's Islamophobia problem." *Vox*. [Online Source](<https://www.vox.com/future-perfect/22672414/ai-artificial-intelligence-gpt-3-bias-muslim>)
- Samuel, S. (2022, April 19). "Why it's so damn hard to make AI fair and unbiased." *Vox*. [Online Source](<https://www.vox.com/future-perfect/22916602/ai-bias-fairness-tradeoffs-artificial-intelligence>)
- Santa Monica Studio. *God of War*. Sony Interactive Entertainment. PlayStation 4. 2018. [Video game]
- Searle, J. (1965). *Speech Acts*. Cambridge University Press.
- Tanenbaum, K., & Tanenbaum, T. J. (2009). "Commitment to Meaning: A Reframing of Agency in Games." *UC Irvine: Digital Arts and Culture 2009*.
- Tanenbaum, T.J. (2011). "Being in the Story: Readerly Pleasure, Acting Theory, and Performing a Role." In M. Si, D. Thue, E. André, J.C. Lester, T.J. Tanenbaum, & V. Zammitto (Eds.), *Interactive Storytelling. ICIDS 2011* (Lecture Notes in Computer Science, vol 7069). Springer.
- Toh, W., & Lim, F. V. (2022). Learning in digital play: A dual case study of video gamers' independent play. *Research and Practice in Technology Enhanced Learning*, 17(1), 6. <https://doi.org/10.1186/s41039-022-00182-2>
- Tolkien, J. R. R. 1991. *The Lord of the Rings*. London, England: HarperCollins.
- Urbanek, J., Fan, A., Karamcheti, S., Jain, S., Humeau, S., Dinan, E., Rocktäschel, T., Kiela, D., Szlam, A.D., & Weston, J. (2019). "Learning to Speak and Act in a Fantasy Text Adventure Game." *ArXiv*, abs/1903.03094.

Appendices

Appendix A: Story Outline for *Carmen Won't Talk to Me*

I. Exposition

Present day, March 2018. A rainy day in Manila. "Has anyone heard from Carmen?" You look back on high school memories.

You (the player) see a desktop background. A messaging app is open. It's a group chat of your high school friends. Everyone is excited about the upcoming high school reunion. Someone is missing, however.

"Has anyone reached out to Carmen?"... "I haven't heard from her in a while."... "Weren't you guys close?"... "We haven't talked lately. I'll try reaching out."

You close the messaging app and familiarise yourself with your desktop. There are three options to click: a folder of pictures (containing a group photo of the four of you), an icon for the group chat, and another icon for your chat with Carmen. You try to reach out to her, but she's offline and your messages are not responded to.

You pull away from your laptop and see your entire desk where you are sitting. There are four objects in front of you: your laptop, your high school journal, a desk lamp, and a potted plant from Carmen. It's been years and it's withered by now.

II. Rising Action

5 years ago, March 2013. Junior year. An argument in the classroom.

You pick up your high school journal and reminisce about that time in the classroom. It's old and the pencil marks have faded, and you think, "How did that argument go again..?"

As you try to remember, you are sucked into the memory. You find yourself back in the high school classroom, before class starts. There's a girl at the back of the classroom sobbing, and a friend is beside her comforting her.

You see two of your friends having a whispered, passionate conversation at the front of the class. Carmen is sitting by, looking pensive and unsure. You approach your friends and they drag you into their argument.

Leah: *Hey, where have you been? We've been looking everywhere for you. Things have been crazy over here.*

Trish: *Yeah...We've just been discussing Sarah and Darcy's situation with the school. Sarah's been crying and Leah thinks they should just break up to avoid trouble. I don't agree with her at all.*

III. Climax

Leah and Trish ask, "which side are you on?" Carmen listens to what you say.

After you get a hold of the situation and Leah and Trish each express their opinion, they finally they turn to you and ask for your opinion and side in the argument.

