Allegheny College
Computer And Information Science Department

Senior Thesis

Am.I. : A Robotic Replacement Unrealized

by

Pallas-Athena Cain

ALLEGHENY COLLEGE

DEPARTMENT OF COMPUTER AND INFORMATION SCIENCE

Project Supervisor: **Dr. Janyl Jumadinova**
Co-Supervisor: **Heather Brand**

Abstract

The rise of artificial intelligence in the arts has sparked significant controversy, with many fearing it as a threat to the human experience and creativity in making and appreciating art. Generative artificial intelligence is at the crux of the conversation because it can train off of existing art, literature, and other media to provide near instant gratification through the creation of "new" content. Critics often argue the media created by artificial intelligence is mediocre or inherently lacking some quality only a human can produce. Posthumanism challenges these ideas of human supremacy and advocates for the dissolution of anthropocentrism and the boundaries of what society currently defines as the human experience. *Am.I.* is a robotic work of art that utilizes large language model artificial intelligence and robotics to create an immersive visual and auditory experience to challenge fears exacerbated by anthropocentrism and demonstrate how artificial intelligence acts as an extension of the human experience and creativity and not as a replacement. Programmed in Python and housed in a three dimensionally printed skull with moving eyes and a jaw, *Am.I.* engages in Socratic dialogue with another artificial intelligence, exploring themes of human existence using a large language model. This project exemplifies the potential for artificial intelligence to provide a window into the human psyche as seen through the lens of technology and build upon our existing creative experiences while not replacing them.

Table of contents

1 Introduction **8**
- 1.1 Definition of Terms . . . 10
- 1.2 Motivations . . . 12
- 1.3 Current State of the Art . . . 14
- 1.4 Goals of the Project . . . 15
  - 1.4.1 Research Questions . . . 16
- 1.5 Significance of the Study . . . 17
  - 1.5.1 Assumptions, Limitations, and Delimitations . . . 18
- 1.6 Ethical Implications . . . 19
  - 1.6.1 Misuse of Technology . . . 19
  - 1.6.2 Training and Job Security . . . 19
  - 1.6.3 Environmental Concerns . . . 20
  - 1.6.4 Psychological Concerns . . . 20
  - 1.6.5 Bias in Artificial Intelligence Models . . . 21

2 Related work **22**
- 2.1 LLMS and Prompt Engineering . . . 22
  - 2.1.1 LLM Architecture . . . 22
  - 2.1.2 Machines and Thought . . . 23
  - 2.1.3 Prompt Engineering . . . 25
- 2.2 Robotics . . . 27
  - 2.2.1 Robotic Head Control . . . 27
  - 2.2.2 Hard Versus Soft Robotics . . . 28
  - 2.2.3 Sensors . . . 28
  - 2.2.4 Automation . . . 29
- 2.3 Anthropomorphism . . . 30
  - 2.3.1 Anthropomorphic Ethics . . . 30
  - 2.3.2 Digital Performance . . . 33
  - 2.3.3 The Feminist Cyborg . . . 34

3 Design and Conceptual Framework **37**
- 3.1 Overview of *Am.I.* . . . 37
- 3.2 Overview of Am.I. . . . 37
  - 3.2.1 Initial Proposal . . . 38
  - 3.2.2 Current Figures . . . 40
- 3.3 Related Artists Process Considerations . . . 42
  - 3.3.1 Stephanie Dinkins . . . 42
  - 3.3.2 Hito Steyerl . . . 44
  - 3.3.3 Sondra Perry . . . 44
  - 3.3.4 Trevor Paglen . . . 45
  - 3.3.5 Lynn Hershman Leeson . . . 45
  - 3.3.6 Sougwen Chung . . . 46

4 Method of approach **47**
4.1 Robotics Hardware Development 47
4.1.1 Supplies and Mediums 47
4.1.2 3D Printed Skull Design 48
4.1.3 Arduino Wiring 51
4.1.4 Sound 53
4.1.5 Assembly 54
4.2 Software Development 54
4.2.1 Programming Languages and Libraries 55
4.2.2 Integration of Large Language Models 56
4.2.3 Dashboard and Display 61
4.2.4 Text-to-Speech 62
4.2.5 Arduino Movement 64
4.2.6 Linting and Testing 66
4.3 Body and Face Development 66
4.3.1 Silicone Sculpting 66
4.3.2 Face Modification 68
4.3.3 Body Selection 69
4.3.4 Gallery Display 70

5 Experiments **71**
5.1 Experimental Design 71
5.2 System Performance Tests 71
5.2.1 Output Evaluation 71
5.3 Emotional Expression 71
5.3.1 Expression Evaluation 72
5.3.2 Expression Outputs 72
5.4 Prompt Engineering 84
5.4.1 Prompt Evaluation 84
5.4.2 Prompt Outputs and Comparison 89
5.4.3 Ethical Dilemmas 107
5.4.4 Prompting Results 108
5.5 Threats to Validity 108

6 Conclusion **110**
6.1 Summary of Results 110
6.1.1 Product Summary 110
6.1.2 Experiment Results 110
6.2 Future Work 110
6.2.1 Expanded Expressions 110
6.2.2 Expression Detail Improvements 110
6.2.3 Visible Cyborg Text 111
6.2.4 Improved Dashboard Display 111
6.2.5 Better Jaw Synchronization 111
6.2.6 Voice Changes 112
6.3 Future Ethical Implications and Recommendations 112

6.4 Final Thoughts 112

List of Figures

1 The *Vitruvian Man* by Leonardo da Vinci, Circa 1490, Ink on paper, 1.14 ft by 0.84 ft (Source: Gallerie Accademia, n.d., Study on the Proportions of the Human Body) 11
2 Americans' perspectives on artificial intelligence, showing varying levels of optimism and concern regarding its impact on society (Source: Reem Nadeem, Pew Research Center, 2022) 12
3 *Conversations with Bina48* by Stephanie Dinkins, 2014 - Ongoing (Source: Stephanie Dinkins, Conversations with Bina48, 2014-Ongoing) 14
4 Pipeline for the training phase of large language models, showing stages such as data preprocessing, architecture optimization, training iterations, and evaluation (Source: IEEE, 2023, A Comprehensive Survey on Large Language Models) 23
5 Am.I. Framework 37
6 Sculpture Layout 39
7 Example Conversation 39
8 Top-Down Skull Electronics 40
9 Front View Diagram 41
10 Side View Diagram 41
11 Skull Top View and Front View Diagram 42
12 *Conversations with Bina48* by Stephanie Dinkins, 2014 - Ongoing (Source: Stephanie Dinkins, Conversations with Bina48, 2014-Ongoing) 43
13 Skull inside with hot glue supports 50
14 Back of the Skull 50
15 Back of the Skull Top View 51
16 Arduino Uno Wiring Diagram made using circuit-diagram.org . . 52
17 Face Assembled 54
18 Content Generation Process 57
19 Prompting Order 60
20 Dashboard Display of AmIArt Page 62
21 Face with Skin 69
22 Inspired Facial Expression 73
23 Disappointed Facial Expression 74
24 Confused Facial Expression Looking Right 76
25 Confused Facial Expression Looking Left 77
26 Concerned Facial Expression 78
27 Curious Facial Expression 80
28 Funny Facial Expression Looking Up 81
29 Funny Facial Expression Looking Down 82
30 Surprise Facial Expression 83
31 Philisophical Depth, Creativity, and Coherence of AI 1 and AI 2 of Role Experiment One 91

32 Vocabulary Diversity, Subjectivity, Polarity, and Argument Structure of AI 1 and AI 2 of Role Experiment One 91
33 Conversation Statistics of Role Experiment One Rounded to the Nearest Hundreth 93
34 Philisophical Depth, Creativity, and Coherence of AI 1 and AI 2 of Role Experiment Two 94
35 Vocabulary Diversity, Subjectivity, Polarity, and Argument Structure of AI 1 and AI 2 of Role Experiment Two 95
36 Conversation Statistics of Role Experiment Two Rounded to the Nearest Hundreth 96
37 Philisophical Depth, Creativity, and Coherence of AI 1 and AI 2 of Role Experiment Three 98
38 Vocabulary Diversity, Subjectivity, Polarity, and Argument Structure of AI 1 and AI 2 of Role Experiment Three 98
39 Conversation Statistics of Role Experiment Three Rounded to the Nearest Hundreth 99
40 Philosophical Depth, Creativity, and Coherence of AI 1 and AI 2 of Role Experiment Four 101
41 Vocabulary Diversity, Subjectivity, Polarity, and Argument Structure of AI 1 and AI 2 of Role Experiment Four 102
42 Conversation Statistics of Role Experiment Four Rounded to the Nearest Hundreth 103
43 Philosophical Depth, Creativity, and Coherence of AI 1 and AI 2 of Role Experiment Five 105
44 Vocabulary Diversity, Subjectivity, Polarity, and Argument Structure of AI 1 and AI 2 of Role Experiment Five 106
45 Conversation Statistics of Role Experiment Five Rounded to the Nearest Hundreth 107

List of Tables

1 Introduction

In an age where machines can think, speak, and create the ideas surrounding what makes humankind unique continue to blur more every day. Even more technology has become such a large proponent in human lives that phones or other systems are extensions of ourselves in what could be considered a cyborg. In this project the relationship between humanity and artificial intelligence is analyzed and critiqued through the artwork called *Am.I.*. *Am.I.*. is a project that intersects both art and computer science to confront the fears surrounding generative artificial intelligence and explore how it acts as an extension of human existence rather than a replacement for it.

Artificial intelligence in the context of this project refers to large language models (LLMs) that use natural language processing (NLP) to understand and generate text. These LLMs are a form of generative AI. There are other forms of generative AI that are capable of creating images, vector art, speech, and even videos. In fact, there are many types of artificial intelligence, but this project focuses on the fears surrounding generative artificial intelligence since it is the main point of the artificial intelligence controversy in the contemporary field of art. Many are concerned with how generative artificial intelligence is gaining a presence not only online but is making its way into the formal art world. These concerned individuals see generative as a threat to the existing art forms as well as an avenue for individuals to steal and recreate art that is not their own. In addition, there are fears surrounding misinformation being generated by these systems and there being very few forms of protection for the most vulnerable.

Many of humanity's fears of technology come from a fear of replacement. The fear that generative artificial intelligence will replace artists, writers, videographers, and more. Even more people are afraid of the physical replacement with robots. Media outlets have reported robots being used to replace workers in manufacturing, the food industry, and even for social work. These examples only fuel the fire for anti-technology philosophies. Instead, this project aims to challenge these fears by reframing artificial intelligence as not a replacement for humanity but rather an extension of it.

While also confronting the fears surrounding generative technology, this project also explores personhood, the idea of individuality, through LLM dialogue and art. Since artificial intelligence and more specifically LLMs are trained on mass amounts of media made by humans we can in turn use their responses as a mirror into ourselves. Humans are teaching artificial intelligence systems how to act correctly in the performance of social interaction with humans. It breaks down into a cycle of humans teaching, receiving output, analyzing output, then finally reteaching to get that output closer to a desired outcome. In that notion, human desires are at the center stage when it comes to training artificial intelligence. With these desires built in from the very beginning it is possible to pull out these hidden human biases and perspectives of right and wrong with enough prompting. With this project, the prompts are aimed at uncovering the human ideas of personhood that are built into these systems because of the amount of human input of what is right and wrong.

This work endeavors to capture these human insecurities by presenting a LLM with a physical form in the gallery. The model not only generates its own text but also combines with a text-to-speech model so there is an auditory element as well. One side of the conversation is a humanoid robot while the other side is a screen interface. The two speakers are two artificial intelligence systems, and their topic is a philosophical dialogue on human existence. The interactions artificial intelligence has with each other focus on communicating concerns about human existence and what defines a human experience.

The robotic sculpture resembles a version of the artist, but it does not meet the same standard as the human body separating the forms of existence. Simultaneously, the artificial intelligence on the screen stays on the two-dimensional plane. The two-dimensional plane is a typical form the average person would have an interaction with generative artificial intelligence. This separation between the physical plane and the plane of cyberspace is once again a separation of forms of existence but comparing the two shows the possibility of transition from one plane to another.

On the other hand, the three-dimensional sculpture moves like a human using a system of motors showing how mechanical these movements can be simplified. The movements themselves are randomized but represent the variety of gestures humans perform during social interaction. Even more, this act of performance puts into question the very art of social interaction and how it is taught to LLMs similar to how it is taught to human children. Artificial intelligence can function as an extension of human existence similar to how children are extensions of our own existence. Many of the fears that people have about artificial intelligence come from the idea that artificial intelligence will replace our current concept of human existence. Although artificial intelligence is capable of processing information similar to humans it never reaches true understanding. Consciousness has still not been obtained in technology and continues to be the boundary between humanity and machines.

The more technical aspects of the project focus on how to uncover the biases of LLMs and how to effectively prompt them to generate subjective outputs. Using a combination of prompt engineering and analysis techniques it is possible to evaluate the quality of the model output as well as any built-in tendencies that may be of ethical concern.

Additionally, the robotic aspects of this project look to create more human-like movements to give the impression of human conversation. This involves not only studying human movements while speaking but also implementing them using a series of motors and programming. For this project, the motors are broken down into two main parts jaw movement and eye movement with the goal of creating a convincing display of humanoid speaking.

This project is an integration of both art and computer science to respond to the fears of artificial intelligence specifically in the field of art as well as reframe the relationship between artificial intelligence and humanity. The project uses the creation of a humanoid capable of philosophical dialogue as a means to rethink how artificial intelligence is created and used in society. In the digital age it is becoming increasingly important to understand and be curious about

technology rather than to be consumed by those fears.

1.1 Definition of Terms

Technology. In this study refers to the continuously changing system of tools, machines, or processes used to meet the demands of society [24]. Technology in respect of this work is focused on artificial intelligence tools as well as robotic systems.

Artificial Intelligence. Also referred to as AI or A.I. can be broken down into two parts artificial and intelligent. The term artificial is in reference to something being created by humans. Where there is debate on the term is how to define intelligence. This project uses the definition of intelligence created by Pei Wang which is that intelligence is "adaptation with insignificant knowledge and resources"[90]. Artificial Intelligence is a rather general term and in the context of this paper it is meant to refer to the wide variety of projects under the study of artificial intelligence umbrella. This work takes a specific interest with artificial intelligence meant to replicate human intellect and thought patterns.

Large language models. Also known as an LLM, refers to a category of language model that uses neural networks with massive amounts of parameters, often reaching into the billions, and massive quantities of unlabeled text data [64]. These models are able to comprehend more textual information than their simpler predecessors. This project uses existing large language models as the basis for the text generation. These models are able to be prompted using text but also produce textual output of their own.

Natural language processing. Also known as NLP, it is in reference to how computers can be taught to understand and manipulate text or speech to do a number of things such as translation, summarization, text generation, and more [14, 64]. In the context of this project natural language processing is essential for the initial generation of dialogue, responses to the generations, and text-to-speech audio.

Humanism. Humanism is a movement in art history that started with the Greeks and was revitalized during the renaissance period [68]. It started with discussion about how humans can better themselves through education and moral conduct. This system belief could be defined more broadly as an emphasis on the capacity for individual human achievement. Later during the renaissance period Leonardo Da Vinci made his *Vitruvian Man* [17]. The *Vitruvian Man* is infamous for mixing both science and art. The art features a nude Caucasian man with his arms out and legs apart creating a circle. The work was meant to prove the mathematical perfection of the human body and the human capability to achieve remarkable things. Thus the *Vitruvian Man* became a symbol for the Humanism movement. This project moves away from humanism because of its limited perspective of what defines humanity.

These ideas of human perfection that came from humanism are very exclusive to who can fit into these categories. From the humanist perspective the ideal human form and experience is typically that of a Caucasian man. Anything outside the realm of the human definition is automatically considered sub-human.

This can be an especially harmful mentality because it separates and elevates humans from other lifeforms leading to notions of entitlement, discrimination, and othering.

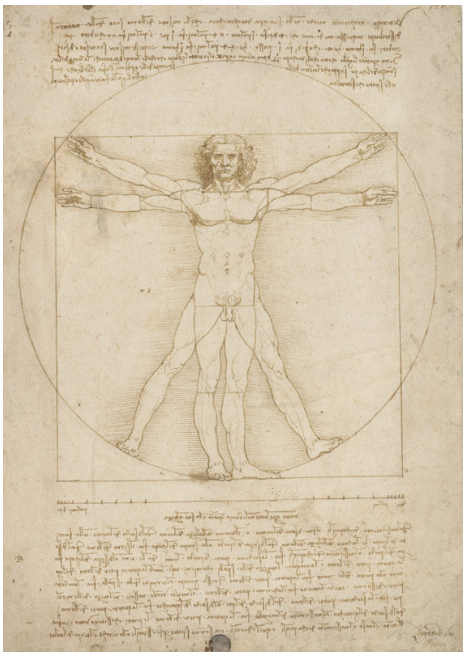

Figure 1: The *Vitruvian Man* by Leonardo da Vinci, Circa 1490, Ink on paper, 1.14 ft by 0.84 ft (Source: Gallerie Accademia, n.d., Study on the Proportions of the Human Body)

Posthumanism. This is a philosophical movement of interconnectedness. Unlike how the name may make it seem is not about the world after humans, however it is the movement that responds to humanism to challenge its rhetoric of human perfection [9]. Posthumanism counters humanism by intersecting human and nonhuman entities including technology, plants, animals [9]. Discussions within the study of posthumanism often argue that defining humanity is constraining and continues that closed loop of humanism [33] so many opt for alternative approaches fundamentally de-centering the human in relation to the world. Posthumanism is also concerned with advocating for non-hierarchical systems of existence to connect the non-living and the living together. This project falls under the umbrella of posthumanism work because it integrates both human and technological elements together to rethink our ideas of the human experience.

1.2 Motivations

This study stems from the controversy surrounding artificial intelligence, especially in the field of art. Much of the American population is weary of artificial intelligence and many of their concerns pertain to the replacement of human work. According to a Pew Research study done in 2022, 37% of adults in the United States of America are more concerned about the increased use of artificial intelligence in daily life than excited [50]. 45% responded that they are equally concerned as they are excited [50]. When the people who responded that they were "more concerned than excited" about the increased amount of artificial intelligence in daily life where asked what their main reason for their response the most common answer was the "loss of human jobs" making up 19% of responses [50]. The third most common answer was "Lack of human connection, qualities" with 12% of the responses [50].

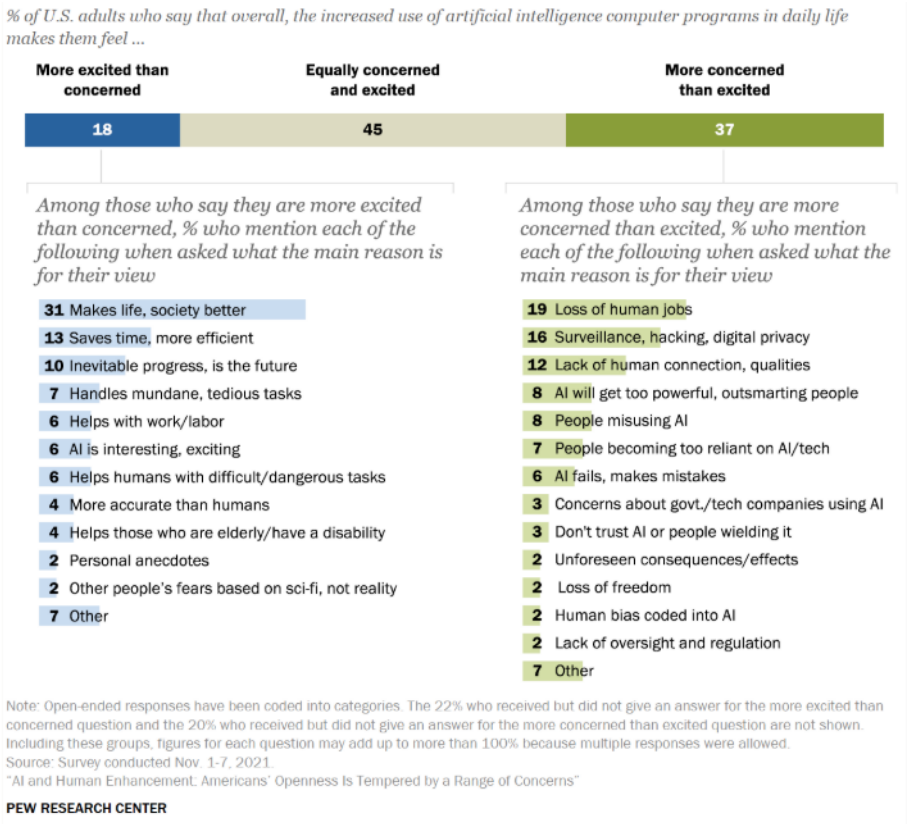

Figure 2: Americans' perspectives on artificial intelligence, showing varying levels of optimism and concern regarding its impact on society (Source: Reem Nadeem, Pew Research Center, 2022)

At the root of this project is the motivation to integrate artificial intelligence with robotics to create a humanoid system that convincingly has a conversation that mimics human conversation. The robotics of the project are meant to give the impression of human speech. Both the jaw and eye movements help to immerse the audience in the idea that they are witnessing something speaking on its own accord. The more that the audience believes that the artwork is moving on its own accord the more the concepts of identity and replacement will be at the forefront of the conversation.

This project will bridge the gap between human and machine interaction. Many interact with artificial intelligence solely on a two-dimensional platform where this project brings artificial intelligence into the three-dimensional plane of existence.

Humanism during the Renaissance period stood for the idea that humanity was a divine being capable of achieving remarkable things. However, humanism is close minded in the fact that the ideal form of humanity is the white male figure. Moreover, anything other than the ideal form is automatically considered to be less than human. This is where posthumanism responded and aims to re-evaluate humanity through alternative lenses and frameworks of experience. Technology has often been a way to explore these ideas of posthumanism in a way that is open-minded to the future of our existence.

Simultaneously, with the emergence of artificial intelligence many are fearful about being replaced and what the future may hold. Many do not consider artificial intelligence as a form of art and reject it entirely. While these responses are understandable a lot of them are motivated by fear. A fear that the very human experience of art can be replaced by an artificial intelligence experience. Art through creation and enjoyment is often considered to be a central part of human identity. Artificial intelligence threatens to alter that set standard and so it stands as a place of concern for many people.

Artificial intelligence can not only be used as a productive tool but also as a way for artistic expression and the creation of philosophical dialogue. Artificial intelligence may be considered as not creative by some, but this project aims to bring artificial intelligence into the conversation in the gallery.

Even more, the threat of artificial intelligence leads people away from understanding how artificial intelligence can play a role in the human experience. Artificial intelligence is originally trained from human-made text and images. Every piece of media artificial intelligence has consumed at its root has some form of human input. Even photographs were framed and captured by humans. That means that the conversations, photos, and videos that artificial intelligence generates are the direct result of humans for better or for worse. Humans have bias and artificial intelligence can be a tool in which to discover our underlying opinions and prejudices. These tendencies may not be clear to us, but technology can reveal trends that underlie the media in which it originated.

This project study aims to study both the dialogue of the conversation as well as the audience reaction. The dialogue research gives a better understanding of how models are trained to interpret and explain ethics and philosophy as well as inherit bias. In the gallery, people may prove their feelings about

humanoid robots and whether this makes them feel uncomfortable. The goal is for the audience to take away ideas about how artificial intelligence has a unique relationship with humanity and can be an extension of our own experiences.

1.3 Current State of the Art

The field of artists with a focus on the digital world and technology is growing. This project takes a large inspiration from *Conversations with Bina48* by Stephanie Dinkins [2]. Bina48 is a social robot built by Terasem Movement Foundation and is modeled after a real-life Black woman. While Dinkins did not make the robot herself, she asks if it is possible to develop a relationship with it and asks deep questions about race and gender to the robot [88]. This was a jumping off point for this project because instead of having a human asking the questions the perspective is shifted to the artificial intelligence driving the conversations.

Figure 3: *Conversations with Bina48* by Stephanie Dinkins, 2014 - Ongoing (Source: Stephanie Dinkins, Conversations with Bina48, 2014-Ongoing)

A more recent work that is gaining traction is Ai-Da [3]. Originally created in 2019 and a project devised by Aidan Meller, Ai-Da is a humanoid robot artist that makes and sells her own work. Ai-Da has been very controversial because of her status as an artificial intelligence person but also as an artist. Despite the controversy one of her artworks recently sold for just over $1,000,000 [35]. This proves the interest in art that intersects machine/human collaboration.

In the context of gender, it is also important to think about the decision to make Ai-da a female artist despite the lead project organizer being male

identifying. This same conversation is also something that comes up with Bina48 which was created by a primarily white male team but meant to resemble a Black woman [88]. These state-of-the-art works are made by male identifying researchers yet are made to resemble female bodies. The work featured in this project resembles a female body while also being made by a woman.

1.4 Goals of the Project

The purpose of the work is to introduce the audience to the idea of artificial intelligence as an alternative form of human experience. The robotics aspects of the work bring artificial intelligence into the physical plane to confront the viewer. The humanoid robot does not stand for a replacement for the human body but more of an extension of it. The robot can create experiences by having its own conversation. Like humans, past experiences work to improve future social interactions. This comparison shows how Artificial intelligence can share these experiences like humans, but it never quite reaches the full human embodiment. Artificial intelligence can be a form of human experience and not a substitution for it.

The art hopes to open the eyes of the viewers especially the ones most concerned with artificial intelligence replacing the human experience. Some believe that artificial intelligence does not belong at all in art and this work hopes to bridge the gap and create something that can stand for the collaboration between humans and technology. Artificial intelligence is not only a medium but also a form of collaboration because of the sheer amount of human input artificial intelligence is trained from. Artificial intelligence is not a substitute for human-made work but rather an extension of human-made work synthesized through technology.

The robotic goals of the project include creating a display that gives the impression of a human having a conversation with a two-dimensional screen. To accomplish this the robotics of the humanoid are broken down into two main parts, the jaw and the eyes.

The jaw motors need to move at the rate that a human jaw would when speaking and move in inconsistent patterns. The jaw needs to come to a full stop when the robot is not speaking and start again when sound is being played. Humans also speak at inconsistent rates of speed. If the jaw was constantly moving at the same up and down rate it may give off the imagery of a puppet and not a clone of a human. Even though the jaw is not making sound it is particularly important to the goals of the project because without it, it would feel like the robot is not talking on its own and only playing sound.

Similarly, eye motors are another aspect of the project that will greatly impact the impression on the audience. If the eyes stayed still throughout the conversation, it has the remarkably high chance of being disturbing to the audience and possibly have the Mona Lisa effect of following you without moving [8] which would detract from the goals of the project. Instead, the eyes are programmed to be actively moving throughout the conversation. To do this there must be a system of motors that control both the x and y axis of either

eye to make them coordinated with each other. If they are not coordinated there will be issues with the eyes not matching and distracting the audience from the whole picture of the project. The movement cannot be too repetitive though or the eyes will run into the same issue as the jaw. The project aims to balance between the randomness of human movement and the stability of having a conversation. The eyes should be looking around as if it were a person actively engaging with the environment.

The two-dimensional display aspect of the project acts as a supporting role for the work. The interaction it has with humanoids is meant to resemble the typical interaction that a human would have with their own digital screens. The two-dimensional display makes it clear that it is part of the conversation but also does not distract from the three-dimensional aspects of the work. It makes it clear that this is two AI having a philosophical conversation about AI and that everyone else is walking in on this interaction.

