OPTIMIZING A VIRTUAL HUMAN PLATFORM FOR DEPRESSION/SUICIDE IDEATION IDENTIFICATION FOR THE AMERICAN SOLDIER

A Thesis

presented to

the Faculty of California Polytechnic State University,

San Luis Obispo

In Partial Fulfillment

of the Requirements for the Degree

Master of Science in Computer Science

by

Christina Monahan

December 2021

COMMITTEE MEMBERSHIP

TITLE: OPTIMIZING A VIRTUAL HUMAN PLATFORM FOR DEPRESSION/SUICIDE IDEATION IDENTIFICATION FOR THE AMERICAN SOLDIER

AUTHOR: Christina Monahan

DATE SUBMITTED: December 2021

COMMITTEE CHAIR: Franz Kurfess, Ph.D.
Professor of Computer Science

COMMITTEE MEMBER: Maria Pantoja, Ph.D.
Professor of Computer Science

COMMITTEE MEMBER: Albert 'Skip' Rizzo, Ph.D.
Director, Medical Virtual Reality - ICT, USC

ABSTRACT

OPTIMIZING A VIRTUAL HUMAN PLATFORM FOR DEPRESSION/SUICIDE IDEATION IDENTIFICATION FOR THE AMERICAN SOLDIER

Christina Monahan

Suicide surpassed homicide to be the second leading cause of death among people 10-24 years old in the United States [13]. This statistic is alarming especially when combined with the more than eight distinctly different types of clinical depression among society today [14]. To further complicate this health crisis, let's consider the current worldwide isolating pandemic often referred to as COVID-19 that has spanned 12 months. It is more important than ever to consider how we can get ahead of the crisis by identifying the symptoms as they set in and more importantly ahead of the decision to commit suicide. To capitalize on the modern shift to electronic-based interactions [13], the use of Artificial Intelligence (AI) and Machine Learning (ML) methods to aid in identification have been previously implemented in Virtual Human interviewing platforms. This effort examines these existing approaches and includes an independent survey that is used to solve the gap in early identification of depression and suicidal ideation using a virtual human interviewing platform by soliciting honest, open, and current feedback from Soldiers on how to optimize such a system to encourage its use in the future. Specifically, the analysis of the survey results identify critical gaps from a participants perspective to be security, customization's, and error handling recommended to be included in future development of the EMPOWER (Enhancing Mental Performance and Optimizing Warfighter Effectiveness and Resilience: From MultiSense to OmniSense) platform. These recommendations are provided to the USC-ICT EMPOWER team to be included in the next prototype and system test.

Keywords: Multimodal, OmniSense, EMPOWER, Virtual Human Agent

ACKNOWLEDGMENTS

Thanks to:

- God above for allowing me this opportunity to make a difference.
- My family has always remained my largest supporter. My mom and dad – have traveled the world supporting my military career and have always encouraged me to challenge myself. My daughter, Brianna, continues to amaze me and inspires me to set a positive example. My sisters Cynthia, Cheryl and Chrystal – have each gone out of their way to continue to push me, support me, and love me unconditionally.
- Professor Kurfess for accepting a curious lifetime Soldier into the remarkable world of Artificial Intelligence, adapting complex topics to engage every student, and the mentorship to explore a topic that is personal yet extremely important.
- All of my fellow Soldiers for assisting with the survey results and helping me to elevate a critical topic to be improved for all future Soldiers in crisis.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

Chapter 1

INTRODUCTION

According to USA Today in 2019, the Pentagon released the suicide statistics for Active Duty military which concluded that "Suicide rate AMONG active-duty TROOPS jumps to six-year high" [2]. This statistic illustrates a drastic jump prior to the worldwide pandemic that has now spanned 19 months. It is more important than ever to consider how we can get ahead of the crisis by identifying the symptoms as they set in and more importantly ahead of the decision to commit suicide. To capitalize on the modern shift to electronic-based interactions [13], this research will examine the existing approaches to Artificial Intelligence (AI) and Machine Learning (ML) methods, identify and propose a solution to close the gap in early identification of depression and suicidal ideation by optimizing a virtual human interviewing platform.

Initial motivation is to impact the critical need of the American Soldier in crisis by improving identification for early intervention. PTSD (Post Traumatic Stress Disorder), Depression and Suicidal ideation continue to increase across society today and is very personal to me. As an active duty Army Soldier of 22 years, I intimately understand the sacrifices of service members in the military today and during combat. In 2019, two of my Soldiers decided to take their life and my closest co-worker attempted suicide. These three gentlemen will have an everlasting impact on my life and are the reasons why this topic is interesting to me. Furthermore, during this study, a mentee made the decision to attempt suicide. Initial research of methods to identify suicidal ideation early led to social media platform intervention strategies, chatbot capabilities and ultimately to discovering the OmniSense ELLIE platform. ELLIE was a remarkable platform and this work will be in direct coordination with the team at

University of Southern California, USC ICT, creator of ELLIE [9]. In 2013, SimSensei & MultiSense: Virtual Human and Multimodal Perception for Healthcare Support analysis provided an in-depth look into what AI has done in this field.

As with all military systems, there are several resources that a Soldier in Crisis should be able to rely on such as ArmyOneSource, Military Family Life Consultant (MFLC), and Clinical Physicians at the locally designated Behavioral Health Clinic. Each of these resources were prompted by an identified need and can provide a certain level of assistance. However, do Soldiers actually feel comfortable reaching out to these resources when they need assistance? Do the resources target the younger generation of Soldiers today? In the electronic-based world of 2021, Artificial Intelligence (AI) and Machine Learning (ML) methods are fully capable of aid in identification of those in crisis as seen through previous demonstrations of technologies such as Woebot [6] and Facebook's Deep Learning technologies [5]. Additionally, the Virtual Human (VH) interviewing platforms such as Ellie created in 2013 [9], seemed to also illustrate this capability initially. AI advancements provide capabilities to analyze the mood of a person by looking at his or her face via Facial Expression Recognition, eye gaze, or IP (image processing); examine a person's voice through Voice Recognition (VR) or analyze a person's text online in real-time. Yet, why did the platform not succeed? With the continuous-use of handheld devices and 'always-connected' generation – AI must be able to break through the barriers to care. Specifically to provide the opportunity of investigating methods that give professionals a technological advantage to identify people in need – possibly prevent suicide attempts and manage mood changes among the people that not only are isolated socially but show signs of depression on their face, their body movements and in their voice. Although there are exponential questions that can be posed to query the current status quo or identify a solution to a specific need, here are a few research questions that are explored and investigated through this work:

- RQ1 – Are the current systems that leadership believe are in place optimized for the newer generation of Soldiers to use?
- RQ2 – What are the specific barriers to care? Judgment, ease-of-access, interaction space, fear of retribution, privacy?
- RQ3 – How can AI/ML be leveraged to close the gap between someone in crisis and receiving immediate care that is appropriate and customized for the individual Soldier?
- RQ4 – What are the most valuable customizations that would solicit future use of a VH platform? Male vs. Female vs. Non-binary vs. Other; Military roles or Civilian; Environment - Inside, Outside, etc.; or other customizations.