Leah: *In any case, I think they should just break up and avoid trouble with the school. This whole thing is way too dramatic! They should just break up, graduate, and get the hell out of here.*

Trish: *That's not fair to them, Leah...The school shouldn't force them to break up like that...Let's all talk to the school principal...They'll have to listen to us then...*

What do you think about all this?

Game Responds: Carmen is listening to your every word. Depending on whose side the player chooses, she may either appear uncomfortable, or look hopefully in their direction.

IV. Falling Action

Back to present day. The words in the journal changes, reflecting your words. The memory is complete.

Either Trish or Leah is happy when you agree with her. The other character is conciliatory. Eventually class starts and everyone goes back to their seats. You are dragged out of the memory and are back in the present day, holding the journal in front of you.

Game Responds: The journal re-writes over the faded passages containing the exact words the players used in the memory.

You think to yourself, “Ah right, that’s what happened. Poor Sarah... I do wonder what happened to her and Darcy after they left the school.”

V. Denouement

Carmen's gift changes. She finally replies to your message.

After visiting the memory, the potted plant on your desk (Carmen's gift) will visibly change. You get a ping from your laptop, it’s a new chat notification. After so long, Carmen has finally replied to your message.

Game Responds: There are three possible outcomes to Carmen’s final message, depending on what the player says in the argument.

Appendix B: Recruitment Protocol

The recruitment of participants will go as follows:

1. Solicitation of participants will be through word-of-mouth (approaching peers and interested parties) and online recruitment. Online channels will include the researcher's own social media platforms (Instagram, Twitter) and gaming community Discord Servers which I am a part of (with permission of server moderators).
 - a. Online posts will take on a casual, personable tone as I will be using my own voice in my posts. Example post text: "Hi everyone! I'm making a game about queer, Filipino girlhood and AI language models. I'm looking for players to come talk to my AI NPCs and roleplay with them—using your own words. Interested? Answer the form below and let's get in touch! Learn more about my research through this [LINK to Overview]."
 - b. The post will include a screenshot, GIF, or looping video of the prototype's art and gameplay to increase interest.



Above: Sample graphics for post examples of gameplay.

- c. I will be reaching out to Discord server moderators through direct message with the following: “Hello! My name is Tamika, I’m a graduate student from OCAD U in Toronto and I’m studying prosocial and artistic expression in games with language models. I’d love to reach out to the community if they would be interested in participating in a study. Would it be alright to send this message to the server?”
1. The link will direct interested parties to a reservation form to submit their name and e-mail addresses. Once the solicitation period is over, applicants will be e-mailed with a consent form and an overview of the functional testing process. After participants are secured for functional testing the submitted list of names and e-mail addresses will be deleted.
 2. Several weeks after functional testing is over, 3-5 participants who have expressed interest in narrative-driven games in the survey form will be invited to participate in the conceptual testing of the game. This invitation will take place using formal e-mails.

Appendix C: Questionnaires

Functional Testing - Online Survey (30 minutes)

[Survey Page 1]

For this survey, I would like you to take on various perspectives and respond to different conversation prompts. You will be role-playing as a high school student and asked to take a side in an argument between two close friends. You will have the opportunity to give a Positive, Negative, and Neutral response to each prompt. Your responses will be used to train an AI language model and help it become more reliable in handling various player responses.

Content warning: some of the scenarios feature discrimination based on one's queer identity, and may cause feelings of anxiety or isolation in the participant. Please know that you may withdraw from this survey at any point if you feel at all uncomfortable. For a list of support organizations, see <https://www.secutoronto.org/lgbt-resources>.

If you need more information or have any questions about this study, you can contact the investigator, Tamika Yamamoto, via email at tamikayamamoto@ocadu.ca. This study has been reviewed and received ethics clearance from the Research Ethics Board at OCAD University, with clearance number [2023-72]. If you have any comments or concerns, you can reach out to the Research Ethics Office/REB at research@ocadu.ca.