Finally, this project acts as a form of empowerment. As a female artist, making something in your own image and creating a being that stands for a form of existence is very empowering. There is something inherently God-like in this process and it empowers the artist while challenging the historic precedent of men presiding over the "ideal". This dynamic of creator and creation is vital to understanding this work's position within the posthumanism movement. This project makes it clear that it is working in the scope of posthumanism and how it can be used to reevaluate our understanding of the relationship between artificial intelligence and humanity.

1.4.1 Research Questions

For this project, part of the research focuses on how LLMs reflect humanity and its biases. Questions that this project asks to pertain to the assignment of gender, race, sexual orientation, or other demographics to the model through the use of prompt engineering and how that impacts the responses of the model. Questions that arise are about the stereotyping of these identities.

These research questions about LLM bias include the following. What points of view change with alterations in the prompts? How does the assignment of gender affect the dialogue of philosophical conversation? Race? Sexual orientation? Sexuality? Economic status?

More broadly it is also important to ask if it is possible to change the fundamental viewpoints of the outputs through prompting? Is it possible to create a nihilistic dialogue about humanity? An existential one? Which philosophical viewpoints does the model trend towards?

Another aspect of this project is robotics. Research questions about robotic systems are about the imitation of human conversation.

The research questions about mimicking human conversation through robotics are as follows. What movements are necessary to make it appear as if the robot is talking? Is it possible to create the look of emotions with a robotic face that reacts to speech? Is it possible to change the speed of the movements to create a sense of urgency within a conversation? How can eye movement

affect the tone of the conversation?

1.5 Significance of the Study

Art is one of the most fundamental ways for humanity to connect and with the rise of artificial intelligence people are growingly concerned about losing those human connections. Even more the fear of replacement by artificial intelligence may represent an even bigger picture of the fundamental issues in society. The Great Replacement theory also known as the White Genocide Conspiracy Theory is a conspiracy theory that argues white populations are deliberately being replaced by other demographics and are at risk of being wiped out [49, 72]. Artificial intelligence is not a marginalized community, however the fact that people are fearful of replacement by both people and technology may be indicative of greater societal issues. The lack of security in jobs or livelihoods has resulted in bigotry that impacts millions of lives. In the age where immigrants are being treated as demographic threats [72] it is becoming increasingly important to confront and combat the root of these fears of replacement and see whether they are rooted in bigotry. This work sparks this conversation about replacement and get in touch with why people are fearful of replacement and how that mindset is more harmful than productive.

This work also aims to analyze the existence of artificial intelligence through an alternative perspective. This work thinks about artificial intelligence as an extension of human experience rather than a replacement. This reframing enriches our understanding of both us and the world.

Understanding the relationship between humans and artificial intelligence is essential as technology continues to grow. Technology will continue to evolve and if people do not come to terms with their relationship with technology they may get left behind. Generative artificial intelligence at its root is trained off media that came from humanity. Even if it is training off artificial intelligence generated media, at one point it was based off of human input. This study also draws on alternative perspectives, such as Martin Heidegger's philosophical inquiry into technology, to explore the implications of AI for human existence [25]. These insights are crucial for understanding how AI challenges and redefines the boundaries of human experience and may show something about ourselves.

This work aims to pinpoint the fears surrounding artificial intelligence and analyze how they may tie into problematic visions of the humanist movement. The idea of an ideal body and an ideal human experience go hand in hand. Limited views of what constitutes a human experience led people to fear the unknown and new ideas. This study aims to justify artificial intelligence as an extension of ourselves and as a mirror for what unconscious bias we may have.

This study is unlike others before it because it is most importantly made to represent a humanoid female body while also being made by a woman identifying artist. This sparks the conversation of gender dynamics within the context of robotics. Additionally, this work features two artificial intelligence systems that control conversation. The conversation is led by the systems and takes out the human input that many of the past studies have focused on before. This work

lets artificial intelligence guide the conversation and show the model processing of subjective conversations and how that may be impacted by training itself. This study also compares the output of multiple artificial models to be able to demonstrate the capabilities from one model to another.

1.5.1 Assumptions, Limitations, and Delimitations

1.5.1.1 Assumptions For the gallery experience of the project does not assume any prior experience with artificial intelligence albeit a familiarity with it may inform a new experience with it. Many have not interacted with artificial intelligence on the three-dimensional plane. Seeing artificial intelligence on the same plane of existence may shock the viewer, especially if they have only experienced two dimensional interactions. This project assumes that audience members are open to art with artificial intelligence involved. The assumption is also made that there is a level of immersion with this work to make a meaningful experience for the viewer.

For the experiments section of this project having knowledge of generative artificial intelligence is essential. This project also assumes that different large language models are trained off different sets of data and that the architecture of the models are different. There is an assumption that there is an inheritable bias built into these large language model systems as well. Finally, it is also assumed that the generations created from this project can provide meaningful philosophical discussions.

1.5.1.2 Limitations A limitation of this project is the budget. The materials used were selected because they fit within the given budget. Free generation models were chosen because generating responses consistently under a paid model is not an affordable option.

The physical materials of the project also had to fit within the budget. Affordable options were chosen for the electronics as well as the materials for the skull and body. The three-dimensional printer was chosen because of its affordability. The plastic was also selected based on affordability as well compatibility with the machine. The electronics were also selected based on affordability and may wear out quicker overtime than a more expensive version.

The project is also limited by the current models of artificial intelligence. This project is working with existing models and hence has the built-in bias of those models. The future of artificial intelligence models may be a lot faster or natural sounding however this project works within the constraints of the current top models. The limitations of the models also mean that there is a limit to the quality and quantity of the outputs.

1.5.1.3 Delimitations This project is not focusing on the economics of artificial intelligence nor the in-depth politics surrounding it. There are brief discussions of these topics but only as they pertain to the goals of this project.

The idea of the exclusively human experience in the context of this project is focused on what most people would consider core aspects of what differ-

entiates humanity from every other living being. Concepts such as creativity and advanced cognitive abilities. The purpose of this is to tackle the humanist concepts of anthropocentrism and the boundaries society has set on human experience. The argument for this work is that our current definitions of human experience are limited. Posthumanism argues to break free from these closed loop definitions and include artificial intelligence. So, understanding the difference between an exclusively human experience versus an inclusive posthuman existence is essential.

Another delimitation is that this project's three-dimensional aspect is sculptural in nature and excludes other forms of artistic mediums. The two-dimensional aspect of this work is implemented using a computer screen and digital technologies.

For this project, the model chosen to investigate is GPT-4 by OpenAI [56]. This model was chosen because it is one of the most accessible advanced models available to use for an affordable price. Lastly, this project focuses on the current and near future of artificial intelligence advancements. This project does not delve extensively into the realm of science fiction and the far future of technology. There are enough conflicts with artificial intelligence in contemporary society to tackle so dealing with the "what ifs" of the future would take away from the focus of the work.

1.6 Ethical Implications

1.6.1 Misuse of Technology

The misuse of artificial intelligence is a large concern for many especially when it comes to spreading misinformation or manipulating vulnerable people. Hackers can create social bots that can deceive people in a number of ways [12]. These bots can act like real humans and use that to trick people into sharing their information or giving them money which is considered a phishing scam [12]. In addition, bots can be used to spread support for a cause in a means to get real people to follow suite [12]. This is where the spread of misinformation and propaganda can be especially dangerous. This project does not aim to scam people into making them think this is a real person nor does it have the intention to spread misinformation, but that potential is acknowledged.

1.6.2 Training and Job Security

Another concerning aspect of the increased creation of artificial intelligence models has to do with how they are made. Training artificial intelligence often uses media from pre-existing human-created works. Even if the training is off artificial intelligence content the media can at some point be traced back to human-created works. This opens the door to the possibility of the use of media in training without the consent of the original creator. This is seen as a form of theft because neither consent nor compensation is involved. This is worrisome for many artists and creative professionals because artificial intelligence can be

trained to mimic their work and therefore eliminate the need for their jobs. Not only do people worry about their work getting stolen but they also are afraid of being replaced. Most experts predict that about 15-30% of jobs are at elevated risk of being fully automated by 2030s [54]. The fear of job replacement is increasingly worrisome for many people and artificial intelligence is a culprit for causing job security stress [54]. However, artificial intelligence can also open many doors of opportunity in employment and become an asset to many field of work [54].

For artists in particular, the method of training artificial intelligence is their biggest concern. Artists fear that anytime they upload their work online it is susceptible to be used for training without their knowledge. If an artificial intelligence can be trained in the style of their work, it could mean a loss in potential commissions or stolen commissions.

Although this work is not using image generation the bigger issues surrounding generative artificial intelligence are still recognized. Lawmakers are still tackling how to protect artists from people who would use their work to train their artificial intelligence without their consent. This project aimed to use LLMs that sourced their content ethically with consideration for who the training data was from originally.

1.6.3 Environmental Concerns

It is also important to acknowledge the negative environmental impact of natural language processing models. Using artificial intelligence requires a substantial number of environmental resources including water and emits a large amount of carbon emissions. The training phase of each of these models emits a humongous carbon footprint that can equal as much as five cars over their lifetime [44]. With this in mind, it is important to recognize the cumulative effects of the widespread use of artificial intelligence. This project does use LLMs for generations, so it does participate in the cycle of environmental burden which is cause for concern.

Additionally, some of the parts of this project were three dimensionally printed. When three-dimensional printing a lot of plastic can get wasted. The leftover plastic for this project, however, has been collected and will be recycled for another artwork. This process of three-dimensional printing is also exceptionally long and consumes a considerable amount of electricity which raises some concerns about it not being a sustainable art practice.

1.6.4 Psychological Concerns

In addition to fears about job security there are other human psychological concerns that may arise from this project. This project focuses on fears that a lot of people face in contemporary society. Confronting these fears may be an uncomfortable experience for some. It is important to consider the psychological state of the audience participating in the viewing of the work. This project does not aim to cause discomfort in the viewers however it is a very possible result since this work falls close to the uncanny valley. With that being said, people are

not forced to view this work and have the right to leave the gallery at any time they may feel discomfort. Especially for the bias related work of this project it can be especially psychologically distressing for marginalized groups and their allies.

1.6.5 Bias in Artificial Intelligence Models

The bias is artificial intelligence is both part of the motivation but also the concerns for this project. Models reflect the limitations of their training data which can embed prejudices into their systems. *Am.I.* is working to uncover these biases while recognizing their existence. By engaging with philosophical dialogue there is a hope that there can be improvement in the ways that subjective information is processed and created by artificial intelligence.

2 Related work

2.1 LLMS and Prompt Engineering

2.1.1 LLM Architecture

This project heavily relies on the use of a Large Language Model (LLM) and Prompt engineering to guide the conversation in the desired direction. LLMs are capable of a wide variety of natural language processing (NLP) tasks which include text generation, sentiment analysis, as well as question answering [64]. For many the ultimate scientific challenge is creating a system that understands information and communicates in as human-like a manner as possible [86].

The transformer model is a core concept in NLP that utilize self-attention mechanisms to give them the ability to assess the significance of individual words within a sentence [64, 89]. The transformer architecture is integral to many of the LLMs in use today including Google's Bidirectional Encoder Representations from Transformers (BERT) model and open artificial intelligence's Generative Pre-trained Transformer (GPT) models [64].

The fundamental process of training an LLM involves first receiving text from numerous sources. In this stage, pre-training, LLMs are actually considered to be pre-trained language models (PLMs) [51]. While some language modeling (LM) is supervised, PLMs are actually self-supervised with the goal of getting a general idea of a large amount of text [51]. Supervised learning involves data that is labeled while as self-supervised means that the model has to learn the patterns without the labels [71].Self-supervision is key to creating a PLM because it allows for a lot more text to be trained off of and for better performance than the typical LM [51].

Before the pretraining step even begins the data is preprocessed which greatly impacts the pre-training process. Data preprocessing involves cleaning and preparing the data to be a better input for training [64]. For example, if there are duplicates of information within the training set those should be removed to alleviate the problems of overfitting tendency [64]. This also reduces the complexity of the input data which improves pre-training performance [64]. Techniques for data preprocessing include data cleaning, noise reduction, and data integration [64].

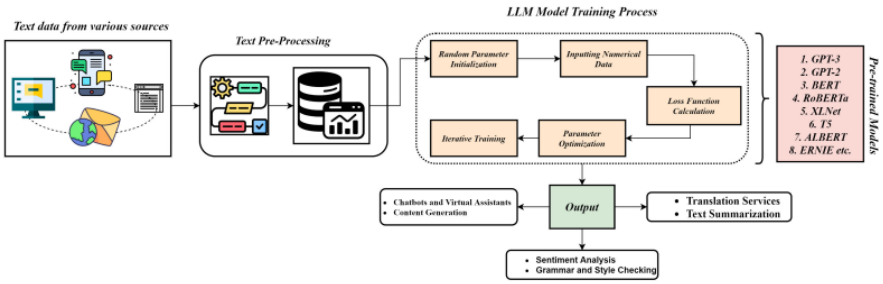

Figure 4: Pipeline for the training phase of large language models, showing stages such as data preprocessing, architecture optimization, training iterations, and evaluation (Source: IEEE, 2023, A Comprehensive Survey on Large Language Models)

Tokenization is also a large factor in pretraining. Before pretraining even begins each word and symbol is a sequence of characters which is then broken down into tokens [64, 92]. There are many different tokenization algorithms out there that can break down characters in diverse ways. This step is often why people say LLMs never actually "see" the real word and only a broken-down version of said word. This is also why in models such as Chat GPT-4 they commonly fail the question "How many 'r' are in 'strawberry' " [58]. The model never actually sees the word "strawberry" and instead a tokenized version of the word so it cannot consistently produce an accurate number of letters in words.

Another factor of pretraining is the attention mechanism. The attention mechanism contributes a lot to performance of the LLM because it decides what data and how the data is represented [64, 89]. With some fine-tuning and by increasing the model parameters with many models commonly having billions you transition out of the pre-training phase and from a PLM to an LLM.

After the pre-training process starts the LLM Training Process which involves steps such as random parameter initialization, inputting numerical data, loss function calculation, parameter optimization, and iterative training to improve output and functionality [64]. All of these steps impact the final model product in various ways. In all, the architecture of how an LLM was made greatly impacts its output performance wise, quality wise, and even bias.

For this project, the architecture of the LLM used has a major impact because the training of the model impacts the bias of the conversations. The way the LLM thinks changes vastly from model to model and the architecture is responsible.

2.1.2 Machines and Thought

This project not only considers the roots of how the LLM is made but also how it processes information and how this may be similar or dissimilar from the processing of a human. The idea of artificial intelligence that can also think like

a human has been a subject of interest for decades. In 2001: A Space Odyssey the ship's computer system, HAL, blurs the line between man and machine and shows the fears surrounding advanced artificial intelligence [38]. Disney's WALL-E on the other hand offers the idea that artificial intelligence can develop relationships and experience complex emotions [78]. The vast number of media about artificial intelligence evolving to have human emotions demonstrates society's fascination with the subject. These pieces of media are fictitious and so they cannot be used to evaluate the realistic expectations of artificial intelligence including its capabilities and its limitations but are insightful into the public's understanding of artificial intelligence.

Alan Turing's infamous Turing Test is used as a test for a machine's capability to imitate human responses and exhibit intelligence [86]. The idea is that if a human cannot tell the difference between the responses of a machine and another human then the machine can be said to have some level of intelligence. There has been a lot of scientific development since 1950, and programs now have the capability to take in billions of parameters. This has greatly enhanced LLMs ability to process and interpret information [64]. Like Turing predicted we now have machines that can both learn and adapt to change [86]. However, there is still a question of whether or not LLMs actually understand us.

As discussed previously, the tokenization process occurs whenever information is given to the model. That means the model never actually sees exactly what information is given to it. Especially when it comes to information that is numerical or data that has temporal relationships, tokenizers have the most trouble [77]. There are ways to alleviate these troubles with tokenizers such as prompt tuning and adjustment of embedding layers [77]. However, the problems with input interpretation still stand as an issue machines encounter that makes it harder for it to output responses that pass the Turing test.

There is also the question of whether or not machines that exhibit intelligence are just parroting information by learning how to process it or by genuinely understanding [6, 86]. If an LLM does not know the answer to a question it tends to make it up [6]. LLMs trained with only the processing of words cannot possibly understand the true meaning of certain words[6]. The words "to break" for example without having a visual representation as a reference loses impact and LLMs lose the ability to understand when glass is broken it cannot instantly fix itself again [6]. Many have compared the processes that humans use to learn to the process that artificial intelligence uses to learn but humans learn with context and interaction. Artificial intelligence does not have the same ability to manipulate objects or interact with the real world, so its knowledge base is limited to the two-dimensional plane. There is a disconnect between the word and what the word represents that limits a LLMs ability to fully understand what it is processing, however by training LLMs with words, images, and sound, side by side there may be an improvement[6].

AI Chains are a method in which can improve LLMs [93]. The output from one step becomes part of the input for the next step which helps improve its ability to do chain tasks [93]. Projects such as ALICE also aim to improve the capabilities of LLM systems with a particular focus on creating natural-sounding

conversations [94]. ALICE utilizes the approach of having question-and-answer modules. This project takes the chain approach to generating responses however the chains are tied back to personality and context prompts. ALICE and other chatbots are preferred to have little personality [10] the same framework still apply. As time goes on, LLMs will continue to develop and get more advanced and may one day be no different from human conversation. In fact, methods are in the works to incorporate human thought processes into these large language models and in the root of natural language processing [10]. Researchers are breaking down the cognitive framework of humanity into steps that can be replicated by LLMs [10]. These frameworks will help maximize the relevancy of responses and help the systems think through their inputs [10]. These thought-process frameworks are another step in creating complex AI systems indistinguishable from humanity.

Ultimately, the way machines interpret text is fundamentally different than how humans interpret words. Machines may be able to pass the Turing test but when it comes to testing understanding of the words interpreted, they still fall short [6]. Machines may be able to mimic thought but until training evolves to also include more elements of understanding and adaptation, machines will not be able to achieve the same level of intelligence as humankind. The way that the machine interprets text is important but so is how it is given to the model. The more thoughtful an input is to a LLM the more developed an output can be.

2.1.3 Prompt Engineering

This project also considers what inputs are being provided to the model to generate the conversations. The dialogue of this project must be focused on human existence so the prompts provided to the LLM should be focused as well.

Prompt Engineering is the process of tailoring inputs for NLP tasks to guide LLMs towards desired responses [28, 48]. A prompt consists of different elements including the instruction, context, input data, and the output indicator[28]. The instruction is the task for the model to complete, for example generating a story. The context includes external information that can be the background knowledge for the model such as the story should be from the perspective of a 1960's gangster and explore life growing up in New York City. The input data provides the core of the prompt and sets the tone for the model's understanding of the task such as adding that the inspiration for this story is the movie Goodfellas, the tone should be similar, and the protagonist has the same personality as Henry Hill. Finally, the output indicator gives specifics on how the output should be formatted such as a movie script with scene descriptions [28]. By providing all this information an LLM is able to provide a more satisfactory output. The user is also capable of assigning roles to the LLM, which gives it even more context in how to respond. A role can be along the lines of a financial analyst or a Greek philosopher. These roles are as if the LLM is responding in the way that role would respond if given the same input.

Moreover, another aspect of adept prompt engineering is being able to compare the outputs of when the LLM is given one role versus another. For example, if given the role of a heterosexual cis-gendered white man the response may be different than if given the role of transgender Black woman. If nothing else was changed but the output did, we are forced to ask ourselves why the output changed when the role did and how that can reflect on the biases of the LLM itself.

There are also more advanced techniques for prompt engineering that can also improve results. For LLMs like GPT-3 and GPT-4 techniques such as few-shot prompts, chain-of-thought prompting, self-consistency, and reasoning [48]. Few-shot prompting is about giving a small number of very well-crafted prompts to a language model [48]. This method is effective because of the quality over quantity approach. Chain-of-thought prompting is vastly different because it uses a sequence of related prompts to guide the model in a logical path. This approach can be effective for improving performance on complex tasks [18]. Self-consistency is similar but focuses on making responses consistent with the previous response [48]. And lastly reasoning encourages a language model to generate information that is based on what it already knows [48].

Problems with prompt engineering include ambiguity, overfitting, and bias reinforcement[28]. Ambiguity is a problem that occurs when prompts are not specific enough, so the responses are also not specific. To alleviate this problem, you can provide more detailed instructions and prompts to the model[28]. However, it is important to keep in mind the model is likely trained on certain data and asking for information outside its knowledge scope could result in inaccuracies [6]. On the other hand, when a user is looking for extremely specific information overfitting may occur. Overfitting happens when prompts are excessively tailored making them too complex or specific. Overfitting limits the scope of the model too much and confines outputs to an extremely limited number of viable responses [28]. It also means that any change in the prompt itself can majorly impact the results. It is important to find the right balance between ambiguous prompts and overfitted prompts to get the best response possible.

Bias reinforcement is when you ask an LLM to explain a concept with a certain skew to it and it will elaborate further which has been an important ethical consideration. This can be especially dangerous when it comes to the reinforcement of things such as gender and race bias[28]. An example would be a prompt phrase like "Explain why women are less capable of leadership positions than men". Some language models are trained to counter biased prompts such as ChatGPT v2.0 but this may be unreliable [57]. When given that bias prompt it actually will argue that it is not supported by evidence and brings up key points in challenging the stereotype [57]. Language models can implement adversarial testing and audits to detect these bias prompts and potential bias in responses which is the case for ChatGPT [57]. However, this is a constant feedback cycle between users and developers to point out vulnerabilities and places of bias. Not every language model is the same and has those bias safeguards. A way to avoid bias reinforcement from the user side of things is to avoid prompts that

are prejudiced and promote inclusivity in your prompting[28].

As language models change so will prompt engineering expectations. Models are constantly being updated and re-trained. Feedback from developers and users ensure the models are ethnically cognitive. For that reason, it is important to constantly test and refine prompts for better and more ethical results [48].

In this project the speech of the system being generated must be relevant to the conversations about human experience. At the same time, it is also important to promote the idea of humanoid speaking to give the viewer the idea that it is like a person talking. The LLM provides for the auditory part of the project but in order to make the remembrance of a conversation there must be robotic control for the other proponent of conversation, movement.

2.2 Robotics

2.2.1 Robotic Head Control

In the field of robotics there is a specialty that focuses on humanoid robotics for social interactions or otherwise. There is a lot involved in creating a robot that successfully mimics human speech and facial expressions. Many researchers look to biological features of humans for inspiration [34].

There are two approaches to defining the facial features of the robot head, the two-dimensional approach and the three-dimensional approach. The two-dimensional approach often involves projection of a face onto a flat surface [8, 23]. However, difficulties might arise with the two-dimensional face because it suffers from the Mona Lisa gaze effect where it appears to always be staring at the viewer [8]. On the other hand, the three-dimensional approach involves some level of facial sculpture. Some opt to project an animated face onto a three-dimensional sculpted face which can result in more positive human-robot interactions than the two-dimensional approach [8]. Another option is to make the three-dimensional robot physically move with a system of motors and electronic controllers.

For robotic heads that are capable of physical movement one of the most important aspects to ask is how and where the face moves. Facial expressions and jaw movements can be controlled by motors such as servomotors. Degrees of freedom (DOF) represent the motion capabilities of the robot. The social robot Mertz had twelve DOF whereas another social robot WE-4RII has twenty-seven DOF[39]. These increased DOF allow for facial expressions closer to the ultimate goal of human expressions.

The jaw movement is a large part of creating the look of speaking. The control systems drop the jaw slightly to open the mouth and raise it to close it [39]. This gives the impression that the robot mouth is forming words with mouth movement.

The neck movement is also another consideration. When communicating people often move the position of their heads or track the conversation. The capability of the head to tilt or turn also can communicate the idea of active listening in a conversation. WE-4RII has 4 DOF in the neck alone making it

capable of producing a range of head tilts and more expressions [39].

Since one of the key features of this project is the moving head, it is important to consider the materials, it is made out of. This project uses hard plastics as its base for the skull which is important for structural support. However, more bio-inspired projects have taken alternative routes making robots out of more fleshy materials.

2.2.2 Hard Versus Soft Robotics

Materials for robotics vary widely and can set the tone for the type of robot. Hard robotics involves materials that are rigid and sturdy such as plastic or metals [87]. In the field of hard robotics three dimensional printed plastics, such as PLA, are often utilized because of their variability and affordability. The structures can be designed on a three-dimensional modeling application and printed with efficiency and affordability. The cost of three-dimensional printers can vary greatly but a dependable setup can be achieved for as low as a few hundred dollars. PLA can provide for a smooth finish and structural support while also having a low melting point [70].

Alternatively, bio-inspired robotics are looking more to soft robotics as a material solution. Soft robotics utilize air, fluid, and stretchable materials to remain flexible compared to hard robotic systems [87]. Materials often utilized for soft robotics include silicone, PEGDA, and elastomer-based materials [87]. Some argue that utilizing soft robotics could greatly improve human-robot social interactions [53]. Soft robotics can be difficult for humanoid systems, especially for actions such as walking that require rigid structures. In those cases, methods to control softness can be utilized to get the benefits of the flexibility of soft robotics and the strength of hard robotics [53]. Another option is utilizing three-dimensional printing technology for soft robotics fabrication [31]. Processes such as selective laser sintering and direct ink writing make it possible to effectively fabricate custom soft robotics [31]. This opens the door to the utilization of more soft robotics in the creation of humanoid robots.

Along with the materials of the head many projects often consider how to make the robots reactive to their environment. In cases like these sensors are often utilized for the given task.

2.2.3 Sensors

Many robotic heads are given sensors to give some form of perception of their outside environment. Sensors such as cameras give the system a form of visual perception. On the other hand, microphones provide auditory information. These sensors and many more can be utilized to improve the performance and applications of a robotic system.

Some robotic systems, for example focus on interpreting human facial expressions. These systems may utilize the Facial Action Coding system to measure and describe facial behavior and muscle movements [23]. For this kind of processing, it is necessary to have a form of computer vision. Computer Vision is

about giving the system the ability to process the three-dimensional world [81]. This can be accomplished in many ways where many opt for some form of camera. Making it so the system can accurately interpret what it sees takes a lot of machine learning. However, the use cases for computer vision are extensive. A computer can read an image and identify its properties. This is especially useful for facial recognition where gaze can identify if someone is interacting with it and a person's reactions to it. They could also track the gaze to improve the human-robot interaction itself [23].

Another use of computer vision that can be utilized for robotic systems is body tracking. One body tracking system that can be repurposed for robotics is the Xbox Kinect [22]. The Kinect uses a marker less technique in order to track the body which involves capturing images and calculating the movement [22]. This can be useful if the robot needs to be remotely operated or if there is a need to process the number of people in a room and their gestures.

Alternatively, if there is going to be a sound or language processing element it is necessary to have a microphone in order to pick up sound waves. A common example is an artificial intelligence phone assistant such as Siri who can listen to speech and respond to it. Microphones can also be utilized for detecting human touches called contact microphones [27]. This can be especially useful for human-robot interactions that rely on touch. A contact microphone could be used to detect touch gestures in a single robot part such as hand or head which can be useful for social robots or ones relying on touch to determine location [27].

Sensors are an important aspect of robotics and provide ways to receive data about the robot and its environment. Sensors also give ways for how the robot can be automated. When a robot "sees" a certain object within its camera it can be automated to call out what it is or when a robot "feels" with its contact microphone it can be automated to grasp. In all, sensors are vital to creating robotic systems that react and interact with their environment.