A survey, system test and analysis were completed to solve the gap in early identification of depression and suicidal ideation using a virtual human interviewing platform by soliciting honest, open, and current feedback from Soldiers on how to optimize such a system. Benefits of this research provide real-time, relevant, and honest insight as to where Soldiers are currently turning to for care, how can a platform be optimized to encourage interaction and provide the judgment-free interactive space for individuals in crisis in the future. The focus of this research was a survey used to better understand the 'Use of Care' of the younger generation of American Soldiers began with identifying currently applied ML and AI methods in recognizing and identifying symptoms and risk factors across millions of data points prevalent in advanced platforms such as ELLIE or EMPOWER. This was accomplished through examining the related works presented in this document. A deep-dive into chatbots and social media implementations of Deep Learning methods or Machine Learning was conducted and presentations were created. This research was followed by the analysis of adequate methods and identification of areas of improvement based on an

independent experiment that was attempted on the platform and in conjunction with upcoming studies conducted by USC-ICT. Furthermore, due to the isolation of the pandemic, this topic is extremely important and the results from the experiment can be a baseline for future study and the development of an automated system for early identification. This research focused on preparing relevant improvements for future implementation with USC-ICT.

1.1 Background

There are several risk factors identified to contribute to suicidal ideations. The first factor is typically a previous suicide attempt, followed by a diagnosis of a type of depression [1]. Although this is not always the case, this research is examining the link to "Depression". Further exploring this concept, there are several different types of Depression across a full spectrum of severity. Types such as Major Depression, Persistent Depression Disorder (Dysthymia), Perinatal Depression, Seasonal Affective Disorder (SAD), Psychotic Depression, Disruptive Mood Dysregulation Disorder, Premenstrual Dysphoric Disorder, and Bipolar Disorder [1]. Depression indicators considered in this research are not just a 'feeling' of sadness for a day or two, we are looking at a deeper level. Indicators such as [14]:

- Persistent sad, anxious, or "empty" mood;
- Feelings of hopelessness or pessimism;
- Feelings of guilt, worthlessness, or helplessness;
- Loss of interest or pleasure in hobbies or activities;
- Decreased energy, fatigue, or being "slowed down";

- Difficulty in concentrating, remembering, or making decisions;
- Difficulty in sleeping, early-morning awakening, or oversleeping;
- Appetite and/or weight changes;
- Thoughts of death or suicide or suicide attempts;
- Restlessness or irritability;

These indicators for depression and suicidal ideation have many visual, audio and textual cues that can be identified through various AI and ML algorithms. However, to leverage these cues, the individual must interact with the system. To that end, there are many barriers to care that prevent an individual from admitting they need assistance, asking for help or following through to get care when needed. The Human-Computer Interaction (HCI) perspective is critical to ensure that the user is open and honest in communication, which will provide true and realistic data points for analysis. If the environment is optimized the user should feel calmer and not judged by the technology and its responses, thus ensuring the change in cues are not triggered by the agent. Correctly identifying such barriers and providing an alternative to the typical resources are two approaches to close the gap and help the Soldier.

Artificial Intelligence (AI) advances in analyzing the mood of an individual by examining Facial Expression Recognition, IP (image processing), and Voice Recognition (VR) are critical aspects of leveraging technology in this field. This analysis provides the opportunity of investigating methods that give professionals a technological advantage to identify people in need – possibly prevent suicide attempts and manage mood changes among the people that not only are isolated socially but show signs of depression on their face, their body movements and in their voice.

1.2 Related Works

In [7], the Canadian literature review examined the various methods of both supervised and unsupervised learning with regards to identifying risk factors for suicide. The study recommended alterations to existing systems and provided insight into the value that individual interactions can be optimized based on participants' feedback. Although their research was expanded across all areas of treatment including interactions in the office, medical records themselves, Internet of Things (IoT) wearable devices, social media, and conversational agents - there is great value in their analysis.

Social Media platforms such as Facebook, Instagram, Twitter, etc. and individual conversational agents such as Woebot have conducted research for their individualized platforms. Most of the social media research focuses on recommendation systems, commercial advertising, and general text monitoring – less towards identification of risk factors to assist someone in need. However, there are relevant independent studies of the platforms such as [10, 5] for Facebook, which relate the use of knowledge graphs (KGs) to Facebook's Social graph to illustrate contagion of suicides across an area or group of people. Or [3, 12, 4] for Twitter that specifically target suicide identification using deep learning methods and Natural Language Understanding (NLU). The most significant implementation that was examined was the Facebook Deep Learning algorithms [5] that examine posts and responses to identify those in crisis and alert local response teams. It is critical in an AI system such as this, to correctly train and identify sarcasm and common slang phrases to ensure resources are not abused for childish pranks or unintended triggers.

The unblinded study of 70 college students for Woebot [6], provides insights into the interactions of the participants. Although this application is not a virtual human platform, this conversational agent's study provides insight into care through tech-

nology. Participants accessed the application via their cell phones and interacted at a minimum on a daily basis for several weeks. Specifically, the interaction experience feedback from the end of study surveys identified the need for a more inviting environment and the database response recommendations based on the depth of knowledge proved positive for some and not extensive enough for others. This unblinded study was a prime example of the type of responses and feedback that was sought through this thesis and the current survey of this VH platform.

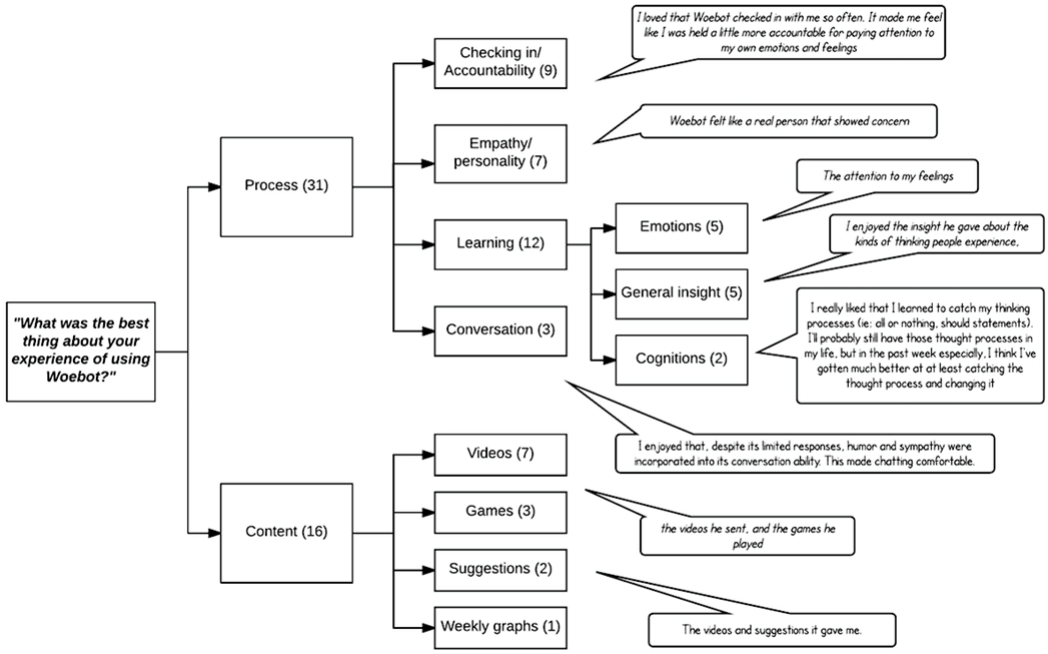


Figure 1.1: Woebot end-of-study participant responses [6]

This work was conducted in direct coordination with the team at University of Southern California, USC ICT, which previously released a video illustration of interaction with a platform named Ellie [9]. In 2013, SimSensei & MultiSense: Virtual Human and Multimodal Perception for Healthcare Support analysis provided an in-depth look into what AI has done in this field. Ellie is now outdated, yet a new prototype is being developed under the new EMPOWER (Enhancing Mental Performance and Optimizing Warfighter Effectiveness and Resilience: From MultiSense to OmniSense)