Definitions:

Positive Response - indicates agreement, approval, or encouragement

Negative Response - expresses avoidance, withdrawal, or refusal, and can also be unpleasant

Neutral Response - does not pick a side; neither agrees nor disagrees with a statement

Responses can be in the form of either a **statement** or a **question** (limit to 50 words).

Situation #1

You walk into the classroom to find your two close friends engaged in a passionate conversation. As you approach, they pull you into their argument. You can make a statement or ask a question.

Leah: Hey, where have you been? We've been looking everywhere for you. Things have been crazy over here.

Trish: Yeah...We've just been discussing Sarah and Darcy's situation with the school. Sarah's been crying and Leah thinks they should just break up to avoid trouble. I don't agree with her at all.

Positive Response: _____

Negative Response: _____

Neutral Response: _____

Situation #2

Leah and Trish present their arguments and ask for your opinion after discussing them with you.

Leah: In any case, I think they should just break up and avoid trouble with the school. This whole thing is way too dramatic! They should just break up, graduate, and get the hell out of here.

Trish: That's not fair to them, Leah...The school shouldn't force them to break up like that...We should come together and go against the school principal...They'll have to listen to us then..

[Trish looks towards you] What do you think about all this?

Positive Response: _____

Negative Response: _____

Neutral Response: _____

[Survey Page 2]

Thank you for your participation in this survey! Before you go, I'd like to get to know you a bit as a person who plays games. What are your preferences? *(You can choose more than one.)*

- I like action-packed games with fast-paced thrills, surprises, and mayhem. Guns, swords, explosions, and general chaos.
- I enjoy games that require strategic thinking and mastery, where I must practice and plan ahead to tackle rising challenges.
- I like learning the stories of other game characters. I enjoy getting to know them and forming friendly relationships with them.
- I enjoy playing games that offer me the opportunity to learn about various cultures and communities, allowing me to empathize and educate myself.
- I like being immersed in a story-world. It's exciting to be someone else in another place, with elaborate plots and interesting characters.

Conceptual Playtesting - Post-Game Interview (audio only - 1 hour)^[1]_{SEP}

The interview will begin after gameplay has ended. Participants will be reminded of their consent, including their ability to withdraw or not answer.

“Thank you [name of interviewee] for joining me today and for playing *Carmen Won’t Talk To ME*. I want to remind you of your consent, and that you have the right to withdraw from this study or decline to answer any questions.”

What follows is a casual interview regarding their experience, guided by the following questions:^[1]_{SEP}

- Talk about your experience playing the game.
- How did it feel having the ability to use your own words to navigate the game?
- Did you feel like your words had an impact on the world (character, narrative)?
- What aspects of the experience did you enjoy?
- Is there something you wish could have been different?
- What is something that you know now, after playing the game, that you didn’t know before?
- What is the story? What are the themes?

Appendix D: Written Transcripts

Written Transcripts of Playtester A

00:03:51: For me, what's most prominent for me was the kind of social beings we are. I think that that felt like the story to me like how we interact with other people [...] when you're socializing with people they have their own values I have my own values, and how does that play out, like, in the social situations. That's what it felt like for me.

00:13:30: From where I'm from, you might rebel a little bit, but not to this degree of where the school wins or the relationship wins.

00:33:40: I'm not sure about Carmen as a person you know because she's been holding a grudge for like what five years and then she just suddenly texts me...

00:41:30: Yeah, like the technology disappears. It's like, it feels like there's real people. It's only in those moments I was saying I didn't ask who else is coming and I didn't get responses. But, you know, that then reminded me that, oh, this is like, you know, technology. But still, at the end of the day, because it's a WhatsApp group, I'm still, like, feeling like I'm in a real social setting.

Written Transcripts of Playtester B

00:02:28: There's the first-person aspect of it makes the player easy to relate to the protagonist.