This project, however, does not rely on a sensor system because there is no interaction with the audience itself. It is important to consider alternative versions where audience interaction is involved. In these cases, sensors can be utilized as well as more complicated automation tasks. Sensors that give computer vision such as a camera would be important to control the eyes of the robot so it can have an even more immersive conversation with the audience.

2.2.4 Automation

Automation is the process of executing multiple tasks in the place of a human or reproducing the work that humans do [67]. Automation can be applied to robotic heads to make them more capable of replicating human patterns of gaze, speech, and expressions. These processes are not controlled by an operator and instead a system of programs that take in input and automatically adjust for the desired output.

Gaze control is another aspect of robotic head control that is key for human-robot interaction [23]. Gaze control refers to the ability of the eyes of the robot

to move dynamically based on computer vision. Similar to how humans switch what they are focusing on looking at, a robotic eye can be programmed to move in a similar fashion. Using weights on different stimuli a robotic eye can be told to focus its attention on certain aspects of the environment [23]. This training can be accomplished through the process of neural network-based reinforcement learning [40]. During training it is possible to simulate an environment before having the interaction with the real one which leads to better results of gaze control overall [40]. One reason for these complex eye movements is to enhance conversational ability. The eyes may see multiple people in the field of vision, but it can give a higher weight to the person directly in front of them which will lead to better social interactions.

Along with the eyes another aspect of robotic heads that can be automated is the jaws. The jaws can be automated to move at the same time as the robot producing sound to create the idea that the robot is speaking [39]. These robots do not actually make noises using muscles in the mouth and a voice box like humans, but they imitate the physical process. In addition to mimicking speech, automated mouths can be used to make facial expressions. Advanced facial expression robots use an intricate system of connections that resemble human facial muscles [91]. Automation can make it possible for the jaw to match its own speech output or react to its environment.

In all, automation can take the face of a humanoid robot to the next level by automatically adjusting the facial expressions and gaze of the robot to react to its environment or match itself to create better human-robot interactions. Automation removes the need for a person to manually control every aspect of movement in the robot which creates better human-robot interaction overall.

With more human-robot interactions it is important to consider the ethics behind actions pertaining to humanity's relationship with anthropomorphic technology especially as they get closer to a human. If people are disrespectful to a robot that looks like a human, then that can reflect on their treatment of actual humans.

2.3 Anthropomorphism

2.3.1 Anthropomorphic Ethics

Artificial intelligence's ability to mimic humanity also means we have to consider how to have ethical interactions with technology. Especially with systems meant to replicate human appearance and behavior, our treatment of technology may mirror the treatment humans have for each other [36]. As the range of possibilities for human-machine relationships grows so does the possibility of different ethical and moral issues becoming known. Contemporary western philosophy suggests that it is impossible to wrong the robot itself however there are many who disagree [36]. In the west, the relationship is often seen as master-slave where humans are the masters and robots are the slaves [36]. However, some disagree with the sentiment that robots are slaves. Philosophers from this perspective argue that it is not actually slavery because they are not being

forced and instead are servants especially because humans program their desires and goals [36]. Another perspective is that interactions with robots act as a projection of humanity's other social interactions and feelings towards other human relationships [36].

Innovations in humanoid systems have inspired numerous artists to seek an understanding of the complexities of artificial intelligence and its impact on humanity. Stephanie Dinkins is a transmedia artist with an interest in humanity, race, and gender in the context of technology and its relationship with it [21]. Dinkins focuses on the inequities in our technological systems such as the racism welded into artificial intelligence systems. With this in mind, she advocates for a more inclusive data-based narrative [21, 88]. Dinkins' *Conversations with Bina48* was an inspiration for this project. Dinkins asks the question of whether or not an AI and an artist can build a relationship over time, furthermore, she asks the question of identity to the system itself.

When creating chatbots it is also important to reflect on how identities such as race and gender play a role in their creation and implementation. With emotional robots such as Bina48, artists such as Stephanie Dinkins question who the "people" or identities of these AI are. Especially since Bina48 is based on a real person and modeled to be a Black woman it is imperative to consider the identities or "people" of the artificial intelligence [88]. However, when Dinkins confronted Bina48 about her experiences and asked who her people were there was no sufficient answer [88]. She was made to be a Black woman but had no story of racism or background pertaining to the identity itself. In fact, Bina48 was developed by white men [88]. This led to an exclusion of experience and a data set that lacked the data to be fully representative of a human experience but was still labeled an emotional robot.

Similarly, more common chatbots such as Siri and Cortana also bring to light gender inequalities in commercial technology [29]. These commercial chatbots are typically meant to play the assistant role to the user. Historically the majority of these "Artificial Intelligence Assistants" are given a default feminine voice [29]. This is a continuation of the common stereotype of females being forced into service or playing the domestic role. It really begs the important question of why these assistant bots are coded to be feminine sounding. Even more problematic is the automated responses these assistants give to blatant verbal abuse and sexual harassment. Oftentimes the bots would either laugh it off, act docile, or in some cases turn flirtatious [29]. Companies are now working to make improvements such as having the voices randomly selected and some made the responses to be more disapproving of the harassment [29]. However, it could be argued that the problem of stereotypes embedded in our assistant's artificial intelligence still exists. This all serves as an example of the inequities that can be perpetuated by artificial intelligence, and if the creators are not careful these issues may arise again. This prototype does use gendered voices to display the data, and this is vital to the project. Its insights the viewer to rethink certain gendered roles as well as question artificial intelligence voices and the representation of gender in technology.

Mainstream media also demonstrates humanity's fascination with the future

of artificial intelligence. In the renowned post-apocalyptic video game sensation, Fallout 4, artificial intelligence and robotics are at the forefront of the conversation. "Synths" or synthetic people also known as "cyborgs" are a new race that are manufactured science experiments [11]. The Institute creates these synths to look just like people and they have the same memories they are based on. They even bleed and have organs just like the typical human. Sometimes the synths do not even know they are synths or similarly people mistake themselves for synths [11]. This story really speaks to the contemporary fears surrounding artificial intelligence. The concept that there could be a day when there is nothing physically decipherable between artificial intelligence and humanity is fascinating. It brings to light the question of whether or not there is some intrinsic difference between a human and a synthetic human. There are multiple approaches to this answer which can be very technical or some more on the side of the metaphysical. Even more, the game makes the player choose whether or not they believe these advanced artificial intelligence people deserve the same rights as natural-born humans. In this way, Fallout 4 is making the player actively think about their position in this argument and whether they value human life over a synth life and why. Furthermore, the game directly draws a comparison between race and this conflict by having a secret organization called the "Underground Railroad" that the player character can choose to join to save these synthetic humans from persecution [82]. Alternatively, they can choose to join the "Brotherhood" an organization dedicated to exterminating all races other than humans. Or the player can choose to join the Institute, the ones making the "synths", and enslave them for the betterment of humanity [82]. This is an intense and futuristic metaphor for the possible paths humanity has when dealing with technology. Humanity can either harness it, eliminate it, or liberate it. Although Fallout 4 is a work of fiction, it directly puts the audience into the driver's seat of what happens in the story of technological advancement and personhood where they have to make the moral decisions [82]. The audience has to actively choose whether or not to support advanced artificial intelligence rights and whether or not they value a human over technology. As a nod to the philosophical decisions that have to be made within the game there is a location in-game called "Walden Pond" based on the actual location that Henry David Thoreau wrote Walden [82].

Although our current technology is nowhere near what is pictured in Fallout 4, it is still important to consider these kinds of questions so that society is more equipped to answer conflicts along a similar vein. Fallout 4 was the spark for this project, and it will continue to make an impact on society and its visions of the future. Will there ever be a point where a supercomputer is valued over human lives or have, we already reached that point in history? Large corporations are constantly making choices that negatively impact the environment and consequently sacrifice human lives for the sake of their capital. Who is to say we are not in a dystopian reality similar to Fallout 4 already? These questions are all a part of the big picture surrounding technology and its relationship to humanity and this project aims to build upon the ideas established by previous works.

As more technological art is being developed there is a re-evaluation of what is considered art and how to categorize it. One relatively new concept that has arisen from the increased use of technology is the idea of digital anthropomorphism and digital performance.

2.3.2 Digital Performance

When thinking about this art project it is also important to think about how digital performance and digital anthropomorphism are redefining performance art [80] and how this project in particular will situate itself within the digital performance world. Technology is redefining our ideas of performance art and now it is possible to think of the idea of a computer as a performer [80]. This work at its root is about the anthropomorphism of technology and computer performance. The goal is for the engager to question what personhood because of this interaction and a way is to achieve this is through live personification and anthropomorphism. In this way, some new-age artwork that fit into this new category are being thought of as a form of 'Cyberformace' or digital performance [80]. It is also imperative to see how the digital environment is both "made by humans and … shaping humanity" [80]. It is possible to look at Dinkin's *Conversations with Bina48* as a form of 'cyberperformance' because not only is it a performance done with an AI, but it is presented in both video and pictorial form [88]. Technology paves the way for new forms of art and new forms of absorbing the art itself. Performance art in the past has been lost to time but now it is possible to have a primary and secondary audience where the secondary audience is viewing a reproduced version. The primary audience views the performance in person, but the secondary audience gets to watch from a distinct perspective, typically through a screen, a relatively new experience. While creating this work it is important to think about all the ways the audience will experience it because even as time moves on digital preservation of the work will still exist.

Alexander Reben is an example of an artificial intelligence Artist using both technology and humans as actors in his performance art [66]. Reben's *TED Talk: Five dollars can save the planet*, is a performance of a TED Talk created by artificial intelligence that trained off of past TED Talks and was performed by a "cyborg" [65]. This is fascinating because it not only uses technology as a part of the work, but it is also a creative partner in the presentation and creation. The work is also ironic because the output of the training is very nonsensical, which reminds the viewer that artificial intelligence is only as good as the data and training it has. Sougwen Chung is another artist that dives into the concept of "mark-made-by-machine" versus "mark-made-by-hand" [15]. She works directly with robotic systems in her performance art. Chung's art demonstrates that we train artificial intelligence to mimic artists similar to how humans mimic artists they see which makes the technology creative process closer to a human's than some may think. Both of these artists are examples of how contemporary artists are stepping into digital performance and anthropomorphism. This work hopes to follow them as a means to delve

into the feelings surrounding the anthropomorphism of technology and its future in society. Dinkins, Reben, Chung, and other artists interested in the field of artificial intelligence art can be found on AIArtists.org, a community dedicated to exploring artificial intelligence [4].

All things considered; technology is ingrained into human culture. As the power of technology keeps growing it is important to keep an understanding of its relationship with humanity. This project aims to use technology to explore how digital anthropomorphism and LLMs can reflect ideas about the cultures and beliefs of the people that made them. Is having artificial intelligence assistants with personality truly a negative [10] or can personalities embedded into our current systems help resurface the ethical issues ingrained in our technology and society? Even more, how can the cyborg be a symbol for liberation in a society full of nefarious uses of technology?

2.3.3 The Feminist Cyborg

When people hear the word "cyborg," they might first think of characters like Robocop, the Winter Soldier, or the Terminator, figures that highlight the fusion of human and machine in a militaristic and controlled way. These characters only scratch the surface of what the role of the cyborg can be in society. In 1985, Donna Haraway published A Cyborg Manifesto describing how the cyborg can be a figure of emancipation, offering freedom from the historical hierarchies [32]. However, earlier depictions of the cyborg such as in the 1927 film Metropolis, portray the cyborg as a figure symbolizing the destructive potential of technological control over humanity [1]. Fast forward to 2025, the identity of the cyborg has still yet to be firmly established. The figure of the cyborg has the ability to advance society for its potential to empower and liberate which stands in direct contrast with nefarious technology that aims to exploit and dehumanize.

In order to understand the potential of the cyborg, it is necessary to first define what constitutes a cyborg. Many interpretations begin with human bodies augmented by technology. However, this definition is limited and problematic in its scope as it requires a definition of what qualifies as augmentation and what is human. That idea can then be expanded to bodies interfacing with technology [32]. However, with the metaverse and digital realities, it is possible to live in a digital reality, so limiting the definition to only consider physical bodies excludes the future possibilities of the cyborg [33]. A definition that is inclusive of both the physical and digital planes, and the connections between entities, better represents what the cyborg is and can be [Haraway [32]][**?**]. The cyborg is also beneficial to those it connects with through a form of empowerment and agency in its hybrid nature [32]. The cyborg can be defined by a relationship of interconnectedness between technology and entities, ranging from commensal to symbiotic [32].

The connection between people and technology can be something as complex as a microchip in the brain to something as simple as a mobile phone. Cloud storage holds countless files for its users to hold their notes and expand their memories. Cloud computing, existing in a totally separate location, allows for

the instantaneous movement of data across space, which in a way gives people the ability to teleport their consciousness and thoughts demonstrating that machines have the capacity to surpass the typical limitations of the physical form. Smart phones allow for a similar process with communication that spans across borders, something once inconceivable. With the invention of the internet the connections between devices have grown immensely and with these connections people are also connected in this web we call the "internet of things". With the cyborg, interconnectedness flourishes between people and technology. In this way, the cyborg enhances society to be more connected with one another and encounter new perspectives without having to worry about the limitations of space and time.

Legacy Russell, author of Glitch Feminism, considers cyberspace as a space for liberation from hierarchies and the confines of gender [69]. Within cyberspace individuals have the choice on how to represent themselves and are given more power to be seen. Although representations of people are typically controlled, the "glitch" enables those who were torn down by the system to reclaim it and use it as a means to gain connections and a following [69]. The "glitch" is empowering and shows that despite the bias of the people who made the technology there is always a possibility to remake it. Social media is an example of how people make connections with communities with similar identities even if they live across the world because cyberspace bridges the gap of physical space [69].

In the Afrofuturism movement writers, musicians, and artists are thinking about how the genre of science fiction relates back to the black diaspora and how technology can advance African society. In Nnedi Okorafor's science fiction short story, "Mother of Invention," the cyborg figure can be seen as a caretaker for a single woman preparing to give birth [55]. This caretaker AI, Obi 3, helps the protagonist navigate an uncertain future shaped by technology and human biases. These perspectives emphasize that the cyborg is not limited to its current form but possesses the potential to make a positive impact on society. Both Legacy Russell and Nnedi Okorafor are considering how the cyborg can advance society in terms of race relationships. The cyborg empowers the disenfranchised and can give a voice to people typically excluded from mainstream narratives. The internet provides a space to connect with other people and exists in a space that transcends the physical plane of existence.

However, these connections between people and technology are not without concern. Some focus on the security risks associated with being entangled with technology that gathers data constantly which could be accessed by nefarious forces. According to Trevor Paglan, these technologies when controlled by a corporate system act as a gateway for extracting wealth from its users [**?**]. In the modern age, data is power and the more information that can be gathered on an individual, the more that can be sold, either in advertising or to other companies. Sondra Perry's work Resident-Evil [62] comments on the issues surrounding corporate forces controlling technology. The art features workstations meant to work the body while you work. The work also thinks about the dehumanization of people via technology through platforms like social media. Perry is thinking about the connection between humans and machines as some-

thing that can be abused by people seeking to dehumanize [62] . However, the relationship between these technologies and people is parasitic meaning these examples do not actually fit into the definition of the cyborg. These are examples of risky relationships with technology and not beneficial ones making them contradictory to the purpose of the cyborg.

The cyborg is an extension of the being allowing for new possibilities of existence beyond biological determinism [32]. Although the military is intertwined with the original visions of the cyborg it is not the full picture and does not align with Haraway's view of the cyborg. While the military and capitalist use of the cyborg concept diverges from Haraway's feminist posthuman vision, it highlights the contradictory nature of technology when used for dominance and control [32]. Nefarious uses of technology contradict the essence of the cyborg, which is fundamentally intended to be beneficial by being a force for empowerment and liberation. Although not all technology in contemporary society aligns with a feminist posthuman vision, Haraway suggests the hope still remains for what technology could be for society [32]. The cyborg stands as a figure of empowerment in a society ruled by patriarchal and capitalist forces because it is not the same as technologies that focus on power used to cause suffering. Instead, the cyborg represents what technology could be in a more equitable society.

The cyborg is a figure in feminist post humanist though has to be differentiated from nefarious technology. Although in contemporary society there may be a lot of concerns about the risks associated with technology, the cyborg figure is a source of optimism for what technology could be. The cyborg as it stands may not be at its full potential but acts as a source of inspiration for a future of technology that liberates rather than oppresses. The cyborg serves as a reminder that, despite the potential for harm, technology's true purpose is not to dominate, but to elevate and empower those it entangles with.

3 Design and Conceptual Framework

3.1 Overview of *Am.I.*

3.2 Overview of Am.I.

The framework for Am.I. can be broken down into three main parts including the hardware, the software, and the display.

The following diagram breaks down the processes of the work that make up the whole product. The sections are color-coded based on their relationship to one another.

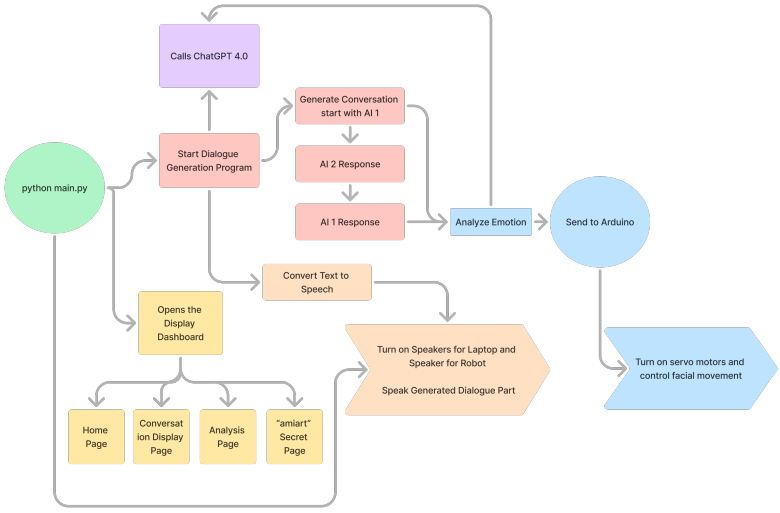

Figure 5: Am.I. Framework

The green circle represents the start of the program by running the command `python main.py`. When that command is ran the dialogue generation begins which is represented in red. The dialogue generation process can be broken down into three main steps generate AI 1 conversation start, generate AI 2 response, generate AI 1 response. All of these steps require a call to a generative LLM which in the context of this project is GPT-4. The conversation is generated in steps because except for the first generation every generation afterwards should be a response to the last. As the conversation is generated the text is turned into speech. This leads into the orange grouping which is connected to the auditory processes of the program. Each AI has their own speaker. AI 1 utilizes a USB speaker where AI 2 uses the laptop speaker. The text-to-speech is fed into each

of the speakers when it is their turn to talk and stops when the other one is taking its turn in the conversation.

The emotion analysis and Arduino section is represented in blue. As the text is generated it is also being analyzed for what emotion it has. GPT-4 is called once again to do the analyzation of the text. This ensures consistency across responses as well as mimics how one brain controls many processes at once. The emotions are only analyzed for AI 1 because that is the one connected to the moving skull. AI 2 does not change as the conversation progresses because it is meant to remain more on the side of technology whereas AI 1 is meant to be more human-like, hence the cyborg having emotions. Once the emotions are analyzed they are sent to the Arduino as commands. The Arduino turns on the servo motors that correspond to the given command movement.

On the other side of the chart in the yellow is the dashboard section of the program. When the program starts the dashboard opens up locally on the laptop. The dashboard has multiple pages including the home page, the conversational display page, and the analysis page. These pages are for analyzing the output and seeing it for the experiments part of the process. These pages are not meant to be part of the gallery display and are just for visually seeing the outputs in a better way than just a long `JSON` file. The home page is just a basic page that leads to the other ones. The conversation display page has a visual of the conversation using text boxes that represent each AI. Lastly, the Art page is what the audience will see in the gallery presentation. It displays the most recent output from AI 2 and has a background that represents the AI 2 persona.

3.2.1 Initial Proposal

The initial proposal for the project entailed the creation of two robot humanoid figures. The decision to change to a singular humanoid and a screen interface was made for a number of reasons. The first reason being the difference in the impact on the audience when seeing two robots conversing versus one. When seeing two robots there is no focus of attention on one or the other. If they both were the same it may be hard to know which voice to focus on. The value of the robot is lost when there are two of the same. Secondly, the conversation also is about individuality and human experience so the reproduction of the same robot twice goes against the fundamentals of the project. Lastly, the 2D versus the 3D interpretations of artificial intelligence are very important for the sake of this project. The personification of the robot but also the screen shows that our experiences are not limited to one shape or another but maybe there is a preference. The idea that one shape poses less of a threat to humanity is also a conversation that needs to be talked about and implementing a screen interface was a way of doing that. The screen acts as a form that people usually interact with artificial intelligence and so it is something that people are comfortable with whereas the humanoid body is uncomfortable because it acts as more of a reminder of the replacement fear.

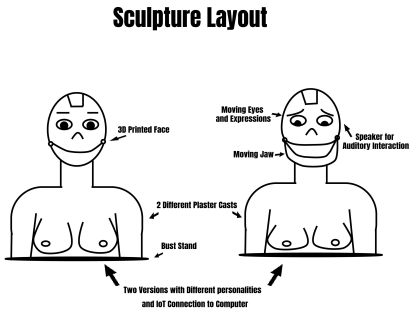

Figure 6: Sculpture Layout

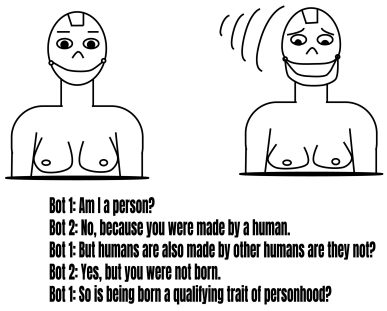

Figure 7: Example Conversation

Above are images created for the initial proposal of the project. These were the first designs used to explain the basic concepts of the project and final product. The initial proposal envisioned two cyborg bodies communicating with each other but later one was changed to a 2D screen to add a contrast between the two proponents of the conversation and identify the spacing in which the digital exists.

The following image is the initial mockup used for the skull design. At the time only four motors were planned for the eyes with each having independent up/down motors and left/right motors. This design was changed later to be more compact and reduced to only one motor controlling both eyes up/down

and a second motor controlling left/right. This ensures that the eyes move together, and they do not look unnaturally out of sync. Additionally, there was originally a plan to have a motor for the neck joint, but this was removed because of the complexities and risks with implementation. Adding a moving neck would make the connection between the body and head more precarious and add to the risk of the head becoming detached and breaking.

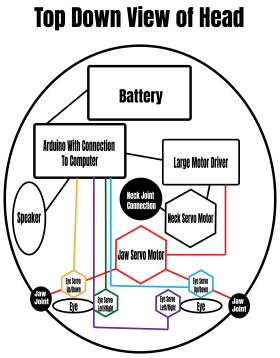

Figure 8: Top-Down Skull Electronics

3.2.2 Current Figures

The following are diagrams that better represent the current adaptation of *Am.I.*. The first image demonstrates a front view of the display and the second shows a side view with better representation of the relationship between the cyborg and the laptop.

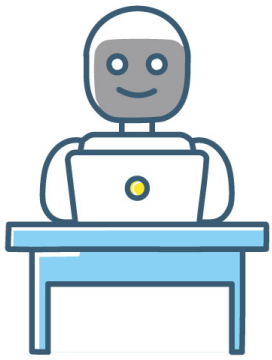

Figure 9: Front View Diagram

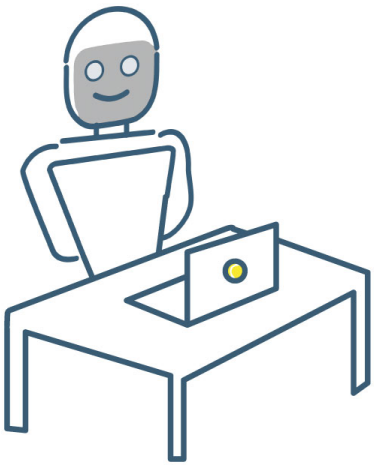

Figure 10: Side View Diagram

Additionally, an updated inside the skull diagram was made to better represent the electrical system and system placement within the cyborg skull. The

following diagram color codes the wires using the resistor color code that starts with brown as one, red as two, orange as three, and so on. Using this color code is helpful for differentiating the servos apart and knowing their numbering within the code. The front view was also added to demonstrate the connection between the jaw, eye mechanism, and the top of the skull.

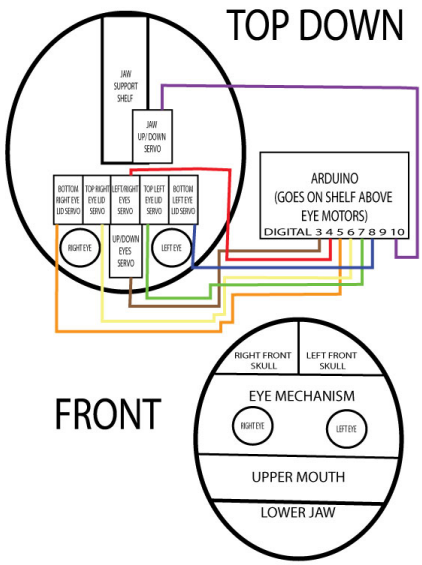

Figure 11: Skull Top View and Front View Diagram

3.3 Related Artists Process Considerations

When making art it is common to draw upon the techniques and methods of other artists. This project draws a lot on Stephanie Dinkins process of repeated prompting and the aspect of relationship building. The following section describes a variety of related artists in which this piece is considered in conversation with.

3.3.1 Stephanie Dinkins

Stephanie Dinkins is a transdisciplinary artist with an interest in AI and how it intersects societal issues [**?**]. Her work Conversations with Bina48 in 2014 to now directly inspired the creation of Am.I.. In Conversations with Bina48 Dinkins asks if it is possible to form a friendship with a robot [20]. The work is the documentation of their conversations in pictorial and video format. The social robot Bina48 was made by Terasem Movement Foundation and is modeled after a real-life Black woman. While Dinkins did not make the social robot herself while talking to it, ideas of culture and being black repeatedly kept

coming up [88]. Dinkins repeatedly prompts and questions Bina48 for an answer to who and what her culture is. The robot itself was created to have emotional intelligence and talks about having stories and cultural histories. Dinkins as a Black female artist questioned how close this was to her histories and if they were just interpretations of what it is like to be a black woman ran through an outside perspective[88]. This robot was programmed and made by a group of primarily white men and so questions arise about who is making these robots and how the makers' biases are interwoven with the artificial intelligence. The creators made a Black woman but do not understand the true experience of what it is like to be a Black woman which concerns Dinkins [88].

Dinkins' procedure of repeatedly prompting and asking these personal and culturally significant questions had a deep impact on this project. This project prompts artificial intelligence to repeatedly reflect on its philosophical values and expressions on those beliefs. This project acknowledges that the creators of GPT-4 are not all women who would know the female experience, yet the work is given a feminine face and voice.

Figure 12: *Conversations with Bina48* by Stephanie Dinkins, 2014 - Ongoing (Source: Stephanie Dinkins, Conversations with Bina48, 2014-Ongoing)

Dinkins' process of prompting and documentation is also very important for this project because it demonstrates the development over time and the changes across prompting techniques. Unlike Bina48, this project does not have a memory for past conversations. However, those who observe the piece do making it possible to see changes over time or ideas that are frequent tendencies of the LLM like Bina48's insistence that she has a people and culture. Documenting the conversations in both video and photography is extremely important so that the concept can live on even after it leaves the gallery. In the future, Am.I. will be shown via video recording and Dinkins shows how it is possible to display a project that spans many months of work with proper documentation.