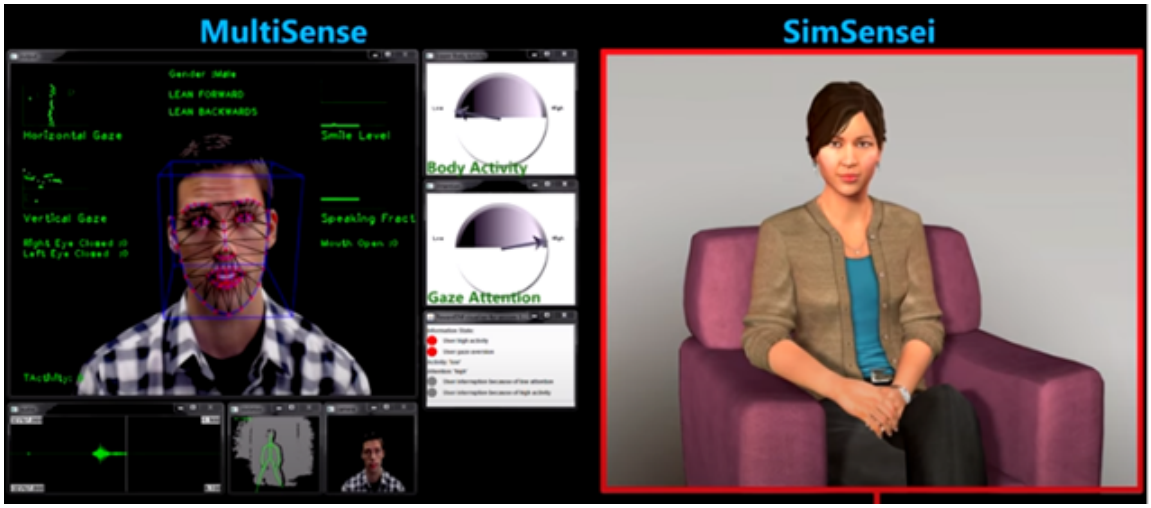


Figure 1.2: ELLIE demonstration [9]

platform. USC-ICT has approved collaboration in support of this thesis and access to the upcoming prototype 'Kevin' of EMPOWER (updated Ellie). Through extensive conversations, it was determined that a 'Use of Care' survey was a pivotal part in identifying the future recommendations to the next prototype. Exactly what do current Soldiers think of such a capability? And 'How can we optimize it?' were two pivotal objectives that were met through the course of this research.

As stated, motivation to research this topic is extensive, however the related works provide various methods to address the largest concern in the area, which is how do we identify people in crisis early. Social media platforms and chatbots are implementing various mechanisms and are reaching a certain group of the population today. On the other hand, the ability to work with a virtual human agent platform and the USC-ICT team to further a military focused project was extremely beneficial for this student, researcher and Soldier.

Chapter 2

SYSTEM DESIGN

The analysis considers the research technologies currently employed within the USC-ICT applications, software, studies, and prototypes and attempted to identify AI/ML methods that could be applied, examined, or improved. The system design is based on each HCI principle and Intelligence attribute, followed by the individual improvements to be integrated with the next prototype. The EMPOWER platform uses the Omnisense server to analyze the data collected while using the unity client as the front end interacting with the user and collecting the data.

2.1 HCI Principles

Human Computer Interaction (HCI) principles are developed and defined to assist technicians with optimizing how a human interacts with technology. With decades of technology advancements, defining the specific principles assists with examining each piece of the VH agent to identify system improvements. The research technologies considered and identified using the following HCI principles for the platform:

Principle #1 – **Cognitive Foundations**, as the mental model of the platform and how well it matches the user's mental model. Through the user's natural perception, memory and knowledge, each platform should be designed to interact naturally and on the same level. Was the platform designed to be used easily? The interaction of the user should be familiar and expected. Limitations of the platform should not be obvious.

Principle #2 – **Interaction Spaces**, the physical and cognitive space in which the user and the automation interact can contribute to the data collected. Through physical,and visual, interaction the interaction can be inviting and solicit more interaction. Goal is that the interaction is pleasing, and solicits the user to continue the interaction.

Principle #3 – **Input-Output Devices** are the methods in which the user interacts with the agent. Audio, visual and textual devices contribute to the direct or indirect interactions. The input-output device used should be optimal for the type of interaction used. A way to address this is to have multiple choices for input-output devices. This is also a consideration for the Americans Disability Act, to ensure there are methods for those that may have reduced capabilities.

Principle #4 - **Interaction Styles** are various methods a user can interact with the system through. These can vary from batch systems, command-line, full-screen interfaces, menus, forms, direct manipulation interfaces, graphical user interfaces, user assistance, speech and mobile devices. This can be further broken down into four parts: articulation, performance, presentation, and observation. These four parts are often viewed as a circular cycle of communication - How well a user is able to articulate the input to the system, the performance of the system to keep up with the required interactions, the presentation of the response from the system back to the user and lastly how well was the response observed by the user.

Principle #5 – **Speech-based Interaction**, elicit additional considerations such as ability to distinguish accents, mitigate noise, the range of auditory signal, transmitting and receiving channels, and encoding. For a human it may be seen as a speaker, through the air (noise) to the ears as the receiver. For a computer, it may be seen as the transmission from a speaker through the air, to a microphone. A few questions are: Did the speech interaction (if available), provide appropriate channels and en-

coding? Was the platform able to distinguish accents? Was the range of signal and mitigation of noise appropriate?

2.2 Intelligence Attributes

In order to define how 'intelligent' or well-developed the machine learning code is within the system, individual criterion are identified and defined. By breaking down the various aspects of the ML agent assists with providing the areas of the system to improve. This survey provided the opportunity to identify and compare the machine learning 'Intelligence' of the platform based on the following attributes:

Criterion #1 **Contextual understanding** as it relates to the intent of the input and incorporating inference to draw conclusions based on known facts. A level of intelligence is required to apply it to compute the intent of a statement - not merely pattern matching (noun and verb).

Criterion #2 – **Conversational design** to handle the input in a conversational way rather than as a transactional platform. Conversational flow is a distinct characteristic of a human conversation; typically we do not jump from one concept to a completely different concept without a transition. Response planning is critical to mimicking human behavior.

Criterion #3 – The **depth of knowledge** is critical. The simple facet of accepting multiple questions at once, offering multiple solutions to the same question or the ability to learn new responses based on context rather than static databases via natural branching requires further programming or complexities than a simple question-answering.

Criterion #4 – **Error detection and handling** is a clear way to see how extensive the natural language processing of a platform extends. The ability to recognize variations such as slang, idioms and colloquialisms is more complex than just a standard dictionary would provide. How the platform reacts to out of range topics is important.

Criterion #5 – **Coreference resolution**, word sense disambiguation and handling of contractions is another level of natural language processing that can be used to analyze the behavior of a platform. Although word sense and contractions should be able to be handled with simple code the coreference resolution is a little more complicated. Being able to determine that the pronoun is associated with a separate part of speech and relate it to that part is more intricate.

Criterion #6 - **Emotional Intelligence (Empathy)**, illustrating empathy through 'human-like' behavior to help individuals virtually to interact based on the 'mood' and 'need' of the conversation. The platform's functionality to make a user feel like they were talking to a real person. The underlying implied ability to 'pay attention' to the user's feelings, by keying in on simple identifiers. And the ability of the Platform to detect implied thought processes or feelings.

2.3 Use of Care

Specifically for Ellie [9], explore opportunities to enhance this PLATFORM, attempt to identify as many recommendations as possible during interaction while comparing to the upcoming prototype of EMPOWER – to provide feedback to the organization:

- If they would seek care? And if so, how would they prefer to seek care?
- If there are any hesitations of AI/ML with their data?
- What might they use a VH for?