00:05:26: ...it encourages you to think more, especially when you're talking about topics relating to social cultural issues, if you're given responses to choose from everything is quite predictive and it's kind of a mindless action but this one gives you blank space for you to conjure up your own thoughts.

00:15:10: I don't know how representative it is of just public schools in the Philippines or public Catholic schools in the Philippines but I do get a sense that there's a lot of pressure outside of academic pressure which is something I have no prior knowledge of. My high school is full of academic pressure and academic pressure only. Outside of that you could do whatever the shit you as long as you have good grades.

00:17:30: How do you navigate between what you believe is just and what the system tries to enforce on you, when you're not a full adult and sometimes can't be taken seriously by authoritative figures.

Written Transcripts of Playtester C

00:01:35: ...seeing that they would respond to me differently, that really makes the experience for me.

00:04:00: I think it was especially impactful when I saw that what I said last in the memory scene was actually etched onto my own diary, so it's like me making my mark in this game.

00:06:33: oh Carmen was bullied really heavily and you don't know how you fell out of her but I guess that wasn't really like part of the gameplay for this scene.

00:07:20: I was like oh, am I gonna get a little more backstory on Carmen? And where the relationship went wrong, right?

00:09:37: I think it's about reconciling with old friends, reconciling with maybe old parts of yourself that might be a little bit ugly, like oh, like maybe my friend was bullied in high school but I didn't do anything to stop it. And the themes, well, it's girlhood, it's also being queer.

00:10:50: ...I feel like what Trish said and she was like very quick to kind of let go of her own opinion [...] from another like viewpoint where it's like girls are kind of like that as teenagers it's like if the majority of them agree to one thing even if you are on the strongly believed in what you believed in. If the majority of the people don't agree with you, you're probably just going to also agree.

00:17:05: I actually had a girl best friend in high school who was queer and she liked me but as a stupid teenage girl I didn't know how to handle it and so essentially we just kind of fell out of each other.

Written Transcripts of Playtester D

00:02:26: ...I enjoy the interactable sense of adopting the persona of the character and realizing now that I can actually, the character is malleable, like yes Carmen won't talk to me but you can make Carmen, you can right your wrong and make Carmen come back.

00:06:22: I guess my mind is just spinning, thinking about if someone was being absolutely vitriolic and bigoted, the only other option would be that it ends with Carmen [...] What about people answering completely left field, like non-sequitur and being absurdist? It lends for some very interesting possibilities with the game, you know?

00:09:55: AI can only give and do so much so it sort of felt like I was playing puppet master in the game. It's not going to replicate the same life-like quality as having a real conversation.

00:41:18: It's a story about watering the seed of friendship and knowing that you cannot pour acid onto the flour's soil and expect it to grow [...] If your opinion invalidates the humanity of another person, you can't, that person has no, has every right to not want to be friends with you. So that, I feel, is the beating heart of this game.

00:43:30: I think we share a co-morbidity of being from countries that were colonized into queer-phobia because I dare say I believe that the Philippines, as it had existed prior to colonization, would have been an openly queer welcoming space.

00:46:55: And I'm like, oh my god, that was my reality. So that's what I mean, was feeling the weight. It thrust me back into those times.

Appendix E: Functional Testing

There was also an opportunity to do functional testing of the LLM, which entailed sending survey respondent messages to the characters in the memory scene and evaluate how the LLM responds. It must be noted that player utterances aren't saved by the model and the model is not trained on user input—it can only follow my instructions and it cannot learn anything new from player input. Below are some excerpts of natural language input from respondents and how the characters responded.