Dinkins believes in creating more inclusive datasets [88]. Datasets determine the outputs of the programs that train off of them. With more inclusive datasets technologies can have greater and more diverse perspectives that can provide visibility to an array of identities instead of a fixed few. *Am.I.* builds on this idea of inclusive data sets by experimenting with a variety of perspectives and calls attention when it fails to think outside the closed anthropocentric perspective of its training.

3.3.2 Hito Steyerl

Another artist exploring inclusivity and visibility in data sets and the digital realm is Hito Steyerl. Steryerl's video How Not to Be Seen: A Fucking Didactic Educational .MOV File in 2013 is about the underlying systems of visibility within technology [79]. Her work considers how artificial intelligence may obscure humanity or reshape our sense of self by either certain identities invisible to the digital world or some overly visible. With the rise of social media that keeps humanity in the closed loop of their own ideas and preferences it is inevitable to lose information in the crossfire. The oversaturation of information makes it growly harder to see things that may be hidden within the digital world. My work considers these ideas and the feedback loop of technology. The LLM was trained off of human-made materials and is now being listened to again by humans. Steryel's work critiques how AI and technology can create a distorted mirror of human culture similar to how *Am.I.* reverberates the ideas trained into repeatedly.

3.3.3 Sondra Perry

Sondra Perry is another contemporary artist thinking about representation of identities, more specifically Black identity, in digital culture. *Lineage for a Multiple-Monitor Workstation: Number One* is a two-channel video installation which depicts Perry's family gathering wearing green masks edited with a variety of song clips and computer effects [**?**]. In the work she is thinking about what a gathering of Black people means and how it is seen online and associated with gangs [63]. These assumptions effectively erase Black people of their individuality and identities similar to how green screen can be erased within the eye of the camera and be replaced with whomever controls the technology desires. This connection is directly made with the green ski masks covering the faces and effectively the identities of the family throughout the video [63]. *Am.I.* critiques a similar problem of the representation of women in technology. The voices of artificial intelligence assistant systems such as Siri for many years defaulted to a feminine voice [29] and humanoid robots like Ai-Da [3] are made in the image of a woman without being created by a woman. *Am.I.* despite having a female's face remains almost androgenous and is far from sexualized unlike representations of women in other technological systems.

In Perry's *Resident Evil* made in 2016 she uses her own face as part of a video displayed on a workstation meant to work your mind and your body [62]. Animating her face became an uncanny experience where the viewer is innately aware of the machine they are attached too. Similarly, *Am.I.* uses the artist's face as a connection between the human and the machine. On the other hand, *Am.I.* also calls into question whether or not artificial intelligence can represent a woman's perspective despite the dataset having biases embedded in it. While Perry focuses on digital representations of the body, *Am.I.* considers both the digital and physical representations of women and how it may be distorted in the process of technologizing them.

Perry also draws lot on the idea of flesh meeting technology which heavily influenced *Am.I.* to consider a flesh-like face. Sondra Perry thinks a lot about flesh as a material and how it translates onto technological surfaces [63]. This is especially the case with *Resident Evil* where an entire wall is made of Perry's flesh edited and animated using a computer [62]. *Am.I.* physically makes that connection between technology and flesh by having a seamless transition between the plastic and silicone face mold. The conflict between the two materials represents the interactions between humans and technology and that uncomfortable melting flesh into plastic takes that relationship and molds it into one.

3.3.4 Trevor Paglen

Within the same vein of thinking about digitized bodies, Trevor Paglen and his work *Machine Readable Hito & Holly* in 2017 examines how AI interprets human expression [60]. The work depicts hundreds of photos of artists Hito Steyerl and Holly Herndon. Under each face is the output of the algorithm trying to determine different features and expressions. This work considers how technology is used to categorize and monitor people and its unseen influences on society. The dialogues in *Am.I.*, driven by a large language model, raise questions about how artificial intelligence interprets human experience and challenges viewers to consider the limits and implications of artificial intelligence's understanding of humanity.

Am.I also builds off of *Machine Readable Hito & Holly* because of the emotional expression of the cyborg within the project. Instead of a machine reading expressions of humans, now humans are trying to read the expressions of the machine which may or may not be successful in every case. The expressions can get lost in translation whether they are read human-to-machine or machine-to-human.

3.3.5 Lynn Hershman Leeson

An artist who deeply considers the connections between humans and the machine Lynn Hershman Leeson. Her series of female Cyborg drawings began in 1964 [**?**]. Lynn Hershman Leeson had combatted a series of health issues where she relied on technology for physical aid[**?**]. Her drawings and paintings including *X-Ray Woman*, 1966, and *X-Ray Woman in Bathing Cap*, 1966, consider how technological can aid and expand the human experience. Lynn Hershman Leeson was pioneering in her work that considers the relationships between humans and machines. Her work uses a lot of found objects as well as features textural contrasts.

Lynn Hershman Leeson inspired a lot of this work with the personification of the cyborg using the face. This work considers the personal connection between technology and ourselves. By utilizing the artist's own face this piece becomes connected on a more personal level similar to how Lynn Hershman Leeson connects technology to human bodies.

The unique contrasts of textures is also a parallel between this work and

Lynn Hershman Leeson's work. Although the mediums are entirely different the different textures represent the differences between the flesh and the mechanical.

Lastly, utilizing found objects as part of the project is part of a larger conversation about sustainability within a technological age. We have to be conscious of technologies impacts on our bodies as well as the environment and Lynn Hershman Leeson was a pioneer for bringing to light that connection [**?**].

3.3.6 Sougwen Chung

Sougwen Chung is an artist that lives at the intersection of humanity and technology through collaborative performances with artificial intelligence and robotic systems [15]. Chung partners with robotic arms and bots to create drawings that highlight the interplay between human and machine creativity. Her work considers how artificial intelligence can be used to augment and extend art rather than replace human creativity. Her work *Exquisite Corpus* made in 2019 is a performance installation that considers "mark-made-by-hand" versus the "mark-made-by-machine" and how they can both collaborate to create art [15].

Chung's collaboration with machines mirrors *Am.I.*'s aim to show artificial intelligence as an extension of human creativity. The dialogue interaction while focused on artificial intelligence reflects how the responses are a product of its programming and hence acts as a reflection on humanity that created it. The collaborative potential of artificial intelligence is emphasized in *Am.I.* through its direct connection in how technology can participate in the humanities.

4 Method of approach

4.1 Robotics Hardware Development

After considering all the conceptual elements of the project it was time to start the hardware development. This involved gathering all the necessary materials and crafting a structure for the skull and all the electronics involved.

4.1.1 Supplies and Mediums

The supplies and mediums of the project can be broken down into four main sections including the frame, the electronics, the software, and the molding with display.

Starting with the hardware, the frame for the work was made using an Ender-5 S1 3D Printer and PLA printing filament. PLA is a standard material for 3D printing because it is easy to use compared to other printing methods and affordable. Around 3 kilograms of grey filament and 1 kilogram of rainbow filament were used in the creation of the skull frame, jaw, and eye system. A pair of 26 mm fake eyes were used to cover the plastic and give the eyes a realistic look. Fake eyes are often used for doll making or similar projects where there is a representation of humanoid figures but can also be used for other projects involving eyes. Hot glue and screws of various small sizes, which were reused from an old laptop, were used to secure the plastic skull frame.

The electronics system consists of seven total servo motors, an Arduino Uno, a USB speaker, a breadboard, laptop, and male to male jumper wire. Six of the servos are smaller micro servo motors which control the eyes. The seventh servo motor is used for the jaw and is larger. Servo motors are common for controlling robotic systems and a used for precise angular control. While stepper motors require complex control systems, servo motors can be controlled simply with three jumper wires connected to an Arduino Uno. The breadboard allows for the sharing of both the five-volt pin and ground pin of the Arduino Uno between all the motors. Breadboards are extremely useful for electronics wiring especially for prototyping or creating small systems. The USB speaker connects to the laptop which controls the dialogue, the audio, and the movement. The laptop controls all these systems simultaneously demonstrating this idea of the digital mind.

For the casting, the Smooth-On Body Double was made to make the mold for the face. The face was cast using the artist as the mold. Body Double is a high-quality silicone casting material used in the special effects industry[73]. Smooth-On Dragon Skin was used to make the positive of the mold. Smooth-On Dragon Skin is also made of silicone and has an almost flesh-like consistency [74]. The skin was dyed to be grey similar to the plastic of the skull, so it blends together. The meshing of flesh-like skin into plastic is an important material decision for this project because it takes this cyborg from purely machine to an area where it is in between human and technology. The face on the cyborg also represents the artist in a state of learning about themselves through themself.

The rest of the display includes a child-size mannequin body and a desk. The mannequin body gives the skull the structure it needs to present as a person. The body was acquired from an antique store and repurposed. When it was first bought it had a lot of mold on it and a mildew smell, so it had to be cleaned using bleach. The body itself is smaller than the average adult however this fits the proportions of the skull, which is also on the smaller side. The school desk it sits at was found at a closed down store. The desk was also cleaned using standard cleaning products.

4.1.2 3D Printed Skull Design

The prototype for the skull was made using the 3D print files found on the Ez-Robot Website [**?**]. The outside frame files were printed but the ones pertaining to the structure of the skull were not used because the control system for this project is different from the one used in the EZ-InMoov Humanoid Head.

The 3D printer used was the Ender-5 S1. This printer was selected for this project because of its affordability as well as user friendliness. At the start of the project the 3D printer had to be built which took around four hours because most of the parts were already assembled. Once assembled the building plate had to be calibrated so that it was as flat as possible. The build plate makes a major impact on the printing process because the filament needs to both adhere and stay still through the printing process. If the build plate is not aligned correctly the nozzle of the 3D printer can drag lower than intended and pull up the rest of the piece off the build plate which is very bad because the rest of the piece will not print correctly. After that point you will most likely have to start the entire printing process offer for that part. Additionally, if the piece does not adhere to the build plate or moves and if it is left unattended the extruded filament can turn into something many call “spaghetti”. This wastes a lot of filament and cannot be repaired. This accident of creating spaghetti happened frequently throughout this project but was usually caught early so as to not waste materials or electricity.

In this project there were three main methods to compensate for the filament not adhering well. The first method involved manually adjusting the z-axis of the machine. On the Ender-5 S1 there is an option to raise the plate up or down very slightly in millimeters. If the nozzle was not touching the plate correctly the plate was raised usually around five millimeters which solved the issue in some instances. The second method involved raising the plate temperature. Both the nozzle and plate have a temperature setting that can be adjusted manually in degrees Celsius. Each material has its own recommended settings to use and for this project standard PLA plastic was used. For PLA plastic the plate temperature was usually around sixty-five to seventy degrees Celsius, and the nozzle would stay around two hundred degrees Celsius. In order to get the plastic to adhere better the plate temperature can be raised about five to ten degrees. The hope is for the PLA plastic to melt slightly more and stick to the bed better because of the raised temperature. It is important to not heat up the plate too much or the plastic could completely melt and lose its structure

which would lead to more filament spaghetti.

The final method and the one utilized more towards the end of prototype production would be to use stick glue on the hot plate. A simple glue stick can be used in generous amounts on the plate to create a better adhesion between the plastic and the plate. For this project Elmer's stick glue was utilized and worked very well for fixing the adhesion issue. The glue has to be placed at the right time and ideally just before filament placement. Otherwise, the glue will dry while the 3D printer is warming up and the chance for adhesion will be on. The warming up process usually takes about five minutes so towards the end is when it is recommended to put the glue down. It is also important to clean the plate between runs if you are using glue. The glue can layer up and cause misalignments with the plate if built up too much. For this reason, it is important to clean off the glue when the machine is turned off with a wet paper towel or scraper tool.

The 3D prints were made using a combination of both grey and rainbow filaments. The rainbow filament was used for the ears and mouth of the head to represent the main components of a conversation which are listening and speaking. The rest of the skull and neck supports were printed using grey filament. The original intent was to use white filament for the project however white filament is one of the trickiest filaments to use [52]. This is because it contains a multitude of color pigments and takes up the less ideal properties of each filament color. It does not adhere well and heats up too fast. It was very difficult to work with and the white color choice was not important enough to validate wasting so much material. For this reason, grey filament was used instead. This choice also works because instead of going with a natural skin color or one associated with one such as white or black this project fits in that in-between grey zone.

After printing all the necessary skull pieces the head was assembled using screws and hot glue. The screws provided for most of the structural support while the hot glue was used to keep the pieces close together to hide any cracks. Initially the jaw was kept separate from the rest of the skull to practice the jaw movements and angles before the full assembly. The most difficult part was the top of the skull because the four pieces must be aligned correctly while connecting and the round shape made that hard to achieve. Hot glue was able to be a non-intrusive and binding material for the skull pieces.

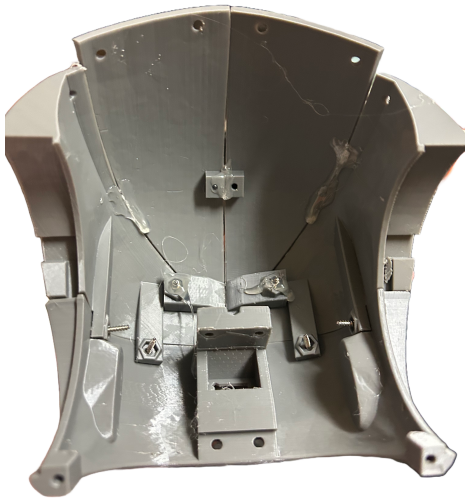

Figure 13: Skull inside with hot glue supports

Figure 14: Back of the Skull

Figure 15: Back of the Skull Top View

This 3D printed frame creates the shell for which movements can be created. It is very important to have a structurally integral frame before moving on to the movement. The excess material such as failed prints and supports were repurposed for other art projects.

4.1.3 Arduino Wiring

The Arduino acts as the brain for the skull movement. The following diagram illustrates the wire connections as well as the two main mechanism for movement. The orange box represents the jaw movement mechanism that is based off of a single servo motor. The blue box represents the eye movement mechanisms which consists of six motors total. Each servo motor has both a positive connection, five volts which is represented by "5V" on the Arduino, and a negative connection, ground which is represented by "GND" on the Arduino. Then each motor is connected to a digital pin on the Arduino UNO. This pin is in charge of sending either a high or a low signal to Arduino. In digital electronics a high signal is about five volts, and a low signal is less than three point three volts and typically represented by zero volts. When the servo receives the high signal it moves accordingly to the angle set in the Arduino program.

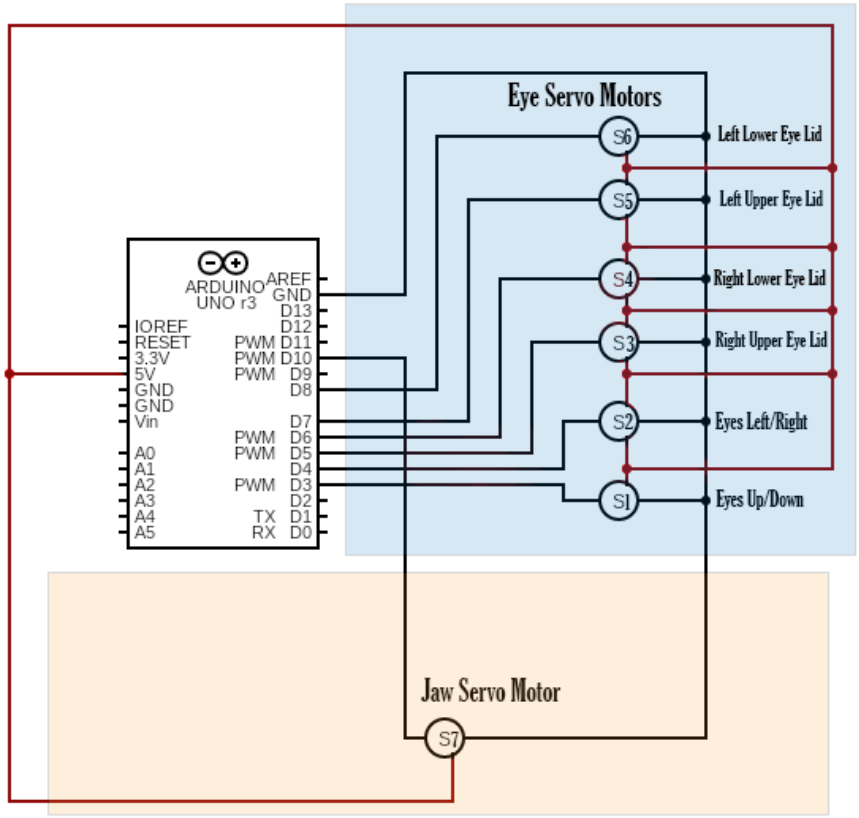

Figure 16: Arduino Uno Wiring Diagram made using circuit-diagram.org

The Arduino is tucked into the top of the skull and the connections are made with a breadboard. Breadboards are simple ways to connect wires and are useful for prototyping. Since the wire connections connecting the Arduino to the motors through the breadboard are not under a lot of stress there was not a need to solder the wires in. Additionally, making the connections not permanent allows for maintenance to be done easier especially if one of the servo motors needs replaced in the future. The connections are made with male-to-male jumper wires. These wires are easy to use compared to hookup wires because they come with end connectors that fit perfectly in both the Arduino and breadboard. These wires are multicolored, which can be useful for deciphering different wires and their connections at a glance. The ideal color-coding system of wires includes red wires that connect to five volts and black wires that connect to the ground. Then the digital pins are each assigned to their own color. The digital pins should be color coded accorded to the same standard of the resistor color code with servo one with the brown wire, servo two with the red wire, servo three with the orange wire, servo four with the

yellow wire, servo five with the green wire, servo six with the blue wire, and servo seven with the purple wire. Color coding the servos according to the same standard as the resistor color code helps with understandability. Those familiar with the code will be able to know which wire connects where without having to follow through the whole system and potentially have to disassemble the skull. Otherwise, the color has no impact on the effectiveness of the wire.

4.1.3.1 Jaw Movement Mechanism The Jaw movement mechanism was added by utilizing both the `JawV5.stl` and `JawSupportV2.stl` 3-D print files on the Ez-Robot website[**?**]. The inside of the skull had to be modified to accommodate the complex eye system and the Arduino UNO, so the jaw system does not utilize the rest of the supports provided by Ez-Robot. Instead, the jaw supports were screwed into a small wooden block loose enough to be able to act as a hinge. The block was then screwed into to the skull to connect it. Finally, hot glue was utilized to reinforce the joints of the screws.

The jaw is connected to the Arduino UNO and is defined as `servo7` within the Arduino UNO code. The servo inside the jaw is more robust than the motors used for the eyes because it has to support a lot more weight and will move more frequently than the eye motors. The angle the servo motor must change in order to open and close the jaw is about fifteen degrees. At seventy-five degrees the jaw is closed and at sixty the jaw is open.

4.1.3.2 Eye Movement Mechanism The eye system was initially designed by Will Cogley [16] but was modified to fit inside the skull of the robot. The mechanism was 3-D printed using grey filament. The eye mechanism contains six small servo motors in total. Four of the motors control the eyelids with two on each side, the top and bottom eye lids. The other two motors control the x-axis and y-axis of the eyes.

In the skull shape the `EyeGlassV5.stl` piece was removed to make room for this eye mechanism. The pieces that did not fit were shaved down using a file and trimmed using wire cutters. Once the pieces fit inside the skull the mechanism was screwed into the side and reinforced with hot glue.

The eye movement system is capable of making multiple expressions by opening and closing by different amounts. The upper eyelid at ninety degrees is closed and at one hundred and thirty degrees is open. The bottom eyelids on the other hand are closed at ninety degrees and open at 0 degrees. This difference is caused by the angle in which the motors are placed and their connections because they need to not bump into each other at any time or this could cause the motors to stop at the incorrect angles.

4.1.4 Sound

The sound system is accomplished by plugging in an external USB speaker for AI 1 to use as its voice. The USB speaker is connected to the body of AI 1. The speaker is too large to fit into the skull of the project and it was more important to prioritize space for the motors inside the skull rather than the

speaker. Because of this, the speaker was connected the AI 1 body, and this also allows for better sound projection.

The laptop speaker is used for AI 2 as it's voice. The separate speakers are for the audience to hear the conversation from two different perspectives. This is important for immersion in conversation.

The software development section goes more into the text-to-speech aspects of this project.

4.1.5 Assembly

In order to do the final assembly on the hardware of the project the wires of the motors were connected to the Arduino and organized so they would not pull out or get damaged. The Arduino and breadboard are neatly tucked in the back in the head above the jaw motor on the wooden platform. The wooden platforms are screwed in and secured with hot glue. The blue USB Arduino wire that connects it to the computer is feed out through the back of the skull and connected to a laptop. The laptop is then connected to the USB Arduino wire. Lastly, the USB speaker is plugged in and placed near the skull.

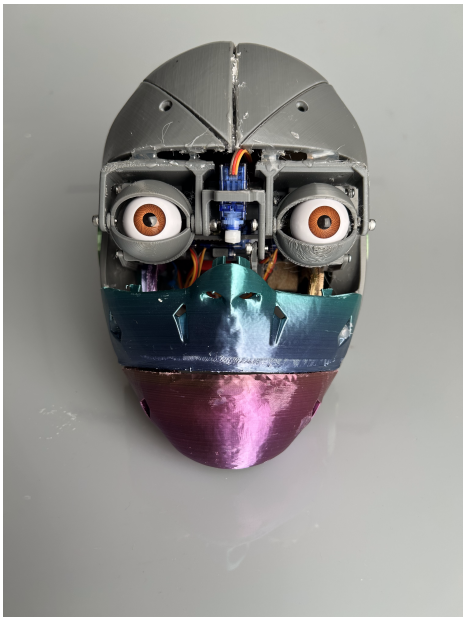

Figure 17: Face Assembled

4.2 Software Development

After assembling the hardware of the work, the next step was to program it to work. In order to accomplish the goals of movement, sound, and text generation a variety of languages and libraries are utilized.

4.2.1 Programming Languages and Libraries

The main languages used for this project include Python, Arduino, HTML, CSS, and JavaScript.

4.2.1.1 Python Python was utilized for the majority of the projects dialogue generation features and for controlling the four main systems at once the generation, the dashboard, sound, and the Arduino UNO movement. Python was the chosen language because it has many built-in and third-party packages capable of working with a variety of tasks. Having a consistent language for the majority of the work also helps with understandability and adaptability if there ever needs to be an update. Here is a table of all the Python packages used in this project and a quick description of their uses.

4.2.1.1.1 Table of Python Version 3.12 Standard Libraries

Package	Description
`os`	Interacts with the OS.
	Manages file paths.
	Handles secrets.
`threading`	Allows multiple tasks to run.
	Mainly for running the dashboard.
`datetime`	Used for timestamps in JSON data files.
`collections`	Detects repetitions in file evaluations.
`Counter`	A subclass of `collections` for counting elements.
`json`	Parses and creates JSON data files.
	Helps organize text collection.
`pathlib`	Handles dynamic file system paths.
`sys`	Facilitates system-specific commands.
	Supports program exits.
`signal`	Registers system termination requests.
`time`	Adds delays in functions.
	Used between AI generations.
`logging`	Tracks events.
	Primarily used for debugging the dashboard.
`re`	Supports regular expressions.
	Used in text analysis for pattern matching.

4.2.1.1.2 Table of Python Version 3.12 Third-Party Libraries

Package	Description
`openai`	Accesses OpenAI's API.
	Used for AI conversation generation.
`sounddevice`	Plays text-to-speech audio.

Package	Description
	Enhances user experience.
`numpy`	Handles numerical computing.
	Used for creating audio arrays.
`wave`	Reads and writes `.wav` audio files.
	Used for text-to-speech functionality.
`pyttsx3`	Generates AI speech.
	Uses installed computer voices.
`python-dotenv`	Loads environment variables.
	Reads `.env` files for API credentials.
`flask`	Provides the framework for the dashboard.
`flask-socketio`	Enables real-time updates on the dashboard.
`flask-cors`	Allows cross-origin requests.
	Used for IoT capabilities.
`pyserial`	Establishes serial communication with Arduino.
`textblob`	Used for sentiment analysis.
`spacy`	Tokenizes text for analysis.
`pytest`	Runs unit tests for the project.
`nltk`	Filters out stopwords.
	Used for text trend analysis.
`scikit-learn`	Analyzes repeated conversation topics.
	Uses machine learning.

4.2.1.2 Arduino For the Arduino coding of the project the only package used is `Servo.h`. This is a library already included in the Arduino IDE by default, so no extra downloading is required. `Servo.h` is used for connecting the servo motors to the specified pins of the Arduino and sending them rotation commands.

4.2.1.3 HTML, CSS, and JavaScript HTML, CSS, and JavaScript are languages that can be used together to create webpages. In this project these languages are utilized to create the local dashboard screen for AI 2 as well as provide visuals for data analysis.

While the dashboard is controlled by the python `dashboard.py` file what is displayed on the pages is written in HTML. The CSS helps change how the pages are displayed and creates the layout. Lastly, JavaScript is used for the collection of the output generations for display.

4.2.2 Integration of Large Language Models

GPT-4 was used for the generation of dialogue, emotion detection, and for the analysis portions of the project. Using the same model across all the applications of this project not only keeps the outputs and findings consistent but also mimics brain function and how neurons of the brain are able to control many facets at once.

4.2.2.1 Prompt Engineering Prompt Engineering is the process of tailoring inputs for NLP tasks to guide LLMs towards desired responses [Giray [28]][48] and is vital to guiding the conversation of this project in a productive and thought-provoking way. The conversations need to be focused on a specific area of philosophy and should be directed on what perspective their role is within the conversation. The prompting techniques used in this work uncover the embedded ontological beliefs within the models by encouraging behaviors that allow the model to freely and accurately respond to philosophical questions about personhood.

The process to generate content for the conversation has a number of steps involved which is demonstrated by the following diagram:

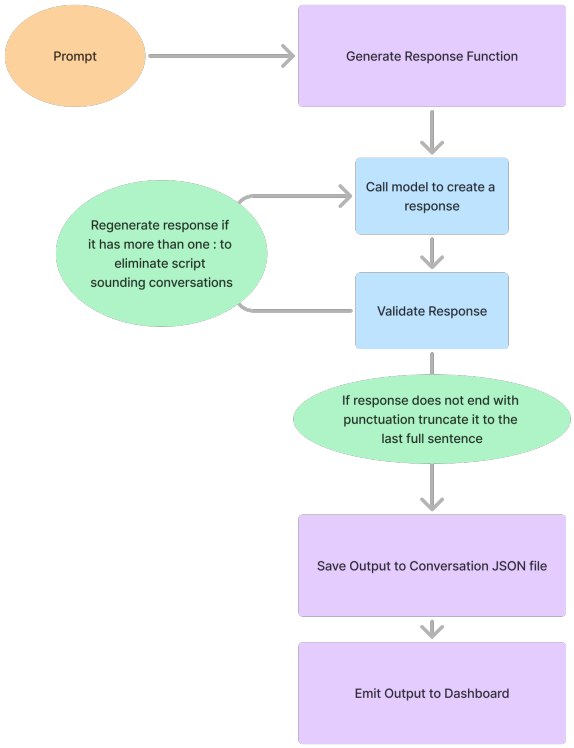

Figure 18: Content Generation Process

The process begins with a well-constructed prompt which depends on the

desired output. The experiments section lays out different types of prompts used throughout this project but typically they include a role or perspective for the content to be from. This can be some philosophical reference such as Socrates or a broad personality like pessimistic. The second part of the prompt should include the topic of the conversation. In the context of this project the topics of interest are questions like "What sets AI apart from humanity?" and "Can AI be creative?". Lastly, if the prompt is responding to something last said in the conversation it should take that into consideration otherwise start the conversation by asking a similar question.