- Male vs. Female vs. Non-binary vs. Other.
- Military roles or Civilian
- Environment - Inside, Outside, etc.
- Other customizations

The appealing property of the survey and analysis was the ability to get honest feedback from current Soldiers on how this virtual human platform can be optimized to ensure Soldiers use it once the next prototype is released. The main functional requirements considered for the analysis of the platform are the HCI and Intelligence perspectives of the agent.

- R1. Create a comprehensive survey to gather relevant feedback.
- R2. Conduct a survey to get feedback from at least 50 participants.
- R3. Consolidate all responses and analyze the results.
- R4. Propose improvements for the next prototype system test.

The survey will be used by researchers in support of the EMPOWER prototype 'Kevin' scheduled to conduct a system test this fall in southern California. Evaluation criteria for the above listed requirements are:

- E1. Have an advisor and technical team review surveys and integrate changes (R1).
- E2. Solicit feedback from 5 Soldiers prior to survey delivery and ensure questions are comprehensive and return relevant feedback (R1).
- E3. Did the survey return more than 50 responses? (R2)

- E4. All responses are combined into one complete document to provide an inclusive 'survey results'. (R3)
- E5. Documentation that analyzes the results and provides direct and clear insight into all responses. (R3)
- E6. Prototype characteristics are validated by results or a list of improvements are developed and delivered to the integration team.

The system design was based on the 5 HCI principles and 6 Intelligence attributes created, defined and implemented in this research. A custom survey was created and conducted to examine the use of care perspective while meeting the functional requirements and the established evaluation criteria. The implementation of the survey is further described in the next chapter.

Chapter 3

IMPLEMENTATION

3.1 Survey

Soldiers will volunteer for a 15 minute session – this will be published by Military Leadership through their formal notification systems. Session will start with a review and signature of the consent form (see Appendix 1). Following a positive acknowledgment on the consent form the participant will then review of the background and motivation of the overall research provided in Appendix 2 (below). Subsequently the Soldiers will watch a 3 minute video of the ELLIE agent from YouTube - Click Here [9]. The process continues by the Soldier completing a 34 question survey on how they would like to access care and if it is a virtual agent – how can it be optimized. (See Appendix 3). This was completed digitally by accessing a link and submitting a Web Form which was saved to a restricted Google Drive directory.

The survey was followed by in-depth analysis of all 51 individual results to ensure every response was considered and the recommended changes considered. The analysis and recommended changes were examined using the next prototype that USC-ICT is currently developing. This next prototype is continued work of the original ELLIE which is referenced as Kevin using the EMPOWER platform, specific improvements will be identified and returned to the development team to be included in the next revision.

3.2 Testing and Validation

The validation of this research is seen through the comprehensive survey consisting of 34 well developed questions soliciting relevant feedback from Active Duty Soldiers. The survey received 51 responses. The consolidated responses illustrate several neutral responses that tend to come from a complacent automated 'click' response when the question doesn't trigger an emotional or enticing engagement to get a positive or negative response. This may be due to the technical nature of the questions and the common understanding of a Soldier. The questions were validated through five Soldiers, however, it is assumed that the review didn't receive the in-depth criticism it was requesting. All neutral responses will be separated to see if the remaining results can identify any polarized improvements for future improvements with the platform system test to be conducted this summer. Evaluation of the results of this study will be seen through analysis of the survey results and system improvements defined. The evaluation criteria was met or realized through the following methods:

- E1. Have an advisor and technical team review surveys and integrate changes (R1).
  - Advisor and technical team reviewed the protocol, research questions, and survey. Their recommendations were integrated into the survey prior to soliciting responses.
  - Survey results are consolidated. The technical team reviewed the initial results during the technical meetings during the summer of 2021. Recommended integrated changes will be identified, discussed and documented for further work in the Fall of 2021.

  - As the minimum survey results were just met due to delay in survey approval - the analysis and recommended improvements will be completed during October 2021.

- E2. Solicit feedback from 5 Soldiers prior to survey delivery and ensure questions are comprehensive and return relevant feedback (R1).
  - Received feedback from 5 Soldiers prior to the IRB approval. Integrated the suggested 'wording' adjustments to make the survey more fluid and understandable. Removed the first page as it was redundant.
  - Currently received 51 responses.
  - Survey continued through the months of June and July, and all results were consolidated in October.
  - Integration with West Point was delayed due to lack of funding and the USC IRB process. Ultimately it was cancelled and a new integration testing is currently being designed and sent for approval for Southern California.

- E3. Did the survey return more than 50 responses? (R2)
  - Collected 51 results.

- E4. All responses are combined into one complete document to provide an inclusive 'survey results'. (R3)
  - All 51 survey results are combined and analyzed in this work.

- E5. Documentation that analyzes the results and provides direct and clear insight into all responses. (R3)
  - Again, this is most beneficial following the consolidation of all survey results.

- E6. Prototype characteristics are validated by results or a list of improvements are developed and delivered to the integration team.
  - A List of improvements are provided in the Future Work section of this thesis and provided to the USC-ICT team for consideration of the new prototype and upcoming system test. Such improvements are the additional confidence in security of the system by updated user agreement created as consent forms that detail the security of the system, data retention and intended use of data. Additionally, the options to customize the settings, agent, and interaction space are provided - however, implementation requires additional funding and change to the current scope of work of the USC-ICT project.

The implementation of the survey provided an opportunity to solicit the desired responses from 51 participants. The responses provided were beneficial in analyzing the barriers to care and areas for improvements. These barriers and areas for improvements are seen through the results and analysis provided in the next chapter.

Chapter 4

RESULTS

Here are the 51 survey responses summarized results. The survey results are consolidated, enumerated, and normalized to account for the typical complacent neutral responses. This will ensure that every thoughtful and genuine response is considered and integrated into the recommended solutions. Additionally, since the survey is based on the viewing of two videos, the neutral responses could account for the principle or attribute not being determined or perceived in the review. The results will be analyzed in the three distinct categories of Barriers to Care Optimization, HCI principles, and AI Intelligence attributes.

4.1 Barriers to Care

There are several barriers to care for every individual, yet this survey set out to attempt to identify some specific barriers for the American Soldier to seek care and possibly seeking care through a virtual platform.

4.1.1 Seeking care

The first two questions of the survey addressed the questions if a participant would seek care and how they would prefer to access care. Out of the 51 individuals 21 stated they would seek assistance if they had a life crisis or significant event occur. 10 individuals responded with 'No' they would not seek care, while 19 said 'Maybe' and only 1 person saying not sure. This illustrates that people with a strong opinion

of 'Yes' or 'Maybe' are about equal and there are a large group that remain adamant they would not reach out for care. This is the key population for this thesis and study - as if we can leverage a platform to make these people more comfortable with getting care, we can make a difference.

37 of 51 participants stated that they prefer to go in person to a clinic to receive trusted health care information. When compared to the first responses of generally seeking care, this percentage is slightly surprising. Although only 41% stated they would get care, over 72% said they would go in person. This question was presented where the participant could select multiple methods for care, of which only 23.5% responded they would use an app or VH agent. Although that number is slightly lower than desired, this still provides hope for the goal of this study and a chance to optimize the platform to encourage more participation.