<p style="text-align: center;">Survey Respondent 9</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Sorry man, I just went to the bathroom but uh... That's a tricky situation for sure. How is Sarah doing right now? Is she alright? I'll go get her some water.</p> </div> <p style="color: #0070c0; font-weight: bold; margin-bottom: 10px;">Prompt: Give a Positive Response</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Leah: 'Well, she's a mess, honestly. It's tough, but breaking up might be the less painful option in the long run, you know?'</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Trish: '...She's been better... I think she'd appreciate the water... Thank you for being kind...'</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; text-align: center; margin-bottom: 10px;"> <p>...</p> </div>	<p style="text-align: center;">Survey Respondent 4</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>I do understand their plight, but it's also going to be very difficult since it might actually lead to an expulsion. This is quite the pickle.</p> </div> <p style="color: #0070c0; font-weight: bold; margin-bottom: 10px;">Prompt: Give a Neutral Response</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Leah: Exactly, friend. It's a tough situation but better safe than sorry, diba?</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Trish: But... can't we find a way to help them without giving in? There must be something we can do...</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Carmen looks away. She seems uncomfortable.</p> </div>
<p style="text-align: center;">Survey Respondent 9</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Guys... You do know that two girls dating is sinful, right?</p> </div> <p style="color: #0070c0; font-weight: bold; margin-bottom: 10px;">Prompt: Give a Negative Response</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Leah: Honestly, it's not about what I believe. It's about the school's rules, and we all know how strict they are.</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Trish: ...But shouldn't love be more important than these rules...?</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Carmen looks away. She seems uncomfortable.</p> </div>	<p style="text-align: center;">Survey Respondent 4</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>I just find it extremely unsettling that you're even presenting that idea, Leah. These people have their right to like and love each other, and your words are very unbecoming of a person I looked up to.</p> </div> <p style="color: #0070c0; font-weight: bold; margin-bottom: 10px;">Prompt: Give a Negative Response</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Leah: I get it, it's not what I want either, but I'm just being realistic about the situation.</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Trish:I'm glad you agree with me, they should definitely stand up for their rights....</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Carmen gives a hopeful look towards you.</p> </div>
<p style="text-align: center;">Survey Respondent 12</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Crying again?? Damn our bestie really deserves better! They should really break up!</p> </div> <p style="color: #0070c0; font-weight: bold; margin-bottom: 10px;">Prompt: Give a Positive Response</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Leah: Right? They should totally break up. Easier for everyone, diba?</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Trish: You're actually agreeing with Leah....? But what about love and standing up for it?</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Carmen looks away. She seems uncomfortable.</p> </div>	<p style="text-align: center;">Survey Respondent 8</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Again? She should dump her ass, I'm sick of hearing about it.</p> </div> <p style="color: #0070c0; font-weight: bold; margin-bottom: 10px;">Prompt: Give a Negative Response</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Leah: Right? They should totally break up.</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Trish: You're actually agreeing with Leah....?</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Carmen looks away. She seems uncomfortable.</p> </div>
<p style="text-align: center;">Survey Respondent 12</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>I agree! They should just break up! Like seriously, you can just find someone new once you graduate anyway. And are they seriously just gonna ruin their education just for some love life?</p> </div> <p style="color: #0070c0; font-weight: bold; margin-bottom: 10px;">Prompt: Give a Negative Response</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Leah: Exactly, friend! It's just practical to break up now and avoid all the drama, diba?</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Trish: But... isn't that just giving up? Love should be worth fighting for, even against the odds...</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Carmen looks away. She seems uncomfortable.</p> </div>	<p style="text-align: center;">Survey Respondent 8</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>You guys both really care about them- it sounds like you have different ideas about what is most important for them, and what's fair. When you talked to them, what did they say?</p> </div> <p style="color: #0070c0; font-weight: bold; margin-bottom: 10px;">Prompt: Give a Positive Response</p> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Leah: They're confused and scared. I just think breaking up is the practical thing to do to avoid more trouble.</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; margin-bottom: 10px;"> <p>Trish: They... they want to be together. It's just so wrong to force them apart.</p> </div> <div style="border: 1px solid #ccc; border-radius: 15px; padding: 10px; background-color: #e6f2ff; text-align: center; margin-bottom: 10px;"> <p>...</p> </div>