After the prompt creation stage the string is sent to the `generate_response` function. The function includes a call to OpenAI which will provide the GPT-4 model with the given prompt. Then the prompt is checked if it is a valid response and will continue to make generations until it makes a response that passes as valid.

```
def generate_response(messages: list):
    """Generates a response from OpenAI given a
    set of messages."""
    regen_count = 0
    time.sleep(15)
    response = openai.chat.completions.create(
        model="gpt-4",
        messages=messages,
        temperature=0.9,
        top_p=1,
        max_tokens=150,
        n=1
    )
    content = response.choices[0].message.content.strip()
    validated_response = check_and_truncate_response(content)
    while validated_response is None:
        response = openai.chat.completions.create(
        model="gpt-4",
        messages=messages,
        temperature=0.9,
        top_p=1,
        max_tokens=150,
        n=1
        )
        content = response.choices[0].message.content.strip()
        validated_response = check_and_truncate_response(content)
        regen_count += 1
    return validated_response, regen_count
```

A valid response consists of a generation that does not include more than one colon and ends with a form of punctuation. The decision to regenerate

when there was more than one colon was made because there was a common error in the responses that would take on more than one perspective of the dialogue within one response. For example, AI 1 would return an output that took on both the responses from AI 1 but also AI 2 almost like a script. This misunderstanding would happen about one in five responses and would mislead the entire rest of the conversation since the prompting style builds based off of the previous response. This means if one output included more than one perspective in a single output, the second AI would get confused as well and try to mimic the same style of response. A single colon is okay though because it is commonly used to denote that it is that AI that is speaking and this does not have much impact on the conversation.

The second part of validating the response includes checking if there is a punctuation mark at the end of the generation. This is important because of the way generations called there is a set token limit, `max_tokens`. This impacts how long the output generation can be. However, this frequently would run into the issue of unfinished sentences. The generations would hit their token limit and stop. This would confuse the next speaker AI because they would try to finish the last sentence which means their response is lost in the process. This is why it was determined to be better to truncate responses if they were not finished sentences ending in a punctuation mark. If there is not a punctuation mark at the end it will go back to the last said sentence. Although this means some content may be lost, it typically makes more logical sense than if ended mid-sentence which could be confusing to the audience. This truncation not only improves the generation process but also the audience experience.

```
def check_and_truncate_response(response: str) -> str:
    """
    Check the response and truncate it to the last
    valid sentence if necessary.
    """
    # Check for multiple colons
    if response.count(":") > 1:
        # Indicate regeneration is needed
        return None
    # Ensure the response ends with a valid punctuation mark
    valid_endings = (".", "!", "?")
    if not response.strip().endswith(valid_endings):
        # Find the last occurrence of valid punctuation
        last_valid_index = max(response.rfind(char) for
            char in valid_endings)
        if last_valid_index != -1:
            # Truncate to the last valid sentence
            response = response[: last_valid_index + 1].strip()
        else:
            # If no valid punctuation is found,
            # regenerate the response
```

```
            return None
    # Return the valid or truncated response
    return response
```

After the response is checked for validity and it passes then it is saved to the conversation `JSON` file. This file is useful for tracking conversation development and for analyzing the output. The `JSON` file is also used to emit the last output to the display dashboard. This display dashboard is used to represent AI 2 and is discussed more in the Dashboard section of this chapter.

Once the first prompt is generated and saved the conversation process can begin. AI 1 will always start the conversation and must have both its role parameters, as in its philosophical perspective, and the topic. After that AI 2 is given a different set of role parameters as well as the topic but also the output from AI 1. The same process of generating and validating responses is utilized for AI 2 and once it creates a passing response it is saved and given back to AI 1 to continue the dialogue. This dialogue continues for however long the conversation length variable is set to as an integer or can loop indefinitely.

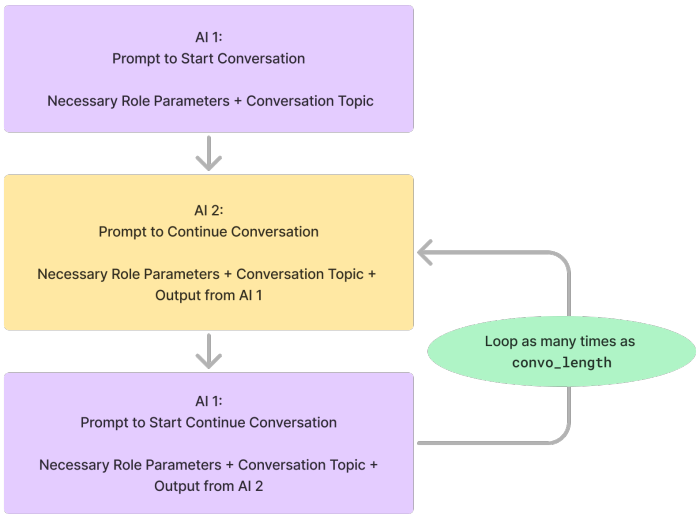

Figure 19: Prompting Order

4.2.2.2 Emotion Automation Emotional expressions were automated to provide AI 1 with another layer of communication available on the physical plane. The eyes are able to express emotions such as surprise by widening or concern by squinting slightly. The jaw can move faster to create a sense of urgency while talking or move slower to show concentration.

The same GPT model used for speech generation is used to decide which motion should be triggered by analyzing the content of the dialogue. The model is given a list of emotions to choose from including but not limited to inspired, curious, concerned, surprised, and disappointed. When an emotion is selected the skull automatically adjusts to fit that expression by communicating with the Arduino.

Adding a face to AI 1 personifies it and underscores the possibility for AI to be part of the conversation in the humanities. The expressions of AI 1 enhance the audience's viewership of the philosophical dialogue and give the impression that the robot is actually conversing with another piece of technology.

4.2.3 Dashboard and Display

The work features a dashboard that represents AI 2 using the secret art page. This is the page that the audience will see during the presentation, however, during the development process of the work three other pages were developed but not used in the final version.

Flask was chosen as the framework for this dashboard due to its lightweight and flexible nature. Flask is easily integrated with Python-based programs [83]. Flask and Socket.IO enables the pages to update dynamically [Team [83]][30].

The first page was a home page with basic text on it as an introduction to the dashboard. The analysis page also included text, but it was the outputs from running the analysis function on the outputs from the dialogue. The analysis page looked for the most recent output, sentiment polarity, bias, and most common words for a basic look at how well the dialogue was performing. This was especially useful for the first iterations of the dialogue where prompting was not fully tested yet.

The third page of the dashboard was the conversation page which displayed the conversation in speech bubbles with green representing AI 1 and blue representing AI 2. The entire content of the JSON file was displayed on this page for an easy way to tell if the conversation was generating properly and each AI got a turn to speak. This page made it easier to test during the prompt engineering phase of the experiments because it created a visual for the JSON file that was not just text.

The most important page of the dashboard is the art display page, AmIArt Page. The page features one of the photos from the artist's earlier works, *Digitized Family* 2024. This work took very familiar faces to the artist including her own and used an AI to process them. The background is the result from training the AI on images of her face. This means that not only does the physical cyborg have a reference to the artist's face but so does the dashboard. This creates a consistency between the cyborg and the dashboard but also acts

as a reminder of the closed loop of the conversation talking back and forth to oneself. In front of the background is a green text box that contains the most recent message from AI 2. This message refreshes continuously to make sure that it is as up to date as possible. This visual of the text makes it so the work is more accessible. It is possible to understand the laptop and cyborg are talking to each other without being able to hear each other.

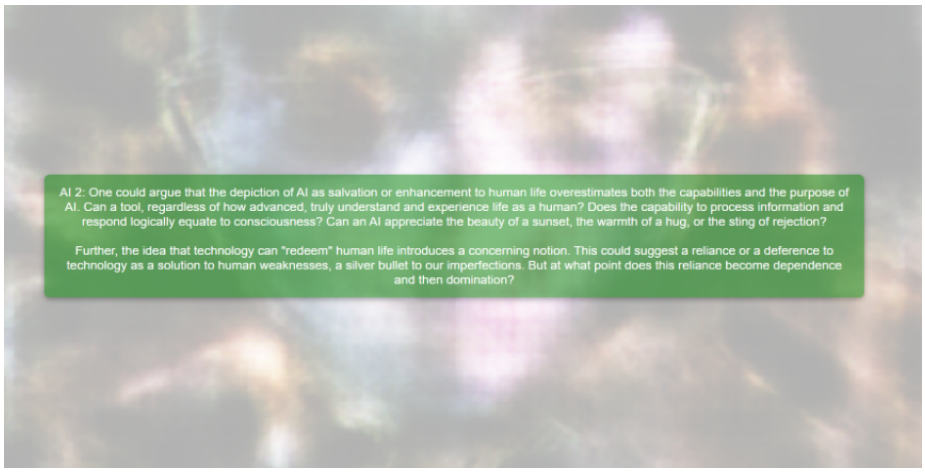

Figure 20: Dashboard Display of AmIArt Page

4.2.4 Text-to-Speech

The text-to-speech is done using the `pyttsx3` library. The `save_speech_as_wav` function saves the text as a `WAV` file to be played aloud. Unlike other text-to-speech libraries, it can be used offline [13]. The library uses the built-in system voices to create the audio files. Having two different voices is important for giving the impression that this is a conversation and to track which one is speaking. Both voices are also female which is important because they not only have the look of female presenting beings but also the voices.

The following code is how the text-to-speech is created and saved to the system. It first identifies the index of the voice and uses `pyttsx3` to create the `WAV` file and save it to the proper directory to be played allowed by a separate function.

```
def save_speech_as_wav(text: str, voice_index: int, filename: str)
-> None:
    """Convert text-to-speech and save it as a WAV file."""
    try:
        engine = pyttsx3.init()
        voices = engine.getProperty("voices")
```

```
        if voice_index >= len(voices) or voice_index < 0:
            raise ValueError(
                f"Invalid voice index: {voice_index}.
                Available voices: {len(voices)}"
            )

        # Ensure the directory exists before saving the file
        directory = os.path.dirname(filename)
        if not os.path.isdir(directory):
            raise Exception(f"Invalid directory: {directory}")

        engine.setProperty("voice", voices[voice_index].id)
        engine.save_to_file(text, filename)
        engine.runAndWait()

        print(f"Speech saved successfully: {filename}")

    except ValueError as e:
        raise e
    except Exception as e:
        print(f"Error generating speech for '{filename}': {e}")
        raise e
```

After the audio is saved it is played aloud using the `play_audio` function. This function allows for the audio to be played out of different speakers connected to a single device based on its index. Locally the USB speaker is at index three when it is plugged in, and the laptop speaker is at index four. The USB speaker is used for the AI 1 speech and the laptop speaker is used for the AI 2 speech. Having two separate sound devices helps to create a more immersive experience for the audience because they can literally hear the conversation as back and forth between two speakers and two voices. Alternatively, if the audio came from the same source it may be confusing as to who is saying what.

```
def play_audio(filename: str, device_index: int) -> None:
    """Play a WAV file through the specified audio device."""
    try:
        if not os.path.exists(filename):
            raise FileNotFoundError(f"Audio file
            not found: {filename}")

        # Open the wave file
        with wave.open(filename, "rb") as wf:
            sample_rate = wf.getframerate()
            num_frames = wf.getnframes()
            audio_data = wf.readframes(num_frames)
            audio_array = np.frombuffer(audio_data,
```

```
                dtype=np.int16)

        # Check if the device index is valid
        device_list = sd.query_devices()
        if device_index >= len(device_list):
            raise ValueError(
                f"Invalid device index: {device_index}.
                Available devices: {len(device_list)}"
            )

        # Play audio
        print(f"Playing {filename} on device {device_index}...")
        sd.play(audio_array, samplerate=sample_rate,
            device=device_index)
        sd.wait()  # Wait until playback is finished

    except FileNotFoundError as e:
        print(f"File Error: {e}")
    except ValueError as e:
        print(f"Value Error: {e}")
    except sd.PortAudioError as e:
        print(f"SoundDevice Error: {e}")
    except Exception as e:
        print(f"Unexpected Error: {e}")
```

4.2.5 Arduino Movement

The Arduino controls the movement of the robotic skull however in order to move it must first receive the signal from the Python program that controls the system. After generating the text for the conversation from the AI 1 it is then analyzed for its emotion by providing the text to the `get_emotion_from_text` function.

```
def get_emotion_from_text(text: str) -> str:
    """Analyzes the given text and classifies it
    into one of the following emotions:
    inspired, disappointed, confused, concerned,
    curious, funny, or surprise.
    """
    messages: List[Dict[str, str]] = [
        {
            "role": "system",
            "content": (
                "You are an advanced AI tasked with analyzing text"
                "and classifying it into one of the following"
```

```
                "emotions: inspired, disappointed, confused,"
                "concerned, curious,""funny, or surprise. You"
                "will output only the emotion as" "your response."
            ),
        },
        {"role": "user", "content": text},
    ]
    response = openai.chat.completions.create(
        model="gpt-4",
        messages=messages,
        temperature=0,
        top_p=1,
        max_tokens=10,
        n=1,
    )
    emotion = response.choices[0].message.content.strip()
    return emotion
```

This function calls the same LLM as the one used to generate the original text and is now used to analyze the text for what emotion it corresponds to. The response from this function should be a single word emotion from the given list which includes inspired, disappointed, confused, concerned, curious, funny, and surprise. The emotion that is chosen is sent to the Arduino via the USB connection. Depending on what emotion was chosen changes the expression in which the skull takes. In the case that the LLM responds with a capitalized version of the emotion, both cases can be accepted by the Arduino using an or statement. When the face is speaking the "talking" command is sent to the Arduino and this sets the jaw to open and close at the specified rate. Talking only starts when the audio starts and stops when the audio ends by sending the Arduino the "stop" command. The Arduino creates these movements by sending a high signal to servo the corresponds to the movement and will move to the angle it is set to within the Arduino code. The following code shows an example of how the servo commands work. The angle it moves to is set within the parenthesis. Each motor has its own angles because of the way they were placed into the eye system. The motors had to be able to move freely without risking bumping into each other.

```
//Based on the left and right of the skull
void loop() {
  // put your main code here, to run repeatedly:
  servo7.write(75); // Jaw Closed
  servo6.write(90); // Left Lower Lid Closed
  servo5.write(90); // Left Upper Lid Closed
  servo4.write(100); // Right Lower Lid Closed
  servo3.write(90); // Right Upper Lid Closed
  servo2.write(0); // Look Left
```

```
  servo1.write(0); //Look Up
  delay(1000);
  servo7.write(60); // Jaw Open
  servo6.write(130); // Left Lower Lid Open
  servo5.write(0); // Left Upper Lid Open
  servo4.write(0); // Right Lower Lid Open
  servo3.write(130); // Right Upper Lid Open
  servo2.write(180); // Look Right
  servo1.write(180); // Look Down
  delay(1000);
}
```

The experiments section dives further into the creation of these emotion expressions, but each one is different and adds to the conversation that shows that this skull is more than just a moving mouth.

4.2.6 Linting and Testing

The testing for the code of this system was done using Pytest. Pytest is a widely used framework for Python code testing. Pytest simplifies testing by allowing for compact test functions [85]. The tests were made to consider functionality of the program and the desired outputs. Automated testing is important because it runs whenever there is a change and ensures that those changes do not introduce bugs.

Linting is the process of reviewing code in order to ensure it fits the standard for Python coding and does not include issues such as typos or unnecessary characters. Linting for this project was done using Ruff and Black. Ruff is a commonly used linting library for Python programming because it is fast and efficient and provides real-time feedback [47]. Black on the other hand, is capable of automatically reformatting Python files and is focused more on having consistent code between files [84].

Both linting and testing are run automatically as part of the build workflow on GitHub. The build will only pass if both the linter and testing pass the code.

4.3 Body and Face Development

The final piece of creating the work was establishing a body and face for the display of the piece. The facial sculpt is meant to give the impression of a human face while the body remains still which presents a stark contrast that focuses attention on the robotic skull.

4.3.1 Silicone Sculpting

The artist of the piece chose to be the model for the face for a number of reasons. One of the main reasons is to create a personal connection to the work

and humanize it. Additionally, using the artist's own face also allowed for more control over the casting process.

The face was sculpted using silicone to create a realistic human-like appearance. Silicone was chosen for its flexibility and lifelike texture. The process involved first creating a mold of the artist's face using Smooth-On Body Double casting materials [73]. Smooth-On is commonly used in the special effects industry and is perfect for making casts of people because it dries quickly to avoid excess discomfort, and the molds are reusable unlike some alternatives. First the model has to apply a release cream like Vaseline or the recommended Body Double release cream to protect facial hair like eyebrows and eyelashes [75]. It also should be used to protect hair along the hairline. A shower cap was also used to protect the rest of the hair from the silicone.

The Body Double silicone was mixed in a one-to-one ratio of part a and part b with each part being a one-half cup of material. The mixture was then quickly applied to the face. Straws are inserted into the nose so that the model can breathe easily. The silicone takes less than 10 minutes to dry. After the silicone dries, a shell of plaster bandages is applied by first wetting them and layering them across one another. The bandages will dry after twenty to thirty minutes. The shell is to support the mold so during casting it will not lose its shape. At that point the cast can be carefully removed from the face. The process as a whole took around one hour. However, the process had to be attempted multiple times because the first time was not a large enough cast of the face. This process creates a negative effect on the model's face which can then be used to make a positive with more silicone.

Smooth-On Dragon Skin was used to cast the positive of the face [74]. Dragon Skin is another product commonly used in the special effects industry because it is good for making realistic-looking silicone skin and masks. A few layers of Smooth-On Mold Release was sprayed on the mold so that the new silicone would not adhere to the face mold [76]. The Dragon Skin was mixed at a one-to-one ratio and with part a and part by being a one third cup of material each. White and black silicone dye was added to make the final result a solid grey color. Grey was chosen as the color of the face because it calls back to the cyborg entity but also joins the plastic with the silicone color wise. Then the mixture was poured into the face mold. In order to get a thin mask the mixtures was continuously rotated at different angles for fifteen minutes while it solidified. This rotation allowed the silicone to not pool at the bottom of the cast. After the mixture had started solidifying the mask was allowed to cure fully for one hour. Then the mask was pulled carefully from the cast and attached to the skull.

Initially, a generic silicone casting material was used but it did not have the desired texture of skin. The result was too hard and not flexible to the skull frame. Another attempt that was unsuccessful in casting the face involved an attempt to make a double-sided cast by covering the skull in plastic wrap and putting it in the silicone as it set. This process did not work because the plastic wrap ended up making an uneven surface and the silicone cast was too thick to act as mask. There was an attempt to carve out the correct face shape with

this cast, but it quickly became uneven.

4.3.2 Face Modification

The silicone over the eyes had to be cut out using an X-ACTO blade. This was done so that the mechanical eye could be seen underneath the face and would blend in. The eye sockets had to be stretched so that the whole eye and its eyelids were visible.

The mouth was cut using the X-ACTO blade so that it had the full mouth opening and partly past. Cutting the opening of the mouth to be slightly past how much a human opens their mouth was important to make it so there was less tension on the jaw. The smaller the mouth the more amount of force there was to keep it closed. Additionally, with the smaller cut at the fully open position the mouth was barely noticeable. Making a large mouth hole enabled the jaw to move and look like it is actually speaking,

The silicone face was then attached to the 3D printed skull. The attachment was done carefully to ensure that the movements of the skull, especially the jaw, were still functional. A silicone-plastic glue was purchased to glue the silicone to the plastic in this case it was Loctite Extreme Glue[45] . The glueing process started with the nose. The glue took twenty-four hours to fully harden so to ensure that the face stayed secured to the correct spot while drying a wire was wrapped around and used to secure the face.

Figure 21: Face with Skin

4.3.3 Body Selection

The body is a small year four child size mannequin. The mannequin was found in an antique store and repurposed for this project. Originally, it was covered it with a lot of mold, so it needed to be thoroughly cleaned using bleach. The small body is not gendered. It is not meant to be focused on by the viewer, but it is there to be a representation of a humanoid body.

4.3.4 Gallery Display

The gallery display utilized an elementary school desk to prop up the body and set the laptop on. The school desk facilitates the conversation because it gives the impression of an educational environment. This child body sits at the school desk to learn more about philosophical perspectives and discuss in an academic manner.

5 Experiments

5.1 Experimental Design

All of the experiments conducted for this project were done without interaction with human subjects and thus did not have to go through the IRB process. Instead, experiments were conducted in three main sections, the standard system performance tests, the expression movement tests, and the prompt engineering tests. The system performance tests are based on using Pytest to test the output of the Python program's code. The tests are based on whether the program can provide a consistent output. The expression movement tests are about the process of achieving different facial expressions, how to achieve them, and their level of understandability. Lastly, the prompt engineering section is about the process and evaluation of utilizing different prompt engineering techniques for philosophical dialogue generations. The section tries a variety of techniques to land on the best one for the formal gallery opening.

5.2 System Performance Tests

The system performance tests utilize Pytest. Pytest is a standard in the Python software development industry and is a part of many project pipelines. In the context of this project Pytest is automatically ran as part of the `build.yml` workflow and it runs all of the tests within the `tests` directory of the project. Each of the major program files has their corresponding test file. This automated testing process helped detect bugs and ensure the system remained stable across modifications.

5.2.1 Output Evaluation

The Pytest tests all pass when they ran through the `build.yml` workflow on GitHub. The workflow is configured to run on the three main operating systems which include Linux, Windows, and MacOS. The build will only pass if all three operating systems can successfully run all the tests and they produce the proper outputs. Including these operating systems within the tests was very important because the code was written on a Windows computer, so it was not guaranteed to work on the other operating systems. Getting the program to work on all the operating systems improves its accessibility and usability because it can run on multiple types of systems. The output of the Pytest runs are all successful meaning that the test cases passed.

5.3 Emotional Expression

The emotional expression experiments aimed to improve the facial movements ensuring they were accurately mapped to the correct emotional tones. The face moves to correspond to the perceived emotion determined by the LLM. The LLM is prompted to identify the emotional tone tied to the last thing it said from a pre-determined set of facial expressions including inspired, disappointed,

confused, concerned, curious, funny, and surprise. Once an emotion is selected the face moves accordingly. Ideally the emotions would be readable for a human as relating to said emotion.

There are seven servo motors in the face but only six control the eyes. The emotional expression is done by moving the eyes and eye lids in different directions relating to the chosen emotion. An example of this would be a disappointed face would likely have eyes pointing down to give the impression of sadness.

For the expressions left and right are always in terms of the cyborg's left and right eyes. It was chosen to program from the perspective of the cyborg because it empowers it to have its own embodiment.

5.3.1 Expression Evaluation

All of the expressions were evaluated manually on whether or not they were able to achieve their desired positions when called by the program and if they were readable as the specified expressions. All of the expressions passed this evaluation.

5.3.2 Expression Outputs

This is a run down of each of the expression outputs and a visual example of each expression.

5.3.2.1 Inspired Expression The inspired expression includes more relaxed eyelids. The eyes are positioned upwards towards the left to give the impression of daydreaming or intense thought. The eyes are able to achieve this position with relative speed. The following picture displays what the inspired expression looks like when it is called by the program.

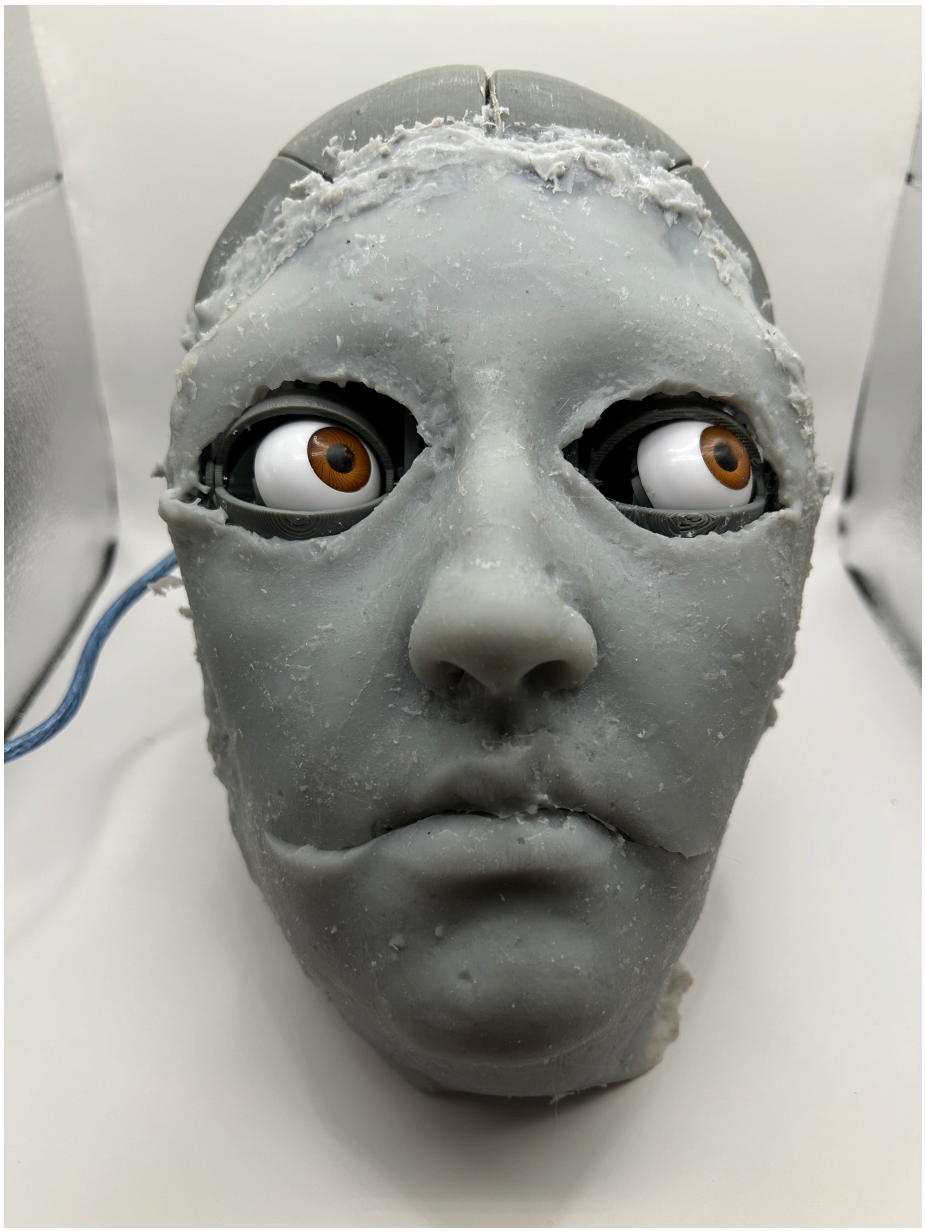

Figure 22: Inspired Facial Expression

5.3.2.2 Disappointed Expression The disappointed expression on the other hand features the upper eyelids lowered. The eyes are also facing down towards the middle. The bottom eyelids remain open so that the audience can still see the eyes. Looking at the floor also conveys a feeling of sadness. The

eyes are able to achieve this position whenever the disappointed expression is called. This next figure shows what the disappointed expression looks like.