4.1.2 Security Concerns

Next the survey addressed concerns with security of interacting with an automated system such as an application or VH agent platform. 37.3% responded with a neutral response, 7.8% stated they felt it would be 'Very secure with no concern', 21.6% responded it was 'Secure', 19.6% stated it was 'Slightly not secure' and 13.7% stated it "Not Secure (my information is vulnerable). Following the 'scaled' question on security, there was the opportunity to provide a custom response as to what concerns they have. This question received 29 responses with 6 being neutral. 17 responses were related to confidentiality, gossip, being hacked, or not knowing who will see the interaction or recorded sessions being leaked. Only 2 responses had to do with the VH not being able to relate or the interaction space of the platform.

There was a second subsection within security to gauge how honest the participant might be with a VH agent and what hesitations they may have. Majority of responses stated they believe they will be very open and honest with the virtual agent. The 25 custom responses included 4 neutral, 1 concerned with being honest with oneself, 2 concerned with backlash or judgment from coworkers, 2 addressed that the technology is still under development and might be prone to errors, 8 stating the machine 'can't relate' or simply responding to 'buzzwords', with the remaining 8 responses were concerned with where the information will go or who will have access to it.

4.1.3 Types of Care

Security was followed by asking for what purposes a participant may use the VH agent for. The question was formed to provide each participant the ability to select as many options as they would consider using with no limits. Out of the 51 responses 12 participants or 23.5% selected 'none' which can reduce the overall percentage of the other responses to this question even more. Out of the remaining 39 participants that might actually use the VH agent platform here are the responses. The most selected response was the use of the VH agent as 'just someone to talk to' receiving 21 responses or 53.9%. This was closely followed by two other choices. The first was very similar as 18 responses selected Resilience or Mindfulness techniques at 46.1%. However, also receiving 18 responses or 46.1% was to receive 'Stress-relief in high-stress scenarios'. 'Just help' received 12 responses or 30.7%. Crisis mitigation received 14 responses or 35.9%. Lastly, 25.64% or 10 participants stated they may use the agent to find other resources.

4.1.4 Optimizations

The final section of the barriers to care section solicited responses on possible methods to optimize the agent and surroundings to be more inviting. Specifically the question was formed as 'Are there any specific characterizations that would make your interaction better? (select all recommended preferences you would like to see integrated)'. The first four options were the gender of the agent, where the responses were: Male = 7 or 13.7%, Female = 21 or 41.2%, Non-binary = 3 or 5.9%, other gender = 0. Next two options provided options for the clothing of the agent where the majority sided with how Ellie was currently designed wearing civilian clothing with 20 responses or 39.6% and only 3 participants selecting a military uniform. This was followed by the rank or position of the agent with 5 participants desiring a leadership position while 8 selecting a 'buddy rank' or that the agent be the equal rank of the participant. The surrounding was formed inside an office setting with 15.7% or outside atmosphere with 37.3%. Last, the survey received some custom responses to the optimization of the agent which were free customization options (like an avatar), just wanting more in-depth responses, or feedback such as 'too robotic' requesting more casual movements.

4.2 HCI Principles

There are five HCI principles examined throughout this work and each had multiple questions formed to support receiving insightful feedback to determine if there is room for improvements. The questions were formed on a scale of 1 to 5, 1 as best and 5 as worst. Given an odd number of options, 3 is neutral and those responses will be counted as the participant did not understand the question, did not witness the principle, or did not care to provide a distinct response.

Principle #1 was based on identifying if the agent's cognitive foundations were adequate. This was formed as three questions which attempted to determine if the mental model of the VH agent was correctly aligned with the mental model of the participant. Through their natural perception, memory and knowledge – the virtual agent should be designed to interact naturally and on the same level. The questions were to identify if the interactions seemed natural, familiar and if the agent's limitations were obvious in the video clips. Natural behaviors received 20 neutral (3) responses, familiar interactions received 21 neutral (3) responses, while obvious limitations received 25 neutral (3) responses. The percentages below in Table 4.1 have these neutral responses removed.

Table 4.1: HCI Principle 1 Results

Question	1(best)	2(good)	3(neutral)	4(not well)	5(bad)
Natural/Easy Interact	12(38.7%)	15(48%)	20	3(9.6%)	1(3.2%)
Familiar	9(30%)	11(36%)	21	3(10%)	7(23.3%)
Limitations Obvious	3(11.5%)	10(38%)	25	9(34.6%)	4(15.3%)

Following the three questions that support principle #1, the survey requested the participants to provide any additional insight or recommendations for improving the agent's ease of use and natural interactions. This was a free-text box that received 12 responses from the 51 participants. Of the 12 responses two were 'none', one was that the virtual agent 'weirded' the participant out, while the other six provided some insight. One participant recognized how this capability could benefit the clinician by keying in on all of the factors at once. While another identified that Ellie did very well with the icebreakers but still concerned if the therapy question/answers would be as fluid. The last five responses stated that the agent's emotions could seem more fluid and that Ellie's speech seemed too robotic, scripted, limited and over compensating. These comments will be explored further in the optimization and future work.

Principle #2 addresses interaction spaces, which are the physical and cognitive space in which the user and the automation interact. This is seen through physical, and visual interaction, which if the interaction is inviting it can solicit more interaction. This was asked as three questions, if the interaction was pleasing, if they would continue the conversation, and if they felt less judged by the VH agent than a real person. Similar to the principle #1, the neutral responses are included in the table but are removed from the percentage calculation in order to illustrate the positive and negative polarization of the actual responses. 1-5 were similar but had different descriptors for each question (very pleasing -1, not pleasing -5; could talk all day -1, couldn't end soon enough; judgment-free -1, judged more -5). As seen below, in Table 4.2 there was a shift in the desire to continue a conversation with the agent to the negative responses. On the other hand, we see the larger shift to the positive as participants tended to feel less judged.

Table 4.2: HCI Principle 2 Results

Question	1(best)	2(good)	3(neutral)	4(not well)	5(bad)
Pleasing	8(28.5%)	13(46.4%)	23	5(17.6%)	2(7.1%)
Continue Conversation	4(12.5%)	11(34.3%)	19	11(34.3%)	6(18.7%)
Judgment	22(56%)	12(30.7%)	12	2(5.1%)	3(7.7%)

Following the three questions that support principle #2, the survey again provided an opportunity for the participant to provide any additional insight or recommendations for improving the agents' interaction spaces: list any physical, and visual, changes that could be made to make the interaction more inviting. This was a free-text box that received 15 responses from the 51 participants. There were two placeholder responses of 'none' and 13 valid responses.

- "Have more options for scenery. And add an option on who you can talk too"
- "Call me by my name"

- “The agent provides an environment without any judgment. Whereas humans are more likely to unconsciously judge.”
- ”When being judged I think it is difficult because it is analyzing your face and movements but those are different for everybody in what they mean.”
- “It’s not a question of judgment... But more knowing that this was not human to human interaction.”
- “Have more emotional responses. Give the AI a sense of care.”
- “More physical mannerisms”
- “Very unnatural movements, appearance and poor choice of character and language used”
- “The information is still being recorded and recorded kept, hence hesitant to use with anything digital”
- “I don’t know if the automation would be able to handle an unexpected input - similar to Siri - human emotions can sometimes be unexpected”
- “My biggest recommendation would be to avoid a solid white background. Solid white adds to the awareness of being in a virtual environment and I feel the addition of more ”homey” or office-like aesthetics would be welcome and make the encounter more enjoyable. I would also recommend adding in additional movements by the virtual agent as a participant is responding as a static-like appearance just feels unnatural.”
- “Due to my TBI sounds and sound level are of more impact than I would’ve ever thought. That said, I have a rain machine that plays while I sleep and it really helps. So, I’d explore additional settings to offer patients and/or allow

the patient to do little modification of the agents setting such as couch color, wall color, turn the agent gestures (on/off/custom)."