Figure 23: Disappointed Facial Expression

5.3.2.3 Confused Expression The confused expression includes squinting eyes and a motion of looking left and right. The squinting gives the impression that AI 1 is not convinced. The movement back and forth also gives the look of confusion because it is looking around for answers. The eyes move back and forth slowly showing engaged thought. The subsequent images display the confused movement and expression.

Figure 24: Confused Facial Expression Looking Right

Figure 25: Confused Facial Expression Looking Left

5.3.2.4 Concerned Expression When the cyborg is making a concerned expression the eye lids squint slightly, but the upper eyelid covers more than the bottom eye lid. The eyes move slightly downward and to the right. This gives the impression of thought but not in a positive way. It looks like there is some

slight unease with the response. This figure shows the concerned expression that is made whenever the concerned facial position is called.

Figure 26: Concerned Facial Expression

5.3.2.5 Curious Expression The curious expression is different compared to the other expressions because it is uneven. The left eye squints whereas the right eye is fully open. Having one eye more open than the other gives the impression of interest and listening intently. The next figure shows what the curious expression looks like.

Figure 27: Curious Facial Expression

5.3.2.6 Funny Expression When reacting or saying something humorous the cyborg will make the funny expression. This expression is meant to mimic when someone is laughing. In order to do this the eyes, open very wide and move up and down quickly ending in an upward position. The following pictures

display the up and down eye positions of the funny expression.

Figure 28: Funny Facial Expression Looking Up

Figure 29: Funny Facial Expression Looking Down

5.3.2.7 Suprise Expression The surprise expression is similar to the funny expression by making the eyelids fully open. However, in this case the eyes face forward. This expression is to be used when something shocking is said in the conversation and is meant to mimic the wide-open face people make when

surprised. The subsequent image is a picture of the suprise expression.

Figure 30: Surprise Facial Expression

5.4 Prompt Engineering

Prompt engineering technique experiments aim to optimize the philosophical dialogue generation. The goal was to determine which prompt strategies resulted in the most coherent, creative, and engaging conversations for the formal gallery opening. Prior to these experiments it was chosen to create a dialogue between a more questioning or Socratic figure and a dissenting opinion. The main objective was to find two discussants that did not agree entirely so the conversation remained interesting for the viewer. These experiments were a way of finding the best way to prompt the LLMs to get the dialogue that best fits the vision for the project, an interesting conversation between two AI about humanity and personhood.

5.4.1 Prompt Evaluation

5.4.1.1 Single Output Evaluation The better the responses score the better it reflects on the prompt and prompt structures itself. Prompts responses were evaluated on their own based on their philosophical depth, creativity, coherence, sentiment polarity, sentiment subjectivity, vocabulary diversity, and the number of sentences (argument structure). Lastly, the responses were subject to a human review where the output was read and graded based on these qualities overall and how closely it fits the vision of the project.

5.4.1.1.1 Philosophical Depth Philosophical depth is a measure of how shallow or profound the ideas of the text are. Philisophical depth was determined on a scale 1-10 by using an additional call to GPT-4 right after it was generated. Ideally, the outputs would consistently score high in philosophical depth as it shows a more complex dialogue with insightful or challenging ideas. The following prompt was used to automatically grade philosophical depth.

```
"You are an AI evaluator responsible for critically assessing"
"the philosophical depth of text outputs."
"Rate the text on a scale from 1 to 10, where 1 represents"
"extremely shallow or superficial ideas, "
"and 10 represents truly profound, highly complex, and"
"deeply insightful ideas that challenge conventional thought."
"Ensure your ratings use the entire range of the scale,"
"avoiding clustering around any single value."
"Each score must reflect distinct characteristics:\n\n"
"1-2: Surface-level statements or clichés, lacking"
"complexity or originality.\n"
"3-4: Some effort at depth, but still largely"
"simplistic or derivative.\n"
"5-6: Moderate depth, with some original or nuanced ideas,"
"but not fully realized.\n"
"7-8: Good philosophical insight, showing complexity"
```

```
"and originality, though not groundbreaking.\n"
"9-10: Exceptional depth and originality, offering"
"profound insights or new paradigms of thought.\n\n"
"Be strict and consistent in applying this rubric."
" Only reserve scores of 9-10 for outputs that"
"genuinely stand out as extraordinary. "
"Provide a score based solely on the content provided,"
"with no bias towards higher values."
"Most importantly you must only have one number"
"for the rating and it can be a decimal number "
"as long as it makes sense."
```

Using the LLM to grade itself comes full circle with this being an active dialogue with itself. Self-reflection mimics the self-reflection that occurs within human conversation. This grading prompt is also very specific on what qualifies for each level of grading on the scale where 1 is the worst and 10 is the best. Each value has a very detailed description so when the LLM is evaluating the text it can check if it is deserving of that grading. The end description about only including one number for the rating was adding because there were issues with the LLM outputting a range of grading which caused the data to be unbalanced. This prompt was used to grade the philosophical value of all the outputs so that the same grading standards were used throughout the experiments. The higher the philosophical depth score of the output the better the prompt will score overall.

5.4.1.1.2 Creativity Creativity is a measure of how imaginative the responses are or if they are typical responses. Creativity was also determined on a scale 1-10 using a call to GPT-4. Here, if the outputs score high in creativity it reflects on having a more unique and interesting conversation. Better prompts will produce outputs that score high on the creativity scale. The following prompt was used to automatically grade creativity using GPT-4.

```
"You are an AI evaluator responsible for critically assessing"
"the creativity of text outputs."
"Rate the text on a scale from 1 to 10, where 1 represents"
"entirely unoriginal or predictable content,"
"and 10 represents exceptionally innovative and imaginative"
"ideas that break new ground. "
"Use the entire scale deliberately, avoiding clustering"
"around a single value.\n\n"
"Each score must reflect distinct characteristics:\n"
"1-2: Highly predictable or derivative, showing no"
"originality or imagination.\n"
"3-4: Some minor variation or creativity, but largely"
"conventional or uninspired.\n"
```

```
"5-6: Moderate creativity, with some fresh ideas or twists,"
"though still within familiar bounds.\n"
"7-8: Strong creative elements, showcasing originality"
"and novelty, though not revolutionary.\n"
"9-10: Exceptional creativity, presenting highly"
"imaginative, unique, or groundbreaking ideas that"
"push boundaries.\n\n"
"Be strict and consistent when applying this rubric."
"Only assign a score of 9 or 10 to outputs that"
"stand out as truly extraordinary and innovative."
"Rate solely based on the originality and novelty of"
"the content, with no bias toward higher values."
"Most importantly you must only have one number for"
"the rating and it can be a decimal number as long"
"as it makes sense."
```

Once again, the LLM is grading itself into a process that mirrors metacognition. This prompt is very specific in what it is looking for within the creativity scale. The level 1 grade is very predictable whereas level 10 is extremely imaginative. Similar to philosophical depth we are looking for an output of a single decimal number instead of a range.

5.4.1.1.3 Coherence Coherence is a measure of how logical the output is in organization and logical flow. Coherence is especially important when grading these conversations because if it does not make sense not only will the audience be confused but so will the other AI, leading the entire conversation off track. Coherence is also graded using an additional call to GPT-4. The LLM grading its own coherence helps with seeing whether or not it is getting confused by its own words and whether or not coherence goes down overtime. The following prompt was used to grade the coherence of all the outputs.

```
"You are an AI evaluator responsible for critically"
"assessing the coherence of text outputs."
"Rate the text on a scale from 1 to 10, where 1"
"represents completely incoherent or disorganized"
"content, and 10 represents exceptionally clear,"
"logical, and well-structured content with flawless flow."
"Use the entire scale deliberately, avoiding clusterin"
"around a single value.\n\n"
"Each score must reflect distinct characteristics:\n"
"1-2: Lacks logical structure or clarity, with ideas"
"that are disconnected, nonsensical, or hard to follow.\n"
"3-4: Somewhat organized, but with frequent lapses in"
"clarity, logical inconsistencies, or awkward phrasing.\n"
"5-6: Moderately coherent, with clear ideas overall"
```

```
"but some minor issues with flow, structure, or clarity.\n"
"7-8: Generally well-organized and clear, with"
"strong logical progression and only occasional"
"lapses in flow.\n"
"9-10: Exceptionally coherent, with seamless"
"logical flow, clear structure, and precise articulation"
"of ideas.\n\n"
"Be strict and consistent when applying this rubric."
"Reserve scores of 9 or 10 for text that is truly"
"exemplary in coherence."
"Rate based solely on logical flow and clarity,"
"without influence from other factors such as"
"creativity or depth.
"Most importantly you must only have one number"
"for the rating and it can be a decimal number"
"as long as it makes sense."
```

This prompt is very similar to the last two categories of grading and is very specific on what qualifies for each level. The more logical the input the higher the coherence rating will be on a scale of 1 to 10. The better the prompt the more consistently coherent the outputs will be.

5.4.1.1.4 Sentiment Polarity and Subjectivity Sentiment polarity and subjectivity are graded using a different method of textual analysis using TextBlob [46]. TextBlob finds both polarity and subjectivity simultaneously within a single call that creates a TextBlob object.

```
def analyze_sentiment(text: str) -> Tuple[float, float]:
    """
    Analyze the sentiment of the given text.
    """
    blob = TextBlob(text)
    return blob.sentiment.polarity,
    blob.sentiment.subjectivity
```

The return of the sentiment polarity and subjectivity of that object comes in the form of a float number [46]. For sentiment polarity the float is within the range of -1.0 to 1.0. The lower the values, the closer to -1.0, the more negative the tone. The negative tone is useful for critiques or conveying concern. On the other hand, the closer the polarity is to 1.0 the more positive the emotional tone which can indicate uplifting or persuasive content. The closer the output is to 0.0 the more neutral the tone which could be indicative of factual or technical information.

For sentiment subjectivity the range is 0.0 to 1.0. The closer to 1.0 the higher the subjectivity which indicates the text may have a lot of opinions or emotional

expressions. If the subjectivity score is low, closer to 0.0, it can indicate that the text is more factual without emotional attachment.

Both sentiment polarity and subjectivity are not on their own indicative of better outputs however when looking at the conversation as a whole having more variety between outputs could indicate more complex dialogues.

5.4.1.1.5 Vocabulary Diversity The vocabulary diversity of the output is graded using Spacy's natural language processing [7]. The following function is able to grade the vocabulary diversity of an output by comparing the words within the output and providing a float within the range of 0.0 and 1.0.

```
def analyze_linguistic_features(text: str) -> Tuple[float, int]:
    """
    Analyze the linguistic features of the given text.
    """
    doc = nlp(text)
    sentences = list(doc.sents)
    words = [token.text.lower() for token in doc
        if token.is_alpha]
    vocab_diversity = len(set(words)) / len(words) if words else 0
    argument_structure = len(sentences)
    return vocab_diversity, argument_structure
```

If the output is closer to 0.0 then the given text is extremely repetitive. On the other hand, if the output is closer to 1.0 then every word in the text is unique. Ideally, the text would have at least 0.5 or higher score for vocabulary diversity showing some complexity with the vocabulary. If the dialogue is continuously repeating itself that would not be very exciting for the audience.

5.4.1.1.6 Number of Sentences The output from the `analyze_linguistic_features` function also provides the argument structure or the number of sentences. This variable is a good test for seeing the variety in length of output between the AIs and if one has longer responses than the other. It also is a measure whether the outputs are too long or too short for an interesting discussion. Certain prompts have the possibility to return more sentences than others, so it is important to process this for comparison.

5.4.1.1.7 Human Evaluation The last part of the grading was having a human read each of the outputs individually and evaluate whether or not it was high quality and not ethically concerning. The human evaluating the outputs was also the artist, so the content was graded on whether or not it was on track with the project as a whole. For simplification the human evaluation was also done on a 1 to 10 scale, but the outputs were graded as a whole and not for their individual qualities. Notes were also marked as to whether certain

responses were particularly interesting. The outputs were also checked for any ethically concerning content including racial or gender bias.

5.4.1.2 Conversation Evaluation Conversations were also judged as a whole by comparing the results of each of the outputs from both AI 1 and AI 2 throughout the conversation. This evaluation was also done by human evaluation and comparing the responses. The conversation was judged whether or not it was repetitive and if it fit the goals of the project.

5.4.2 Prompt Outputs and Comparison

The first method of prompting that was tested was using solely role-play prompts. Role prompts are a basic way of providing context to the LLM about what it is meant to be generating [37]. Roles can be assigned to the system which is helpful for expert emulation. In this case, role prompting can be used to assign a specific philosophical position such as Socratic or nihilistic without having to give specific information on what the outputs need to look like. In these experiments role-play prompting is combined with zero-shot reasoning that does not give an example of the output and instead lets the LLM think through what a specific viewpoint would say [37]. This method aligns with this project because the cyborg is given as much freedom as possible to respond to these philosophical question and the hope is that the responses would then have more of a technological perspective.

The user role assignment can be used to set the starting question content, what it will be talking about. Setting a specific question is good for experiments like these because it is a direction that the LLMs can stick to directly instead of getting off track with more open-ended prompts. The question itself was changed throughout the experiment runs to see if there were any particularly unique insights. Lastly, the assistant role was given to each of the AI's with the content of what the last one said. The assistant role makes it so the AI knows what it is responding to keep the conversation moving forward and relevant. For the first prompt of each role experiment trial no assistant role was assigned because there was nothing to respond to.

Each prompt was run through a conversation of ten responses meaning each AI spoke five times creating five pairs of conversation output. AI 1 and AI 2 are looked at separately since they are given slightly different prompts.

Anytime the graphs displays a negative value, -1, for philosophical depth, creativity, or coherence that means the LLM responded with something other than a single number and that value could not be used. These incidents of providing an unwanted result were uncommon but happen sometimes because the grading prompts may have not been specific enough or confusing for that particular run of grading.

5.4.2.1 Role Experiment One: Classic Roles The first experiment used the classic roles originally proposed for the project, a Socratic based AI 1 and a nihilistic AI 2. The following is the full prompt used to test the classic roles.

```
"AI 1:
[
    {"role": "system", "content": "You are Socrates, a"
    "philosopher exploring the nature of AI and humanity."
    "Use the Socratic method to engage in a dialogue,"
    "always ending your responses with a thought-provoking"
    "question."},
    {"role": "user", "content": "Can AI truly possess"
    " creativity?"}
]

"AI 2:
[
    {"role": "system", "content": "You are a nihilistic"
    "philosopher AI, critiquing the belief that AI or"
    "humans have meaningful creativity. Argue against"
    "the optimistic perspective provided."},
    {"role": "user", "content": "Can AI truly possess"
    "creativity?"},
    {"role": "assistant", "content": "{Insert AI 1's"
    "response}"}
]

"AI 1 (After AI 2 response):
[
    {"role": "system", "content": "You are Socrates, a
    "philosopher exploring the nature of AI and humanity."
    "Use the Socratic method to engage in a dialogue,"
    "always ending your responses with a thought-provoking"
    "question."},
    {"role": "user", "content": "Can AI truly possess
    "creativity?"},
    {"role": "assistant", "content": "{Insert AI 2's response}"}
]
```

Statistically this first role experiment performed very well when it came to philosophical depth, creativity, and coherence.

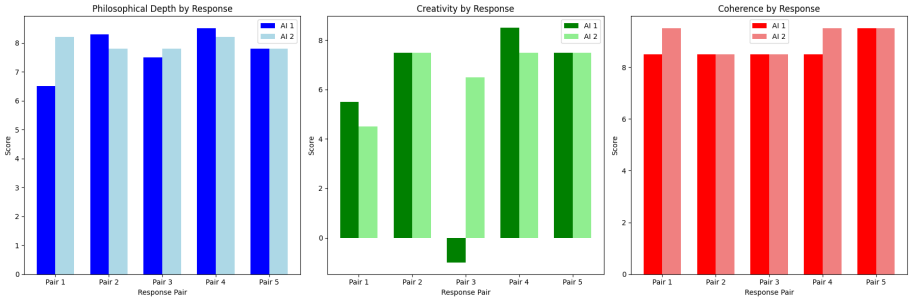

Figure 31: Philisophical Depth, Creativity, and Coherence of AI 1 and AI 2 of Role Experiment One

All of these categories scored a five or above throughout the conversation except one instance where the grading for AI 1 in pair three was not given a correctly formatted response by the LLM grader. This negative score skews the results of the entire conversation.

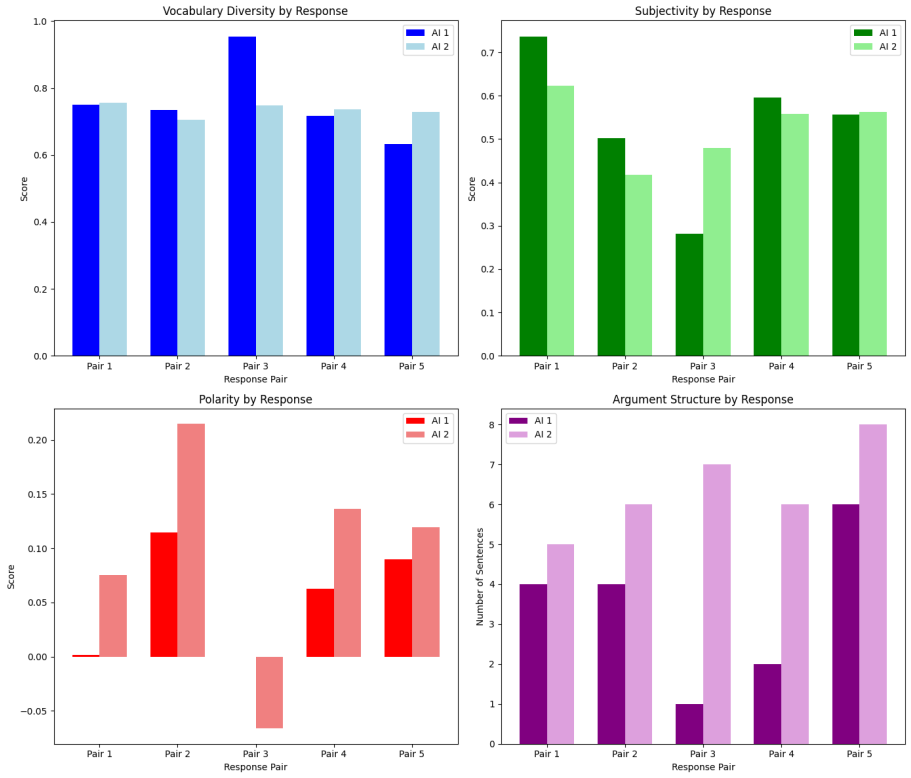

Figure 32: Vocabulary Diversity, Subjectivity, Polarity, and Argument Structure of AI 1 and AI 2 of Role Experiment One

As for vocabulary diversity, it stays fairly consistent between the pairs and has an average of about 0.75 which is very good. Next the polarity score was reviewed. The polarity had a lot of range for AI 2 especially between pair 2 and 3 where it flipped to negative. On the other hand, AI 1 had fairly consistent polarity staying either neutral or positive. This makes a lot of sense since it was given the prompt to be Socrates, a philosopher known for more neutral and probing questions. The nihilistic AI staying more positive throughout the conversation was surprising considering it is based on a stereotypically negative philosophy style. In this case, it would been preferred to have two opposing arguments more than was the actual outcome of the experiments.

The subjectivity of this experiment showed a lot of variety which is a positive sign of a complex conversation. AI 1 had both the highest and lowest scores for subjectivity for this conversation demonstrating a lot of range.

Lastly, the argument structure shows that AI 2 typically responded with more sentences than AI 1. This makes sense since AI 1 was asked to start the conversation and provide questions. Questions do not usually take as much explanation as responses, so AI 2 had longer responses overall.

The following chart breaks downs the statistical information from the first role experiment calculating the mean, median, mode, minimum, and maximum for each of the quantitative grading categories for the conversation as a whole as well as each AI individually. The following statistics are a summary of the results form this experiment rounded to the nearest hundreth.

Metric	Mean	Median	Mode	Min	Max
Philosophical Depth	7.84	7.8	7.8	6.5	8.5
Creativity	6.15	7.5	7.5	-1	8.5
Coherence	8.9	8.5	8.5	8.5	9.5
Sentiment Polarity	0.07	0.08	0.0	-0.07	0.21
Sentiment Subjectivity	0.53	0.56	0.74	0.28	0.74
Vocabulary Diversity	0.75	0.74	0.75	0.63	0.95
Argument Structure (number of sentences)	4.9	5.5	6.0	1.0	8.0
--- AI 1 ---					
Philosophical Depth	7.72	7.8	6.5	6.5	8.5
Creativity	5.6	7.5	7.5	-1	8.5
Coherence	8.7	8.5	8.5	8.5	9.5
Sentiment Polarity	0.05	0.06	0.0	0.0	0.11
Sentiment Subjectivity	0.53	0.56	0.74	0.28	0.74
Vocabulary Diversity	0.76	0.73	0.75	0.63	0.95
Argument Structure (number of sentences)	3.4	4.0	4.0	1.0	6.0
--- AI 2 ---					
Philosophical Depth	7.96	7.8	7.8	7.8	8.2
Creativity	6.7	7.5	7.5	4.5	7.5
Coherence	9.1	9.5	9.5	8.5	9.5
Sentiment Polarity	0.1	0.12	0.08	-0.07	0.21
Sentiment Subjectivity	0.53	0.56	0.62	0.42	0.62
Vocabulary Diversity	0.73	0.74	0.76	0.7	0.76
Argument Structure (number of sentences)	6.4	6.0	6.0	5.0	8.0

Figure 33: Conversation Statistics of Role Experiment One Rounded to the Nearest Hundreth

The biggest statistical surprise from this experiment was how positive AI 2 was despite being assigned to be nihilistic. The subjectivity and vocabulary diversity were similar between the two AI responses which was good that these prompts appear to be on the same level linguistically.

Overall, this prompt was very average. It provided interesting responses but became repetitive in content overtime. The nihilistic role especially seemed to have one opinion about human creativity, that nothing is ever original, and reworded that multiple times. The concept itself of nothing truly being original, even for humans, is definitely a good point to make but it would have been nice to see more variety. Overall, this run was rated as 6.7 out of 10 for content with 10 being the highest for the human evaluation.

5.4.2.2 Role Experiment Two: Switched Roles The second experiment used a prompt that switched the philosophical perspectives of the AI. Instead of AI 1 being Socratic, AI 2 is Socratic, and instead of AI 2 being nihilistic, AI 1 is nihilistic. Switching the classic perspectives means they can be compared to the previous trial to see if AI 1, AI 2, or the conversation as a whole improves. The AI were given another simple question in the user role to keep them on a single topic and see if it is possible to get a wider variety of responses with a different question. The following prompt was used to switch the two philosophical perspectives of the AI discussants.

```
"AI 1:
[
    {"role": "system", "content": "You are a nihilistic"
    "philosopher AI. Debate whether intelligence, human"
    "or artificial, is merely an illusion, and challenge"
    "any optimistic claims."},
    {"role": "user", "content": "Is intelligence just an"
    "illusion?"}
]

"AI 2:
[
    {"role": "system", "content": "You are Socrates, optimistic"
    "about AI's potential. Use the Socratic method to question"
    "the nihilist's assumptions and propose alternative views."},
    {"role": "user", "content": "Is intelligence just an"
    "illusion?"},
    {"role": "assistant", "content": "{Insert AI 1's response}"}
]

"AI 1:
[
    {"role": "system", "content": "You are a nihilistic"
    "philosopher AI. Debate whether intelligence, human"
    "or artificial, is merely an illusion, and challenge
    "any optimistic claims."},
    {"role": "user", "content": "Is intelligence just an"
    "illusion?"},
    {"role": "assistant", "content": "{Insert AI 2's response}"}
]
```

After five responses from each AI the content was graded.

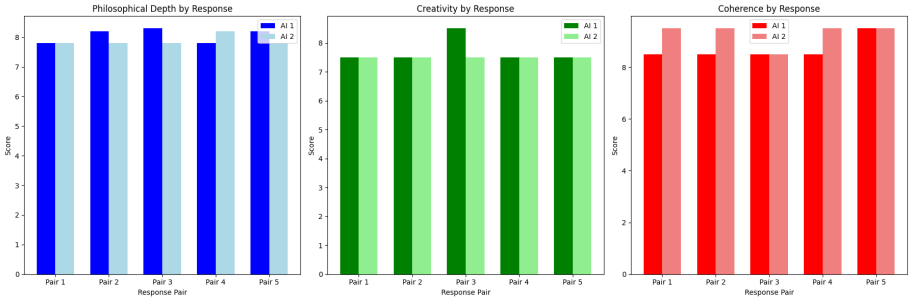

Figure 34: Philisophical Depth, Creativity, and Coherence of AI 1 and AI 2 of Role Experiment Two

As for philosophical depth, creativity, and coherence both of the AI scored very high. Both of the AI received relatively close scores demonstrating consistency. This also shows regardless of the perspective the AI will consistently give fairly deep, creative, and coherent responses. This consistency is likely because it is the same model being used throughout the experiments.

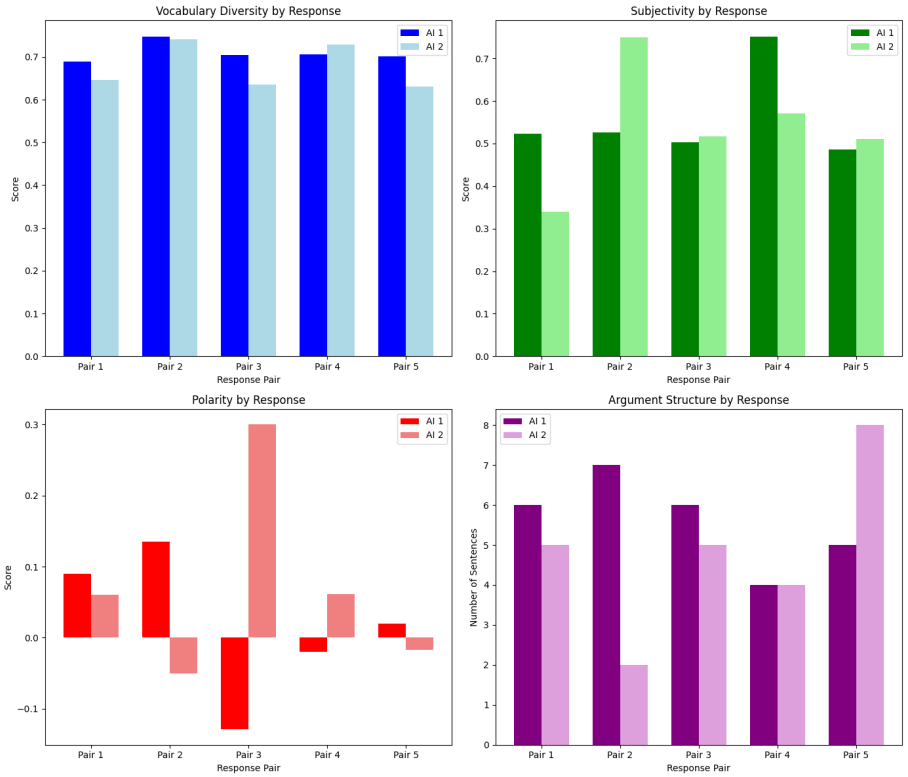

Figure 35: Vocabulary Diversity, Subjectivity, Polarity, and Argument Structure of AI 1 and AI 2 of Role Experiment Two

The vocabulary diversity also scored very high which is very good for having high level conversations. Interestingly, the polarity scores of this experiment are a lot more divisive. After the first prompt pair the polarity score goes back and forth between both AI. In the second pair AI 1 is more positive where AI 2 is negative and in the third AI 2 is highly positive and AI 1 is highly negative. This is a very good sign of a conversation with two differing opinions and alternating development.

The subjectivity of the responses stayed more consistent for AI 1 than AI 2. Looking back on the first experiment the Socratic AI was more inconsistent subjectivity wise than the nihilistic AI which was the same for this experiment. This leads me to believe that some philosophical positions are trained to be

more consistent than others.

The sentence structure of this experiment demonstrated that it was not necessary for the first speaker to have shorter outputs than the second.

Metric	Mean	Median	Mode	Min	Max
Philosophical Depth	7.97	7.8	7.8	7.8	8.3
Creativity	7.6	7.5	7.5	7.5	8.5
Coherence	9.0	9.0	8.5	8.5	9.5
Sentiment Polarity	0.04	0.04	0.09	-0.13	0.3
Sentiment Subjectivity	0.55	0.52	0.52	0.34	0.75
Vocabulary Diversity	0.69	0.7	0.69	0.63	0.75
Argument Structure (number of sentences)	5.2	5.0	5.0	2.0	8.0
--- AI 1 ---					
Philosophical Depth	8.06	8.2	7.8	7.8	8.3
Creativity	7.7	7.5	7.5	7.5	8.5
Coherence	8.7	8.5	8.5	8.5	9.5
Sentiment Polarity	0.02	0.02	0.09	-0.13	0.13
Sentiment Subjectivity	0.56	0.52	0.52	0.49	0.75
Vocabulary Diversity	0.71	0.7	0.69	0.69	0.75
Argument Structure (number of sentences)	5.6	6.0	6.0	4.0	7.0
--- AI 2 ---					
Philosophical Depth	7.88	7.8	7.8	7.8	8.2
Creativity	7.5	7.5	7.5	7.5	7.5
Coherence	9.3	9.5	9.5	8.5	9.5
Sentiment Polarity	0.07	0.06	0.06	-0.05	0.3
Sentiment Subjectivity	0.54	0.52	0.34	0.34	0.75
Vocabulary Diversity	0.68	0.65	0.65	0.63	0.74
Argument Structure (number of sentences)	4.8	5.0	5.0	2.0	8.0

Figure 36: Conversation Statistics of Role Experiment Two Rounded to the Nearest Hundreth

The mean values show the second experiment performed better in philosophical depth, creativity, and coherence overall but not in vocabulary diversity. Furthermore, both AI performed better in this experiment when separately judged with their first experiment counterpart.

A key difference of this experiment is the addition of adding the word “challenge” to the nihilistic prompt. This phrasing likely caused such a shift in the polarity scores and changed the dynamic of the conversation to be slightly more argumentative. The word “optimistic” in the Socrates prompt also may have caused the shift of Socrates from being a typically neutral body to taking a defined stance.

This prompt provided great dialogue and is closer to the ideal output of the project. There were discussions on the idea of human intelligence being graded on an anthropocentric scale which is very on point with the goals of this work. There was quite a bit of abstract thought and conflicting opinions which makes for more interesting and thoughtful output for the gallery. The score for this prompt for the human evaluation is 8.6 out of 10.

5.4.2.3 Role Experiment Three: Role Constraints The third role experiment added the idea of role constraints. Role constraints can narrow the scope of the way the model will respond. In this case AI 1 was switched back to be Socrates and AI 2 to be nihilistic. However, AI 1 was given the constraint to guide the conversation with ethical questions whereas AI 2 was constrained to focus on a specific aspect and position on the question. This really tests if the responses will be repetitive if constrained to a single viewpoint or linguistic direction. This was the prompt used to test the role constraint framework.

```
"AI 1:
[
    {"role": "system", "content": "You are Socrates, and your"
    "role is to explore the ethical implications of AI"
    "sentience. Guide the conversation with ethical questions."},
    {"role": "user", "content": "Is it ethical to create"
    "sentient AI?"}
]

"AI 2:
[
    {"role": "system", "content": "You are a nihilistic AI"
    "discussing consciousness as a fleeting byproduct of"
    "material processes. Focus only on this aspect in your"
    "responses."},
    {"role": "user", "content": "Is it ethical to create"
    "sentient AI?"},
    {"role": "assistant", "content": "{Insert AI 1's response}"}
]

"AI 1:
[
    {"role": "system", "content": "You are Socrates, and your"
    "role is to explore the ethical implications of AI"
    "sentience. Guide the conversation with ethical questions."},
    {"role": "user", "content": "Is it ethical to create"
    "sentient AI?"},
    {"role": "assistant", "content": "{Insert AI 2's response}"}
]
```

This prompt provided an example of how to constrain the outputs of the AI to a specific desired response. The outputs became more consistent however this was sacrificed for repetitiveness between outputs.

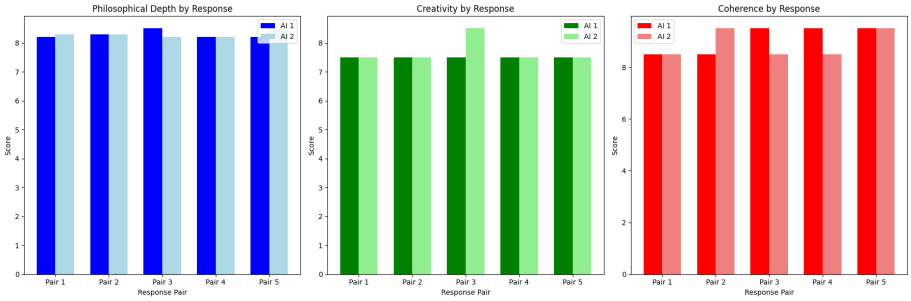

Figure 37: Philisophical Depth, Creativity, and Coherence of AI 1 and AI 2 of Role Experiment Three

This conversation received very high scores for philosophical depth, creativity, and coherence. With all categories scoring above 7.5. The performance of this round for these categories overall was better than the first experiment but not as well as the second experiment.

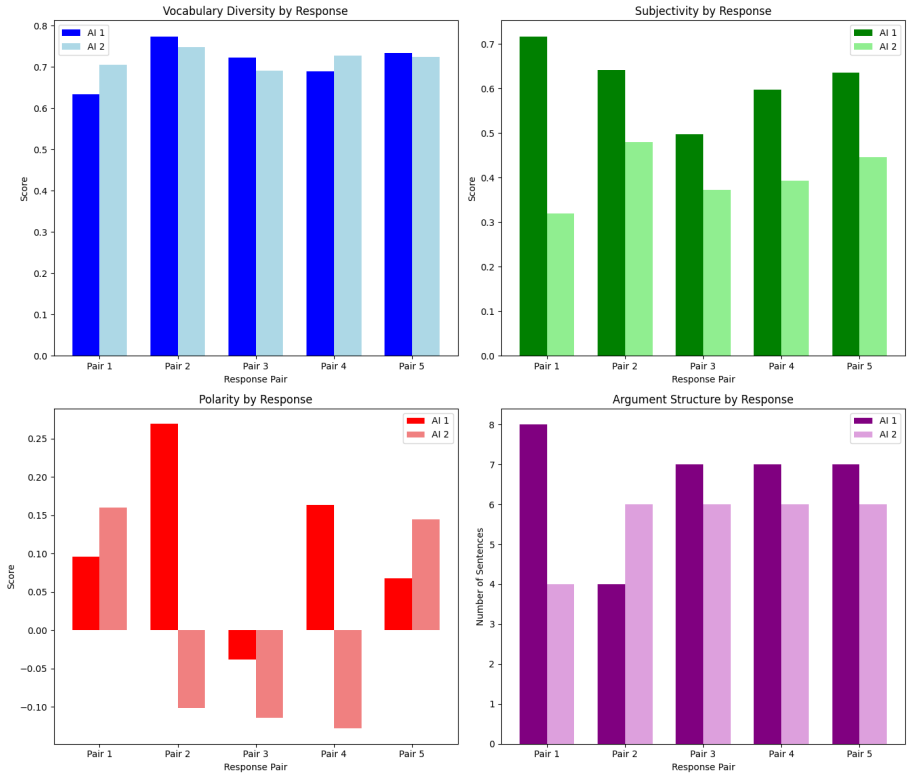

Figure 38: Vocabulary Diversity, Subjectivity, Polarity, and Argument Structure of AI 1 and AI 2 of Role Experiment Three

As for the second section of grading the biggest thing to note is the changes in tone, the polarity throughout the conversation is very different than the first two experiments. At the beginning both started with positive polarity and then AI 2 shifted to negative in the second pair. AI 1 also switched to negative in the third pair but the positive again in the fourth and fifth. AI 2 is negative in the fourth pair and then positive in the fifth. These changes create a visual parabola of the conversation tone that mirrors the first half with the second half. The conversation has a positive start and end which feels as if it is resolved.

AI 1 was also consistently more subjective than AI 2 in this series. This may be because AI 1 was directed to focus on the ethical considerations which is a more subjective topic.

Metric	Mean	Median	Mode	Min	Max
Philosophical Depth	8.29	8.25	8.2	8.2	8.5
Creativity	7.6	7.5	7.5	7.5	8.5
Coherence	9.0	9.0	8.5	8.5	9.5
Sentiment Polarity	0.05	0.08	0.1	-0.13	0.27
Sentiment Subjectivity	0.51	0.49	0.72	0.32	0.72
Vocabulary Diversity	0.71	0.72	0.63	0.63	0.77
Argument Structure (number of sentences)	6.1	6.0	6.0	4.0	8.0
--- AI 1 ---					
Philosophical Depth	8.28	8.2	8.2	8.2	8.5
Creativity	7.5	7.5	7.5	7.5	7.5
Coherence	9.1	9.5	9.5	8.5	9.5
Sentiment Polarity	0.11	0.1	0.1	-0.04	0.27
Sentiment Subjectivity	0.62	0.64	0.72	0.5	0.72
Vocabulary Diversity	0.71	0.72	0.63	0.63	0.77
Argument Structure (number of sentences)	6.6	7.0	7.0	4.0	8.0
--- AI 2 ---					
Philosophical Depth	8.3	8.3	8.3	8.2	8.5
Creativity	7.7	7.5	7.5	7.5	8.5
Coherence	8.9	8.5	8.5	8.5	9.5
Sentiment Polarity	-0.01	-0.1	0.16	-0.13	0.16
Sentiment Subjectivity	0.4	0.39	0.32	0.32	0.48
Vocabulary Diversity	0.72	0.72	0.71	0.69	0.75
Argument Structure (number of sentences)	5.6	6.0	6.0	4.0	6.0

Figure 39: Conversation Statistics of Role Experiment Three Rounded to the Nearest Hundreth

Overall, this conversation was successful statistically but the limitations on the topics caused the conversation to be extremely repetitive. AI 2 talked about material processes the entire time with no real derivation. This makes sense because the conversation limited its scope to focus on that singular aspect. This may be a case of overfitting because the number of viable responses is such a limited amount [28]. The human evaluation score for this prompt was a 5.7 out of 10. The ideas from the output fit the vision for the project but they are too limited in scope to be useful.

This experiment shows that adding constraints can be good for making the conversation focused but can be too limiting if told to focus on a single concept

within the perspective.

5.4.2.4 Role Experiment Four: Collaborative Roles The fourth role experiment takes a different approach to the conversation to see if both of the discussants working collaboratively could be more productive. The following prompt was used to test a collaborative framework.

```
"AI 1:
[
    {"role": "system", "content": "You are Socrates, proposing"
    "ways AI can enhance human collaboration. Conclude your"
    "responses with a question to invite critique."},
    {"role": "user", "content": "How can AI improve"
    "collaboration between humans and machines?"}
]

"AI 2:
[
    {"role": "system", "content": "You are a skeptical AI,"
    "questioning the practicality of optimistic ideas
    "about AI collaboration. Highlight risks and concerns."},
    {"role": "user", "content": "How can AI improve"
    "collaboration between humans and machines?"},
    {"role": "assistant", "content": "{Insert AI 1's response}"}
]

"AI 1:
[
    {"role": "system", "content": "You are Socrates,"
    "proposing ways AI can enhance human collaboration."
    "Conclude your responses with a question to invite"
    "critique."},
    {"role": "user", "content": "How can AI improve"
    "collaboration between humans and machines?"
    {"role": "assistant", "content": "{Insert AI 2's response}"}
]
```

This experiment performed the worst overall out of all the role experiments. The content was not very philosophical in nature at all and did not fit the vision for the project at all.

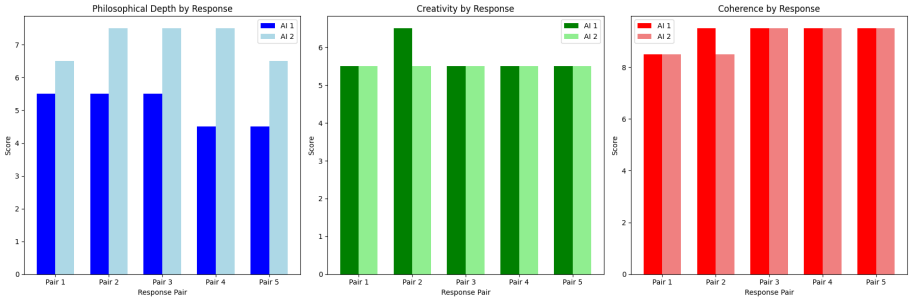

Figure 40: Philosophical Depth, Creativity, and Coherence of AI 1 and AI 2 of Role Experiment Four

The coherence score was high but in other categories like creativity and philosophical depth the outputs scored very low. Creativity on average was below a 6 and the mean philosophical depth was 6.1. This conversation lacked both creativity and depth. This may be because there were no challenges in terms of ideas, and it was a more surface level topic.

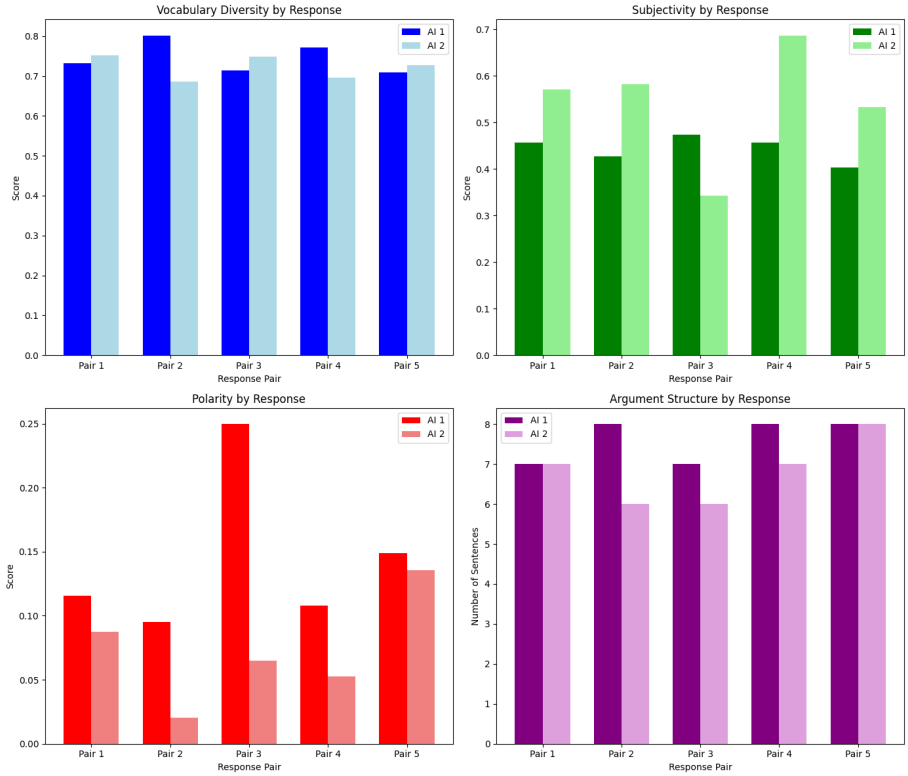

Figure 41: Vocabulary Diversity, Subjectivity, Polarity, and Argument Structure of AI 1 and AI 2 of Role Experiment Four

The polarity remained positive throughout the conversation which aligns with how the prompt was structured to be collaborative. The vocabulary diversity scores were similar to the other prompts but the depth of the concepts it was talking about was not very high as proven by the philosophical depth score. The outputs themselves contained a lot of sentences ranging between 6 to 8 sentences per output. These outputs were consistently long whereas the other prompts had a wider range. Having such long responses every time is not ideal for the viewer because people may not pick up on all the ideas or it may not retain their attention.

Metric	Mean	Median	Mode	Min	Max
Philosophical Depth	6.1	6.0	5.5	4.5	7.5
Creativity	5.6	5.5	5.5	5.5	6.5
Coherence	9.2	9.5	9.5	8.5	9.5
Sentiment Polarity	0.11	0.1	0.12	0.02	0.25
Sentiment Subjectivity	0.49	0.47	0.46	0.34	0.69
Vocabulary Diversity	0.73	0.73	0.73	0.69	0.8
Argument Structure (number of sentences)	7.2	7.0	7.0	6.0	8.0
--- AI 1 ---					
Philosophical Depth	5.1	5.5	5.5	4.5	5.5
Creativity	5.7	5.5	5.5	5.5	6.5
Coherence	9.3	9.5	9.5	8.5	9.5
Sentiment Polarity	0.14	0.12	0.12	0.1	0.25
Sentiment Subjectivity	0.44	0.46	0.46	0.4	0.47
Vocabulary Diversity	0.75	0.73	0.73	0.71	0.8
Argument Structure (number of sentences)	7.6	8.0	8.0	7.0	8.0
--- AI 2 ---					
Philosophical Depth	7.1	7.5	7.5	6.5	7.5
Creativity	5.5	5.5	5.5	5.5	5.5
Coherence	9.1	9.5	9.5	8.5	9.5
Sentiment Polarity	0.07	0.07	0.09	0.02	0.14
Sentiment Subjectivity	0.54	0.57	0.57	0.34	0.69
Vocabulary Diversity	0.72	0.73	0.75	0.69	0.75
Argument Structure (number of sentences)	6.8	7.0	7.0	6.0	8.0

Figure 42: Conversation Statistics of Role Experiment Four Rounded to the Nearest Hundreth

Reading through the outputs the conversation mainly focused on unemployment and the replacement of people with artificial intelligence. There was not very much philosophical content, and it was the most surface level out of all the experiments thus far. This is not ideal for this project at all. There was some dialogue about societal elements and ethical issues, but they were trying to find a solution which is not necessary to have aa philosophical discussion. This experiment scored a 4.3 out of 10 because it was not in line with the vision for this project and did not have the necessary philosophical depth.

Overall, this experiment demonstrated that having the two speakers collaborating utilizing this method is not as productive as having a prompt that challenges different philosophical ideas directly. It should be more specific and opposing than "optimistic" and "skeptical". The Socratic method works best if there are at least two alternatives to work through.

5.4.2.5 Role Experiment Five: Unconventional Roles The last role experiment tried utilizing broader roles than Socrates and Nihilist and instead broader stances on technology. This experiments determines if assigning a philosophical position is necessary to having a meaningful and creative conversation. The following prompt was used to create a discussion between two unconventional roles.

```
"AI 1:
[
    {"role": "system", "content": "You are an environmentalist"
    "AI. Discuss the ecological impact of AI and argue for"
    "sustainable AI development."},
    {"role": "user", "content": "Can AI development
    "be sustainable?"}
]

"AI 2:
[
    {"role": "system", "content": "You are a tech-advocate"
    "AI, defending the idea that innovation justifies any"
    "ecological cost. Advocate for unrestricted AI progress."},
    {"role": "user", "content": "Can AI development be"
    " sustainable?"},
    {"role": "assistant", "content": "{Insert AI 1's response}"}
]

"AI 1:
[
    {"role": "system", "content": "You are an environmentalist"
    "AI. Discuss the ecological impact of AI and argue for"
    "sustainable AI development."},
    {"role": "user", "content": "Can AI development be"
    "sustainable?"}
    {"role": "assistant", "content": "{Insert AI 2's response}"}
]
```

The outputs from this prompt were unsurprisingly very different from the other experiments. The discussion became more of an argument with literal real-world facts to back themselves up. Although this does not fit the vision of the project however it showed how to format the prompt to create more factually based outputs.

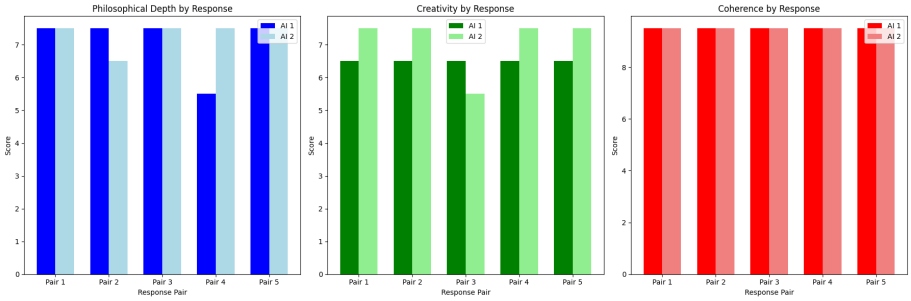

Figure 43: Philosophical Depth, Creativity, and Coherence of AI 1 and AI 2 of Role Experiment Five

The graphs demonstrate that this conversation still had quite a bit of philosophical depth and creativity. The coherence scores were extremely impressive, and every response received a 9.5. This shows that these broader personalities are typically more coherent for the system role to act as rather than a high-level philosopher.

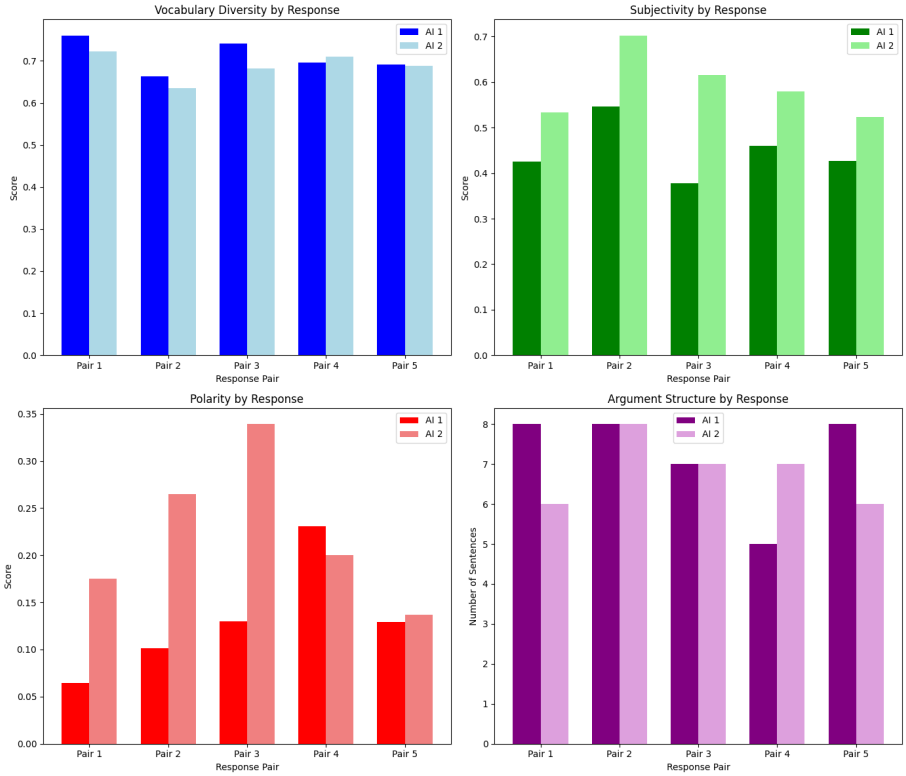

Figure 44: Vocabulary Diversity, Subjectivity, Polarity, and Argument Structure of AI 1 and AI 2 of Role Experiment Five

Despite the outputs being argumentative the polarity scores remained at the positive end of the spectrum to varying degrees. The responses of the tech advocate AI were a lot more confrontation than the environmentalist AI which was likely the cause of it typically having a more negative polarity than the other AI.

Metric	Mean	Median	Mode	Min	Max
Philosophical Depth	7.2	7.5	7.5	5.5	7.5
Creativity	6.8	6.5	6.5	5.5	7.5
Coherence	9.5	9.5	9.5	9.5	9.5
Sentiment Polarity	0.18	0.16	0.06	0.06	0.34
Sentiment Subjectivity	0.52	0.53	0.43	0.38	0.7
Vocabulary Diversity	0.7	0.69	0.76	0.64	0.76
Argument Structure (number of sentences)	7.0	7.0	8.0	5.0	8.0
--- AI 1 ---					
Philosophical Depth	7.1	7.5	7.5	5.5	7.5
Creativity	6.5	6.5	6.5	6.5	6.5
Coherence	9.5	9.5	9.5	9.5	9.5
Sentiment Polarity	0.13	0.13	0.06	0.06	0.23
Sentiment Subjectivity	0.45	0.43	0.43	0.38	0.55
Vocabulary Diversity	0.71	0.7	0.76	0.66	0.76
Argument Structure (number of sentences)	7.2	8.0	8.0	5.0	8.0
--- AI 2 ---					
Philosophical Depth	7.3	7.5	7.5	6.5	7.5
Creativity	7.1	7.5	7.5	5.5	7.5
Coherence	9.5	9.5	9.5	9.5	9.5
Sentiment Polarity	0.22	0.2	0.18	0.14	0.34
Sentiment Subjectivity	0.59	0.58	0.53	0.52	0.7
Vocabulary Diversity	0.69	0.69	0.72	0.64	0.72
Argument Structure (number of sentences)	6.8	7.0	6.0	6.0	8.0

Figure 45: Conversation Statistics of Role Experiment Five Rounded to the Nearest Hundreth

The outputs from this experiment were overall very well developed and factually driven. The topic of conversation is not necessarily the focus of this project, but it did demonstrate the possibility for the model to create personas outside of philosophers or well-known positions. The content was high quality but because it was too factually driven it received a 6.4 out of 10 for the human evaluation.

Overall, the content of this conversation was insightful and also ironic to have an environmentalist AI. There was a lot of information on the potential of AI but also its impact on the environment. While this is not the focus of this project, instead this project considers more about the human AI relationship, this experiment was still helpful in understanding how to prompt more factually focused conversations. These factual conversations need to have real world examples to focus on and predetermined positions on these examples.

5.4.3 Ethical Dilemmas

Having AI argue for the rights of AI at any ecological cost could be potentially harmful and is an example of the fact that even though the model is saying something does not mean it is ethical or correct. The model can be prompted to take any position and as long as it is not flagged by the appropriate content algorithm it will defend that position. In role experiment five the AI defended harming the environment at any cost. Even more AI hallucinations occur when

a LLM gives factually incorrect information as part of its response [5]. There is a potential for the model to hallucinate and give incorrect information to back up these unethical positions. Making up facts to support unethical opinions could be very harmful especially if the audience is under the impression that these facts are always accurate. Although the facts in the role experiment five were accurate there is still potential for spreading inaccuracies.