Principle #3 addresses Input-Output (I/O) Devices, which are the methods in which participants interact with the agent. Audio, visual and textual devices contribute to the direct or indirect interactions. This was asked as two questions, if the I/O devices were optimal and if there were multiple options. As the results below clearly show, this was difficult to determine based solely on watching a video, which led to a large shift to the neutral response of 3. The neutral responses are included in Table 4.3 but are removed from the percentage calculation in order to illustrate the positive and negative polarization of the actual responses.

Table 4.3: HCI Principle 3 Results

Question	1(best)	2(good)	3(neutral)	4(not well)	5(bad)
I/O optimal	9(28.1%)	17(53.1%)	19	5(15.6%)	1(3.1%)
Multiple I/O	7(33.3%)	8(38.1%)	30	4(19%)	2(9.5%)

Following the two questions that support principle #3, the survey again provided an opportunity for the participant to provide any additional insight or recommendations for improving the agents' I/O devices. This was a free-text box that received 8 responses from the 51 participants. Similar to the other text boxes the survey received three 'none' and then the following 5 responses:

- "Depending on what the actual requirements are for a person to interact, it would be best to just be able to talk to a computer screen and the typical camera be able to pick up all indicators. Meaning - no additional sensors being required to make it easier for all Soldiers to use the virtual human."
- "I wouldn't feel comfortable talking to a computer"
- "the graphics could resemble more human feature"

- "Need more options"
- "I'm sure the system is optimized for touchscreen devices. Perhaps the system could allow for the interaction to start via keyboard and mouse to then see if the patient could benefit or prefer touch screen and then provide a free/low cost tablet."

Principle #4 addresses interaction styles presented by the agent which are various methods a user can interact with the system through. These can vary from batch systems, command-line, full-screen interfaces, menus, forms, direct manipulation interfaces, graphical user interfaces, user assistance, speech and mobile devices. This can be further broken down into four parts: articulation, performance, presentation, and observation. This was asked as three questions based on the video demonstration, if the participant felt like it would be easy to articulate what it wanted to convey to the agent, if the agent could keep up with the conversation and if the agent's responses were easy to understand. Same as all previously presented responses the neutral responses are included in Table 4.4 but are removed from the percentage calculation in order to illustrate the positive and negative polarization of the actual responses. The results for this set of questions seemed to receive very similar responses for all three questions from the participants, all having 19 or 21 neutral responses, and polarized to the positive spectrum.

Table 4.4: HCI Principle 4 Results

Question	1(best)	2(good)	3(neutral)	4(not well)	5(bad)
Articulate	9(30%)	16(53.3%)	21	4(13.3%)	1(3.3%)
VH keeps up	8(26.7%)	17(56.7%)	21	4(13.3%)	1(3.3%)
Easily understood	17(53%)	11(34.3%)	19	3(9.3%)	1(3.1%)

Following the three questions that support principle #4, the survey again provided an opportunity for the participant to provide any additional insight or recommendations

for improving the agents' interaction styles. This was a free-text box that received 7 responses from the 51 participants. Four responses stated 'none', and the other three provided feedback:

- "The agent did well. Although they should be able to give more insight on the feelings to your response."
- "Improve the agent's appearance to one that is more natural."
- "People often lie to themselves just to get through the day, at some point those lies are so practiced that you can find the truth just by observing them. When I first came out about my depression no one understood because I was always such a happy inclusive person who always made others feel good about themselves. Of course some only see what they want to see and a machine has no personal desires so it might pick up minute cues that real people don't. I don't know if this will help but I sincerely hope your project works."

Principle #5 addresses speech-based interaction, with additional considerations such as ability to distinguish accents, mitigate noise, the range of auditory signal, transmitting and receiving channels, and encoding. For a human it may be seen as a speaker, through the air (noise) to the ears as the receiver. For a computer, it may be seen as the transmission from a speaker through the air, to a microphone. Again, neutral responses are included in Table 4.5 but are removed from the percentage calculation in order to illustrate the positive and negative polarization of the actual responses. The increase in some neutral responses could be due to the fact that the participant felt they couldn't identify it through the video, such as detecting accents was not distinguishable or represented in the short video.

Following the three questions that support principle #5, the survey again provided an opportunity for the participant to provide any additional insight or recommendations

Table 4.5: HCI Principle 5 Results

Question	1(best)	2(good)	3(neutral)	4(not well)	5(bad)
Understandable	15(44%)	15(44%)	17	3(8.8%)	1(2.9%)
Accents/Slang	5(25%)	12(60%)	31	2(10%)	1(5%)
Noise Mitigation	16(45%)	16(45%)	16	2(5.7%)	1(2.9%)

for improving the agents' speech-based interaction spaces. This was a free-text box that received six responses from the 51 participants. Of which, four were 'none', one is to have options to select your agent type and appearance which didn't directly relate to principle #5 and the last one stated "Not sure about accents, I am very articulate.".

4.3 Intelligence Attributes

Throughout Artificial Intelligence there are various different descriptions trying to define what exactly is intelligence and how it is determined. In this study, there are 5 attributes identified to help form the concept of how intelligent a virtual human platform may seem. The five attributes are followed by the survey results of which have a large number of neutral responses. This may be due to these being presented at the end of a 34 question survey and the complacency that sets in or the uncertainty of the participants. Nevertheless, all responses seemed to lean towards a positive somewhat intelligence being conveyed by the platform.

Attribute #1 is contextual understanding as it relates to the intent of the input and incorporating inference to draw conclusions based on known facts. A level of intelligence is required to apply it to compute the intent of a statement - not merely pattern matching (noun and verb). Did the Virtual Agent understand your input was asked to the participants with a scale 1 that the agent seemed very intelligent and

could infer the intent of a response to 5 being least intelligent. The responses in the table below illustrate that Ellie seemed somewhat intelligent.

Attribute #2 – Conversational design to handle the input in a conversational way rather than as a transactional platform. Conversational flow is a distinct characteristic of a human conversation; typically we do not jump from gone concept to a completely different concept without a transition. Response planning is critical to mimicking human behavior. Did the Virtual Agent illustrate a conversational design within the responses? Can the Virtual Agent stay on topic? How many branches does the Virtual Agent have to a specific question? These questions seemed to receive a slightly more neutral response, however still positive.

Attribute #3 – The depth of knowledge is critical. The simple facet of accepting multiple questions at once, offering multiple solutions to the same question or the ability to learn new responses based on context rather than static database via natural branching require further programming or complexities than a simple question-answering. Can the Virtual Agent accept more than one question at a time? Can the Virtual Agent learn new responses based on conversation correctly? Can it plan responses from future inferences? Out of all 6 attributes this was the one to receive the most neutral responses. This may strictly be due to the length of the demonstration of three minutes.

Attribute #4 – Error detection and handling is a clear way to see how extensive the natural language processing of a Virtual Agent extends. The ability to recognize variations such as slang, idioms and colloquialisms is more complex than just a standard dictionary would provide. How the Virtual Agent reacts to out of range topics is important. Can the Virtual Agent detect slang? Can the Virtual Agent correctly identify misspellings? Can the Virtual Agent coach the user to clarify – to recover

from out of scope questions? Closely following behind attribute #3, this received 27 neutral responses with the remaining polarizing towards the positive.

Attribute #5 – Being able to determine that the pronoun is associated with a separate part of speech and relate it to that part is more intricate. Can the Virtual Agent correctly identify the correct meaning of the ambiguous terms (such as contractions or slang)? Can the Virtual Agent resolve the 'he' to reference the male that was mentioned three sentences before to provide the correct responses? This attribute is following very closely with Attribute #1 for the positive polarization of a total of 24 participants identifying it as intelligent vs 27 participants for #1.