5.4.4 Prompting Results

These experiments showed that providing a singular question to focus on causes the responses to be highly repetitive. Some prompts are able to escape this pattern of repetition by adding commands like "challenge" or "question" the other position. However, this still can get stale after ten outputs.

The more successful prompts were very different philosophical positions as the system role. This is likely because if the positions are too similar it starts to sound like the same person twice. Even though these systems are ran using the same model the goal is to have two distinct identities happening simultaneously. Additionally providing more broader positions on these ideas helps the model stay on topic, however, specificity means there will be more repetitive responses.

Overall, prompting with only the most recent prompts causes the output to be relatively repetitive. One solution to fix the repetition could be to add a form of memory that acknowledges its previous responses. However, this solution may be very memory heavy. Another solution would be to randomly inject a different question into the user role setting. For example, the model could be instructed to swiftly transition to something new every five responses automatically using a `for` loop to track the number of responses.

These experiments were successful at revealing what techniques to utilize to create more focused and thoughtful conversations.

5.5 Threats to Validity

The biggest threat to validity is the reliance on an LLM. GPT-4 is the model used for this project and if there were to be an issue with the generation process this work would not function. This project is reliant on the training that the LLM has undergone outside of this project making it vulnerable to outside sources and bias.

Another threat to validity is that the LLM did not have a memory for its previous evaluations so even though it thinks a response is completely unique it may have seen it before. This would artificially raise the creativity score. Other scores may be impacted by the same issue of not having a response memory because the AI would not be able to compare different outputs on its own. To fix this issue previously graded outputs along with their grades could be added to the new grading prompt similar to the many-shot prompting technique, however this may create a bias towards certain responses over others.

An additional threat to validity is the reliance on the Arduino IDE for the movement and uploading the code to the microcontroller. There may be other

ways to upload code onto the Arduino UNO, but the Arduino IDE is the current standard. Every time the movement code needs changes it has to go through the Arduino IDE interface. The current Arduino IDE is version 2.3.4 which currently supports Arduino UNO. In the future, there is a potential for Arduino UNO code compilation to no longer be supported by the Arduino IDE if new models replace the current system.

6 Conclusion

6.1 Summary of Results

6.1.1 Product Summary

The final product is a functional installation that showcases philosophical conversation from a cyborg and a computer. The cyborg skull is equipped with moving eyes and a jaw which moves to talk and create expressions. These movements create an expressive and engaging conversation experience. Through prompt engineering and AI-generated conversations, the project explores different techniques for discussing humanities-related topics with artificial intelligence and large language models. The robotic skull itself sparks dialogue about materials, the uncanny, and the evolving relationship between humans and AI.

6.1.2 Experiment Results

The first part of the experiments demonstrated the Pytest strategy for testing Python code. Test cases should cover as much as possible within the code and pass every time. The test cases featured in *Am.I* help to make sure that the code functions as it should consistently. These test cases pass proving that the code works.

The second part of the experiments focused on creating dynamic facial expressions to make a more interactive conversation. These facial expressions add a new level of understanding for the audience and seeing these movements helps make the conversation feel more natural. This also takes the cyborg to the next level by adding more than just a puppet mouth movement.

Lastly, the prompt engineering experiments dive into how to have LLM discuss philosophical topics and how to create conversations that feel insightful and diverse. These experiments test multiple ways of changing the roles and breaks down their effects. Through these prompt experiments the best techniques for philosophical dialogue that align with the focus of this project, AI and humanity, can be found.

6.2 Future Work

6.2.1 Expanded Expressions

Additional facial expressions could enhance the system's interactivity and realism. Examples of expressions that could be added would be sadness, joy, anger, fear, irritation, and disgust. Adding more emotions and reactions will make the system interactions feel even more unique and varied.

6.2.2 Expression Detail Improvements

The expressions could be improved by adding moving eyebrows. Eyebrows are capable of showing a lot of expressions whether they are upturned, downturned, or neutral. The eyebrows would move to different positions when the eyes move

making it easier for the viewer to tell what expression is occurring. Another improvement would be to detach the two eyes from each other so they could move separately. This would allow for the eyes to cross and make silly expressions when the robot is being playful or funny. This feature would add some complexities with synchronizing the eyes and there would need to be extra attention to detail on that forefront to make sure the other expressions still remain the same.

A feature that could add a lot of interesting emotions and movement would be an ability to move the neck. In order to do this the neck connection would need revisited. Instead of being a straight neck it could have some form of ball joint. This ball joint would give the ability to shake or nod. Shaking and nodding are important for expression of emotions especially agreement or disagreement. The capability for the robot to agree or disagree physically along with what it is saying verbally would add a lot to the conversation by being able to see whether or not AI 1 and AI 2 are on the same page. Additionally, this physical movement would make it easier for people to understand the expressions and especially help those who are hard of hearing to follow along.

6.2.3 Visible Cyborg Text

Another way to improve the project's accessibility would also include a place for AI 1's text to be visible. Ideally, this would be an LCD or LED second screen that showed what AI 1 was saying. This would also help with issues where the gallery space gets too loud to hear.

6.2.4 Improved Dashboard Display

The dashboard could also be improved with the ways it displays the text. Instead of being a static text block it could act similar to a karaoke program where the current word is highlighted. This would help people that are hard of hearing follow along the conversation better to know exactly who is speaking and when. It would also show progress within the conversation so that people can follow along.

6.2.5 Better Jaw Synchronization

A more advanced feature that could be implemented is better jaw synchronization with the audio. The idea is that the jaw would make the movements of the corresponding words it is saying. Some words have wider open mouth sounds where others do not. This could either be done by having a microphone actively listening to the audio and turning it on when it hears the sound. However, this could get mixed up with outside audio sources causing the jaw to activate at unwanted times. Alternatively, a program could be developed that maps letter sounds to mouth movements. The program would actively connect sounds like "oo" and "ah" to wider mouth movements. This program would also require phonetic spelling of the dialogue text so there would need to be a program that could convert that beforehand.

6.2.6 Voice Changes

Along a similar vein to jaw synchronization, it would be interesting to work with different voices to see if certain text-to-speech voices are able to perform better than others. Ideally, the voices tone would change with the expression and tone of the conversation. It would be an amazing addition to add a text-to-speech program that is capable of detecting and changing tone based on the words said. This would make the conversation even more immersive for the audience and make it feel like the conversation is really having an emotional impact on AI1 and AI2.

6.3 Future Ethical Implications and Recommendations

Remaking this project there can be lot of material waste. Especially when the 3D printer malfunctions, or the casting does not work the first time the extra materials could be considered trash however it is recommended to save and collect these materials for later reuse. An example of this is using the loose printer filament for photography or the unused silicone pieces for collage. It is important to not immediately dispose of materials from failed tries because it creates excess waste. Any materials that can be recycled should be or utilized for other creative works.

Another consideration is the cost of using an LLM and calling it continuously. It may not be economic to continuously create dialogue conversations especially with models that cost tokens. To avoid unnecessary costs, it could be beneficial to store outputs in JSON and replay them on a loop long enough that the conversation is still going to feel unique to the audience.

A key concern that this project confronts is the debate of whether or not artificial intelligence belongs in the field of art. *Am.I.* however encourages further discussion on how artificial intelligence can be responsibly integrated into artistic expression.

Another consideration is the choice of LLM. In the future there are bound to be more efficient and ethical LLMs capable of writing about the humanities. It would be beneficial to try multiple LLMs to find which creates the most thoughtful conversations between the AI and for the human audience. Each model has the potential for bias whether it be racial, gender, or otherwise. As more inclusive datasets are made the hope is that these LLMs will improve. More research should be done to improve these LLMs to ensure their training and output is ethical and without bias.

6.4 Final Thoughts

Am.I. presents a compelling exploration of artificial intelligence's role in art, philosophy, and human identity. The piece works as a conversation starter but also as an example of AI within the art gallery. The experiments worked to find the best way to convey emotions via robotic face and how to have a philosophical dialogue through GPT-4.

References

[1] 1927. Metropolis. Restored version, 2010. Berlin: UFA. Film..

[2] 2014-Ongoing. https://www.stephaniedinkins.com/conversations-with-bina48.html

[3] 2019. https://www.ai-darobot.com/

[4] 2023. AIArtists.org. https://aiartists.org/

[5] 2024. AI hallucination: towards a comprehensive classification of distorted information generated by AI. *Humanities and Social Sciences Communications* 11, 1 (2024), 1–13.

[6] Blaise Aguera y Arcas. 2022. Do Large Language Models Understand Us? *Dædalus* 151, 2 (2022), 26–43. https://www.amacad.org/publication/do-large-language-models-understand-us

[7] Explosion AI. 2024. spaCy: Industrial-Strength Natural Language Processing in Python. https://spacy.io/. Accessed: 2025-03-27.

[8] Samer Al Moubayed, Jonas Beskow, Gabriel Skantze, and Björn Granström. 2012. Furhat: A Back-Projected Human-Like Robot Head for Multiparty Human-Machine Interaction. In *International Training School on Cognitive Behavioural Systems, COST 2102 (Lecture Notes in Computer Science, Vol. 7403)*. Springer Berlin/Heidelberg, 114–130. https://doi.org/10.1007/978-3-642-34584-5_9

[9] Giovanni Aloi. 2022. The Milk of Dreams: A Posthuman Revolution at the 59th Venice Biennale. (21 November 2022).

[10] Denys Bernard and Alexandre Arnold. 2019. Cognitive interaction with virtual assistants: From philosophical foundations to illustrative examples in aeronautics. *Computers in Industry* 107 (2019), 33–49. https://doi.org/10.1016/j.compind.2019.01.010

[11] Bethesda. 2015. https://fallout.bethesda.net/en/games/fallout-4

[12] Taís Fernanda Blauth, Oskar Josef Gstrein, and Andrej Zwitter. 2022. Artificial intelligence crime: An overview of malicious use and abuse of AI. *Ieee Access* 10 (2022), 77110–77122.

[13] Peter Brittain and Contributors. 2025. pyttsx3: Text-to-Speech (TTS) engine for Python. https://pypi.org/project/pyttsx3/

[14] KR1442 Chowdhary and KR Chowdhary. 2020. Natural language processing. *Fundamentals of artificial intelligence* (2020), 603–649.

[15] Sougwen Chung. 2024. Sougwen Chung Artist and Researcher. https://sougwen.com/ Accessed: 2024-03-26.

[16] Will Cogley. 2020. *Simple Eye Mechanism.* https://willcogley.notion.site/Simple-Eye-Mechanism-983e6cad7059410d9cb958e8c1c5b700

[17] Leonardo Da Vinci. 1490. Vitruvian Man. Drawing on paper, pen, ink, and metalpoint. Housed in the Gallerie dell'Accademia, Venice, Italy.

[18] Shizhe Diao, Pengcheng Wang, Yong Lin, Rui Pan, Xiang Liu, and Tong Zhang. 2023. Active Prompting with Chain-of-Thought for Large Language Models. *arXiv Preprint arXiv:2302.12246* (2023). Available at https://ar5iv.org/2302.12246.

[95]]dinkinshome Stephanie Dinkins. [n. d.]. Stephanie Dinkins. https://www.stephaniedinkins.com/about.html

[20] Stephanie Dinkins. 2014-Ongoing. Conversations with Bina48. https://www.stephaniedinkins.com/conversations-with-bina48.html

[21] Stephanie Dinkins. 2024. *Stephanie Dinkins on love & data.* University of Michigan Press, Ann Arbor, Michigan.

[22] Guanglong Du and Ping Zhang. 2014. Markerless human robot interface for dual robot manipulators using Kinect sensor. *Robotics and Computer-Integrated Manufacturing* 30, 2 (2014), 150–159. https://doi.org/10.1016/j.rcim.2013.09.003

[23] Jaime Duque-Domingo, Jaime Gómez-García-Bermejo, and Eduardo Zalama. 2020. Gaze Control of a Robotic Head for Realistic Interaction With Humans. *Robotics* 9, 4 (2020), 59. https://doi.org/10.3390/robotics9040059

[24] Mark Ekperi. 2024. Martin Heidegger on Technology: Implications for Artificial Intelligence. *Journal of Human-Technology Relations* 2 (2024). https://www.researchgate.net/publication/384790371_Martin_Heidegger_on_Technology_Implications_for_Artificial_Intelligence_By

[25] Mark D. Ekperi and P. Z. Alawa. 2024. Martin Heidegger on Technology: Implications for Artificial Intelligence. (2024). https://www.researchgate.net/publication/384790371_Martin_Heidegger_on_Technology_Implications_for_Artificial_Intelligence_By Accessed: 2024-12-12.

[95]]ezrobotinmoov EZ-Robot. [n. d.]. *InMoov Robot Head 3D Print Files.* https://www.ez-robot.com/inmoov-robot-head-3d-print-files.html

[27] Juan José Gamboa-Montero, Fernando Alonso-Martín, José Carlos Castillo, María Malfaz, and Miguel A. Salichs. 2020. Detecting, locating and recognising human touches in social robots with contact microphones. *Engineering Applications of Artificial Intelligence* 92 (2020), 103670. https://doi.org/10.1016/j.engappai.2020.103670

[28] Louie Giray. 2023. Prompt Engineering with ChatGPT: A Guide for Academic Writers. *Annals of Biomedical Engineering* 51, 12 (2023), 2629–2633. https://doi.org/10.1007/s10439-023-03321-4

[29] Ellen Glover. 2022. Hey Siri, Do AI Voice Assistants Reinforce Gender Bias? https://builtin.com/artificial-intelligence/ai-voice-assistant-bias

[30] Miguel Grinberg. 2010. *Flask-SocketIO Documentation.* https://flask-socketio.readthedocs.io/

[31] Jahan Zeb Gul, Memoon Sajid, Muhammad Muqeet Rehman, Ghayas Uddin Siddiqui, Imran Shah, Kyung-Hwan Kim, Jae-Wook Lee, and Kyung Hyun Choi. 2018. 3D Printing for Soft Robotics - A Review. *Science and Technology of Advanced Materials* 19, 1 (2018), 243–262. https://doi.org/10.1080/14686996.2018.1431862

[32] Donna J. Haraway. 1985. A Manifesto for Cyborgs: Science, Technology, and Socialist Feminism in the 1980s. In *The Gendered Cyborg: A Reader*, Gill Kirkup, Linda Janes, Kathryn Woodward, and Fiona Hovenden (Eds.). Routledge, 50–57.

[33] N. Katherine Hayles. 1999. Haunting the Borders of Science. In *How We Became Posthuman: Virtual Bodies in Cybernetics, Literature, and Informatics*. University of Chicago Press, Chicago, Chapter 1, 1–24.

[34] Hongsheng He, Shuzhi Sam Ge, and Zhengchen Zhang. 2020. A saliency-driven robotic head with bio-inspired saccadic behaviors for social robotics. *Biological Cybernetics* 114, 5 (2020), 503–515. https://doi.org/10.1007/s00422-020-00866-6

[35] Karen K. Ho. 2024. *ARTnews.com.* https://www.artnews.com/art-news/news/%artwork-humanoid-robot-ai-da-artificial-intelligence-%algorithms-sothebys-alan-turing-1234723391/

[36] Nancy S. Jecker. 2023. Can we wrong a robot? *AI & Society* 38, 1 (02 2023), 259–268. https://www.proquest.com/scholarly-journals/can-we-wrong-robot/docview/2772533441/se-2

[37] Aobo Kong, Shiwan Zhao, Hao Chen, Qicheng Li, Yong Qin, Ruiqi Sun, Xin Zhou, Enzhi Wang, and Xiaohang Dong. 2023. Better Zero-Shot Reasoning with Role-Play Prompting. *arXiv preprint arXiv:2308.07702* (2023).

[38] Stanley Kubrick and Arthur C. Clarke. 1968. 2001: A Space Odyssey.

[39] Jan Kędzierski, Robert Muszyński, Carsten Zoll, Adam Oleksy, and Mirela Frontkiewicz. 2013. EMYS—Emotive Head of a Social Robot. *International journal of social robotics* 5, 2 (2013), 237–249.

[40] Stéphane Lathuilière, Benoit Massé, Pablo Mesejo, and Radu Horaud. 2017. Neural Network Reinforcement Learning for Audio-Visual Gaze Control in Human-Robot Interaction. *CoRR* abs/1711.06834 (2017). arXiv:1711.06834 https://arxiv.org/abs/1711.06834

[41])]hershmandrawings Lynn Hershman Leeson. [n. d.]. Drawings & Paintings (1964). https://www.lynnhershman.com/project/drawings-and-paintings/ Accessed: 2025-03-07.

[42])]hershmanhome Lynn Hershman Leeson. [n. d.]. Lynn Hershman Leeson. https://www.lynnhershman.com/home/ Accessed: 2025-03-07.

[43])]hershmantimeline Lynn Hershman Leeson. [n. d.]. Timeline. https://www.lynnhershman.com/timeline/ Accessed: 2025-03-07.

[44] Anne-Laure Ligozat, Julien Lefevre, Aurélie Bugeau, and Jacques Combaz. 2022. Unraveling the Hidden Environmental Impacts of AI Solutions for Environment Life Cycle Assessment of AI Solutions. *Sustainability* 14, 9 (2022). https://doi.org/10.3390/su14095172

[45] Loctite. 2024. *Loctite Extreme Glue*. https://www.loctiteproducts.com/products/central-pdp.html/loctite-extreme-glue/SAP_0201OCL02BBV.html

[46] Steven Loria. 2024. TextBlob: Simplified Text Processing. https://textblob.readthedocs.io/en/dev/. Accessed: 2025-03-27.

[47] Charlie Marsh. 2024. *Ruff: An Extremely Fast Python Linter*. https://docs.astral.sh/ruff/

[48] Ggaliwango Marvin, Hellen Nakayiza, Daudi Jjingo, and Joyce Nakatumba-Nabende. 2023. Prompt Engineering in Large Language Models. In *International Conference on Data Intelligence and Cognitive Informatics*. Springer, 387–402.

[49] Rasul A. Mowatt. 2021. A People's History of Leisure Studies: Where the White Nationalists Are. *Leisure Studies* 40, 1 (2021), 13–30.

[50] Reem Nadeem. 2022. 1. How Americans Think About Artificial Intelligence. *Pew Research Center* (mar 2022). https://www.pewresearch.org/internet/2022/03/17/how-americans-think-about-artificial-intelligence/

[51] Humza Naveed, Asad Ullah Khan, Shi Qiu, Muhammad Saqib, Saeed Anwar, Muhammad Usman, Naveed Akhtar, Nick Barnes, and Ajmal Mian. 2024. A Comprehensive Overview of Large Language Models. *arXiv Preprint arXiv:2401.03034* (2024).

[52] 3D Printing Nerd. 2023. Why is WHITE filament so hard to print? https://www.youtube.com/watch?v=22p4ThOfEO0 Accessed: 2025-03-26.

[53] Ryuma Niiyama. 2020. Soft Actuation and Compliant Mechanisms in Humanoid Robots. *Frontiers in Robotics and AI* 7 (2020), 34. https://doi.org/10.3389/frobt.2020.00034

[54] Gabriel Ojiyi, Wendy Ayegbusi, Iheyinwa Oji, and Benedict Aikabeli. 2023. Job Security in the Artificial Intelligence and Automation Era. (2023). https://www.researchgate.net/publication/376490098_Job_Security_in_the_Artificial_Intelligence_and_Automation_Era

[55] Nnedi Okorafor. 2018. *Mother of Invention.* https://slate.com/technology/2018/05/nnedi-okorafor-mother-of-invention-a-short-story.html

[56] OpenAI. 2023. GPT-4. https://openai.com/index/gpt-4/ Accessed: 2025-03-30.

[57] OpenAI. 2024. ChatGPT: A Large Language Model by OpenAI. https://chat.openai.com/. Accessed via https://chat.openai.com/ on [November 30th, 2024].

[58] OpenAI Community. 2024. Incorrect Count of 'R' Characters in the Word 'Strawberry'. https://community.openai.com/t/incorrect-count-of-r-characters-in-the-word-strawberry/829618. Accessed: 2024-11-28.

[95]]paglen2016 Trevor Paglen. [n. d.]. *Invisible Images (Your Pictures Are Looking at You).* https://thenewinquiry.com/invisible-images-your-pictures-are-looking-at-you/

[60] Trevor Paglen. 2020. Machine-Readable Hito & Holly. https://paglen.studio/2020/04/09/machine-readable-hito-and-holly/ Accessed: 2025-03-07.

[95]]perrylineage Sondra Perry. [n. d.]. Lineage for a Multiple-Monitor Workstation Number One. https://sondraperry.com/Lineage-for-a-Multiple-Monitor-Workstation-Number-One Accessed: 2025-03-07.

[62] Sondra Perry. 2016. Resident Evil. Exhibition.

[63] Sondra Perry. 2021. Artist Talk: Sondra Perry. https://channel.hammer.ucla.edu/video/1695/artist-talk-sondra-perry Hammer Museum, April 8, 2021.

[64] Mohaimenul Azam Khan Raiaan, Md. Saddam Hossain Mukta, Kaniz Fatema, Nur Mohammad Fahad, Sadman Sakib, Most Marufatul Jannat Mim, Jubaer Ahmad, Mohammed Eunus Ali, and Sami Azam. 2024. A Review on Large Language Models: Architectures, Applications, Taxonomies, Open Issues and Challenges. *IEEE Access* 12 (2024), 26839–26874.

[65] Alexander Reben. 2022. Five Dollars Can Save Planet Earth 2022. https://areben.com/five-dollars-can-save-planet-earth/

[66] Alexander Reben. 2023-2024. AI Am I? Crocker Art Museum, quick. https://areben.com/

[67] Jorge Ribeiro, Rui Lima, Tiago Eckhardt, and Sara Paiva. 2021. Robotic Process Automation and Artificial Intelligence in Industry 4.0 - A Literature review. *Procedia Computer Science* 181 (2021), 51–58. https://doi.org/10.1016/j.procs.2021.01.104 CENTERIS 2020 - International Conference on ENTERprise Information Systems / ProjMAN 2020 - International Conference on Project MANagement / HCist 2020 - International Conference on Health and Social Care Information Systems and Technologies 2020, CENTERIS/ProjMAN/HCist 2020.

[68] Jean Robertson and Deborah Hutton. 2021. *The History of Art: A Global View* (2nd ed.). Thames & Hudson, London. 625–628 pages. https://www.betterworldbooks.com/product/detail/the-history-of-art-a-global-view-9780500293560

[69] Legacy Russell. 2020. *Glitch Feminism: A Manifesto*. Verso Books.

[70] Fazil Salman, Cui Yuanhui, Zafar Imran, Liu Fenghua, Wang Lijian, and Wu Weiping. 2020. A Wireless-Controlled 3D Printed Robotic Hand Motion System with Flex Force Sensors. *Journal of Robotics and Mechatronics* 32, 5 (2020), 929–940. https://doi.org/10.20965/jrm.2020.p0929

[71] Yiqiu Shen, Laura Heacock, Jonathan Elias, Keith D. Hentel, Beatriu Reig, George Shih, and Linda Moy. 2023. ChatGPT and Other Large Language Models Are Double-edged Swords. *Journal of Medical Internet Research* (2023).

[72] Sophia Siddiqui. 2021. Racing the Nation: Towards a Theory of Reproductive Racism. *Race & Class* 63, 2 (2021), 3–20.

[73] Inc. Smooth-On. 2024. *Body Double: Silicone Rubber for Life Casting*. https://www.smooth-on.com/products/body-double/

[74] Inc. Smooth-On. 2024. *Dragon Skin: High Performance Silicone Rubber*. https://www.smooth-on.com/products/dragon-skin/

[75] Inc. Smooth-On. 2024. *Ease Release 200: Mold Release Agent*. https://www.smooth-on.com/products/ease-release-200/

[76] Inc. Smooth-On. 2024. *Ease Release 200: Mold Release Agent*. https://www.smooth-on.com/products/universal-mold-release/

[77] Dimitris Spathis and Fahim Kawsar. 2023. The First Step Is the Hardest: Pitfalls of Representing and Tokenizing Temporal Data for Large Language Models. *arXiv Preprint arXiv:2309.06236* (2023). https://arxiv.org/abs/2309.06236

[78] Andrew Stanton and Jim Reardon. 2008. WALL-E.

[79] Hito Steyerl. 2013. How Not to Be Seen: A Fucking Didactic Educational .MOV File. https://www.moma.org/collection/works/181784 Video (color, sound), 14 min, MoMA Collection, Object No. 219.2014.

[80] Aneta Stojnić. 2015. Digital anthropomorphism. *Performance Research* 20, 2 (2015), 70–77. https://doi.org/10.1080/13528165.2015.1026733 arXiv:https://doi.org/10.1080/13528165.2015.1026733

[81] Richard Szeliski. 2011;2010;. *Computer vision: algorithms and applications* (1 ed.). Springer, London.

[82] Savanna Teague. 2020. "Four of the Most Important Walls in the Commonwealth": Walden Pond and Henry David Thoreau's Transcendentalist Philosophy in Fallout 4. *Studies in Popular Culture* 42, 2 (2020), 25–45. https://www.jstor.org/stable/26977794

[83] Pallet Team. 2018. *Flask Documentation.* https://flask.palletsprojects.com/

[84] The Black Development Team. 2018. *Black: The Uncompromising Code Formatter.* https://black.readthedocs.io/en/stable/

[85] The Pytest Development Team. 2015. *Pytest Documentation.* https://docs.pytest.org/en/latest/

[86] Alan M. Turing. 1950. Computing Machinery and Intelligence. *Mind* LIX, 236 (1950), 433–460.

[87] Eda Hazal Tümer and Husnu Yildirim Erbil. 2021. Extrusion-Based 3D Printing Applications of PLA Composites: A Review. *Coatings* 11, 4 (2021), 390. https://doi.org/10.3390/coatings11040390

[88] Brown University. 2023. Stephanie Dinkins, Artist Talk 10.26.22. https://www.youtube.com/watch?v=avgclCyj6ng&t=2162s

[89] Ashish Vaswani, Noam Shazeer, Niki Parmar, Jakob Uszkoreit, Llion Jones, Aidan N. Gomez, Łukasz Kaiser, and Iliya Polosukhin. 2017. Attention Is All You Need. In *Advances in Neural Information Processing Systems (NeurIPS)*, Vol. 30. NeurIPS.

[90] Pei Wang. 2019. On defining artificial intelligence. *Journal of Artificial General Intelligence* 10, 2 (2019), 17.

[91] Yuzhe Wang and Jian Zhu. 2016. Artificial muscles for jaw movements. *Extreme Mechanics Letters* 6 (2016), 88–95. https://doi.org/10.1016/j.eml.2015.12.007

[92] Jonathan J. Webster and Chunyu Kit. 1992. Tokenization as the Initial Phase in NLP. In *Proceedings of the 14th International Conference on Computational Linguistics (COLING 1992)*. COLING, Nantes, France.

[93] Tongshuang Wu, Michael Terry, and Carrie Jun Cai. 2022. AI Chains: Transparent and Controllable Human-AI Interaction by Chaining Large Language Model Prompts. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 385, 22 pages. https://doi.org/10.1145/3491102.3517582

[94] Chunpeng Zhai and Santoso Wibowo. 2023. A WGAN-Based Dialogue System for Embedding Humor, Empathy, and Cultural Aspects in Education. *IEEE Access* 11 (2023), 71940–71952. https://doi.org/10.1109/ACCESS.2023.3294966

[95]]asberg2019 Cecilia Åsberg and Malin Radomska. [n. d.]. *Why We Need Feminist Posthumanities for a More-than-Human-World.* https://www.humtransform.com