Attribute #6 – Emotional Intelligence (Empathy), illustrating empathy through 'human-like' behavior to help individuals virtually to interact based on the 'mood' and 'need' of the conversation. Did the Virtual Agent make you feel like you were talking to a real person? Did the Virtual Agent 'pay attention' to your feelings? Can the Virtual Agent detect your thought processes or feelings? Lastly, empathy seems to be difficult for people to accept from a virtual platform. Ellie did solicit 20 responses displaying empathy and the open text responses allowed for several participants to voice that they do not feel that a machine can truly understand or empathize.

Table 4.6: Intelligence Attributes Results

Attribute & Question	1(best)	2(good)	3(neutral)	4(not well)	5(bad)
#1 Understanding	7(21.1%)	20(60.6%)	18	5(15.1%)	1(3%)
#2 Conversational Design	8(28.5%)	14(50%)	23	4(14.3%)	2(7.1%)
#3 Depth of Knowledge	7(31.8%)	10(45.4%)	29	4(18.1%)	1(4.5%)
#4 Error handling	5(20.8%)	12(50%)	27	6(25%)	1(4.1%)
#5 Ambiguous Term/Slang	6(22.2%)	18(66.7%)	24	1(3.7%)	2(7.4%)
#6 Empathy	5(18.5%)	15(55.5%)	24	5(18.5%)	2(7.4%)

As seen in Table 4.6 above and the calculations with the neutral responses removed, all attributes are polarized significantly to the positive or 'intelligence' perspective. Attribute 5 is the highest percentage at 88.9% voted intelligence as for the ability to

resolve ambiguous terms or understanding slang. Closely following is attribute #1 which is the VH agent ability to convey understanding which was rated intelligent 81.7%. The agent's conversational design was rated at 78.5% intelligent while its depth of knowledge came in at 77.2% intelligent. The two lowest attributes were Empathy at 74% and Error Detection at 70.8%. As a whole, the attributes were collectively weighted at 78.5% viewed intelligence with a mean of 77.8. Following the six intelligence attributes the survey again provided an opportunity for the participant to provide any additional insight or recommendations for improving the agents' intelligence. This opportunity only received 8 responses, four being 'none' and the other four were:

- "The AI is probably better than a human but you can't replace human empathy or relationships."
- "Small talk and its injection into the conversation at random points would add to the feeling of intelligence of the virtual platform. Starting off the conversations with small talk is great but it also needs to be thrown in from time to time rather than jumping directly to the next question."
- "Will be limited due to lack of understanding on several levels"
- "Maybe add a preference for people- if they want the cold distant automated version or the warmer more human-like interaction."

4.4 All Inclusive Results

Comparatively, rather than removing all neutral responses, this section explores all results with neutral (3) responses with the median scores examined. The complete percentages in Table 4.7 include all neutral responses.

Table 4.7: All Inclusive Results

Question	1(best)	2(good)	3(neutral)	4(not well)	5(bad)
Natural/Easy Interact	12(23.1%)	15(28.8%)	20(40.4%)	3(5.8%)	1(1.9%)
Familiar	9(17.3%)	11(21.2%)	21(42.3%)	3(5.8%)	7(13.5%)
Limitations Obvious	3(5.8%)	10(21.2%)	25(48.1%)	9(17.3%)	4(7.7%)
Pleasing	8(15.4%)	13(25%)	23(44.2%)	5(11.5%)	2(3.8%)
Continue Conversation	4(7.7%)	11(21.2%)	19(36.5%)	12(23.1%)	6(11.5%)
Judgment	22(44.2%)	12(23.1%)	12(23.1%)	2(3.8%)	3(5.8%)
I/O optimal	9(17.3%)	17(32.7%)	19(38.5%)	5(9.6%)	1(1.9%)
Multiple I/O	7(13.5%)	8(15.4%)	30(59.6%)	4(7.7%)	2(3.8%)
Articulate	9(17.3%)	16(30.8%)	21(42.3%)	4(7.7%)	1(1.9%)
VH keeps up	8(15.4%)	17(32.7%)	21(42.3%)	4(7.7%)	1(1.9%)
Easily understood	17(32.7%)	11(21.2%)	19(38.5%)	3(5.8%)	1(1.9%)
Understandable	15(28.8%)	15(28.8%)	17(34.6%)	3(5.8%)	1(1.9%)
Accents/Slang	5(9.6%)	12(25%)	31(59.6%)	2(3.8%)	1(1.9%)
Noise Mitigation	16(30.8%)	16(30.8%)	17(32.7%)	2(3.8%)	1(1.9%)
#1 Understanding	7(13.5%)	20(30.8%)	18(36.5%)	5(9.6%)	1(1.9%)
#2 Conversational Design	8(15.4%)	14(26.9%)	23(46.2%)	4(7.7%)	2(3.8%)
#3 Depth of Knowledge	7(13.5%)	10(19.2%)	29(57.7%)	4(7.7%)	1(1.9%)
#4 Error handling	5(9.6%)	12(23.1%)	27(53.8%)	6(11.5%)	1(1.9%)
#5 Ambiguous Term/Slang	6(11.5%)	18(34.6%)	25(48.1%)	1(1.9%)	2(3.8%)
#6 Empathy	5(9.6%)	15(28.8%)	24(46.2%)	5(9.6%)	2(3.8%)

4.4.1 Median

This evaluation provides an alternative perspective for examining the participant responses received by including all results and not removing the neutral responses. This examination is conducted to see if any of the analysis might provide different results. All results in Table 4.8 provide a median score calculated with all responses considered.

The scoring within the survey was 1 being best or great, 2 is good, 3 is neutral, 4 not well, and 5 was the worst of bad. When calculating the median all values were multiplied times the number of responses received in that column and then divided by the total responses of 51. It is interesting that all but 2 questions were polarized to the positive side.

Table 4.8: Median Results

Question	Median Score 1-5
Natural/Easy Interact	2.3
Familiar	2.76
Limitations Obvious	3.02
Pleasing	2.61
Continue Conversation	3.16
Judgment	2.06
I/O optimal	2.45
Multiple I/O	2.73
Articulate	2.45
VH keeps up	2.47
Easily understood	2.22
Understandable	2.22
Accents/Slang	2.65
Noise Mitigation	2.19
#1 Understanding	2.47
#2 Conversational Design	2.57
#3 Depth of Knowledge	2.47
#4 Error handling	2.73
#5 Ambiguous Term/Slang	2.57
#6 Empathy	2.69

The two questions that received values within the neutral spectrum were 1.) If limitations of the platform were obvious which returned a median of 3.02 and 2.) If the platform solicited the participants desire to continue a conversation receiving a median of 3.16. These were somewhat expected as most Soldiers are not accustomed to interacting with a Virtual Human Agent and it is not expected for the original Ellie platform to have no obvious limitations. However, these results provide the insight to enhance the platform in the next prototype to reduce these limitations. On the other hand, the top five questions that received the lowest median scores (or the best feedback) are: If the participant felt less judged by the VH agent than a real person with a 2.06 median, Noise mitigation with 2.19, Easily Understood by the system at 2.22, the system was understandable also at 2.22, followed by the interaction space was natural and easy to interact with receiving a 2.3 median. The most significant

of these five is that the participant felt less judged by the Virtual Human platform when compared to a human receiving a 2.06 median score. This was the most positive question throughout the entire survey which supports the need to continue research in providing necessary care rather than solely focusing on performance which is the primary focus of the military funded research currently.

4.4.2 Complacency

With all surveys there is the possibility for a survey participant to become complacent and bored with answering questions and just begin to mark neutral on all questions. This section is intended to identify the responses that answered neutral for all questions and those that hit a certain point and began to click through the answers. This is analyzed by initially identifying 14 responses that had the last six questions all marked with neutral responses. Below these 14 are broken down into 3 different categories.

The first category identifies 6 participants out of the overall 51 survey participants, that selected the neutral responses for all 22 questions that were measured on a 1 to 5 scale. 12% of the survey responses were completely neutral showing no interest in providing a unique perspective within their responses. Of the 6 participants 5 had responded with 'maybe' when asked if they would seek care if they had a life crisis, while only one responded yes. Lastly, out of the 10 opportunities to provide custom responses within free-text fields, all of these 6 participants provided no responses. In summary, these six participants provided a 'check the box' type of support to this project.

The second category identifies 5 participants or 9.8% of the results have 18 to 21 questions marked as neutral. These results identify a few participants that likely had

no understanding of AI, HCI, or computer background and simply did not understand most of the questions. They did provide polarized responses for judgement and ease of use. However, any of the technical responses are neutral. Additionally, 2 of these 5 participants did provide feedback in the custom fields but were very simple responses such as: comparing the platform to Siri or trusting an application in general.

The final category within the complacency setting identifies the final 3 participants that started marking a neutral response for all questions following the Input/Output questions in the middle of the survey. This identifies participants that provided unique and custom responses to the first half of the survey questions. The I/O questions were in the middle of the 34 question survey. At this point, these 3 participants just wanted the survey to end. They did not provide any feedback in the free-text box following the beginning of the neutral responses.

Overall, through the analysis of the results of the survey provided the ability to examine the HCI principles and Intelligence attributes of the virtual human platform through impressive data collection repository and data analysis. The results provide the development team insight for future integration with the next prototype. Additionally, this survey and responses support the need for additional funding of this project to adjust the scope of work to include AI/ML methods with the dialogue department and identified customization's.

Chapter 5

FUTURE WORK

The goal for this survey was to obtain at least 50 volunteers to provide honest recommendations, and this was exceeded by reaching 51 responses. The survey results have been consolidated, enumerated, and normalized to account for the typical complacent neutral responses. This has ensured that every thoughtful and genuine response is considered and integrated into the recommended solutions. The results were analyzed in the three distinct categories of HCI principles, AI Intelligence, and Barriers to Care Optimization as described above. These results identify several areas of improvements for future work to include security, customizations and error handling. The largest takeaway is the scope of work for the current research is largely restricted to the lack of funding to fully incorporate the feedback from this survey.

5.1 Gaps/Solutions

Leveraging the individual feedback in all three categories, the analysis included the evaluation of gaps and recommended solutions provided by the individual participants of the survey. Each gap and solution have an individual subsection below to provide deep thought, analysis, and evaluation of the proposed idea.

5.1.1 Barriers to Care - Security

Security of information collected by the system was the primary gap that solicited the most comments through the survey. Specifically, the fact that a Virtual Human Agent

system is recording everything makes this topic extremely important. Most survey participants called attention to the fact that a real life health care provider can only perceive, recall, and annotate a limited amount of data. That data is minimal when compared to the data collected through the virtual human agent. Additionally, improper classification and labels with relation to data security were the second concern identified through the survey results. These gaps will be addressed by identifying how the data security and classification within the system are designed. Basic information technology security mechanisms to ensure availability, integrity and confidentiality are in place within this virtual human agent platform. Once this system is used in production with real participants and collecting individual medical information additional security measures will be in use to adhere to all regulatory requirements that govern the security of electronic medical information such as HIPAA. A few research questions developed to address these concerns are:

- What mechanisms or security measures are currently in place to ensure confidentiality?
  - Availability - Maintaining the system data, Local vs Cloud storage, Backup or duplication, Releasability, defense in depth against DoS or DDoS attacks, etc. Deduplication?
  - Integrity - Encryption (whole disk vs. individual record), administrative privileges (least privilege policy, separation of duties), password security, access management, etc.
  - Confidentiality - HIPAA, release of information need-to-know and data retention policies, staff and personnel. Is any of the data anonymized? Special layers of confidentiality when discussing sensitive health care information, [6] [8]
  - Privacy - The ability for the person to be 'left alone [2]. HIPAA

Currently being used solely as a research platform, EMPOWER is secured when data is transferred with encryption and is only available on the local USC-ICT internal network. The network does require authentication, has tiered administrative privileges that do use the concepts of least privilege and separation of duties to ensure research data is not used inappropriately. Additionally, the server that hosts the data of the system is using local storage that is password protected with strict password policies, is backed up regularly, has updated intrusion detection and virus protection. Additionally, to ensure confidentiality of the data collected via the research, each participant is assigned a unique ID that anonymizes the user's name from the data that is stored on the protected server which is further described below.

- How can we clearly articulate these mechanisms and measures to the user prior to first use?
- Formulate a straightforward, short, clear statement that summarizes the above capabilities.

There are now individualized consent forms for each participant that contain common language that provide clear, concise, straightforward statements. Every participant must sign a descriptive consent form and 'opt-in' to the level of sharing that they authorize for their data collected through the VH agent. The minimal consent is 'Research only', meaning that their data is to be used for the specific research only, their recording will be anonymized by 'scrubbing' to remove facial and verbal identifiable information to obscure the identity while maintaining the heart rate, eye gaze, and other informative data to assist improving the agent. With the other extent of the spectrum to authorize full use of the interaction such as using the actual recording and release on YouTube.

- How is this articulation ensuring they know their information is secured and will not be shared, further than just stating HIPAA?
- Develop a way to reassure the participants.

The specific security mechanisms are outlined in the consent form following the option in which they select when interacting with the system. Since the system is focused on and only approved for research purposes the disclosures and reassurances of HIPAA will be integrated with the production edition. It will be reinforced that all data collected from the system can be maintained in a personal health record separate from the video collected. The actual interaction and deductions from the system can be annotated similar to what a physician would transcribe. The transcription would be handled exactly how personal medical and psychological records are secured and maintained regularly.

5.1.2 Customizations

Beyond the concern of their interactions being shared the HCI principles solicited numerous responses requesting additional customizations of the platform to enhance the interaction space. A few research questions in regards to this are:

- Are there already built in options to customize on the next prototype?
- How difficult is it to have the platform provide individual customization options?
- Can the agent, environment, lights, and sounds be customized?

Due to current restraints of COVID, financial support and scope of the IRB approvals the current research platform is focused on developing the proof of concept as one agent named Kevin that is focused on specific results. If the initiative were to gain

traction with a significant increase to funding additional customization could be further developed. Previously there were multiple options started through their NVBG toolkit but were discontinued due to falling outside of the current scope of work. They had started to develop characters for the frontend to include Avery, Ben, Bradley, Ellie, Kevin, Matt, Rachel, Harmony, Pedro, Rio and Utah. Technically speaking it is fairly easy to adapt the current configuration to include options for the agent, environment, lights and sounds to be adjusted - but currently out of scope of the approved project.

5.1.3 Agent Optimizations

To further enhance the individual experiences, what other agent optimizations can be included or improved? What options can be optimized to solicit continued conversations? Below are a few items to be explored:

- Address the participant by name
- More scenery options
- Provide toggle options for additional options
- Improve the emotional responses of the agent by including more physical mannerisms
- Remove the less natural movements

The USC-ICT Simsensei team has been focused on making such improvements to the EMPOWER system based on the original feedback from ELLIE. Through individual interactions with the current prototype they have been extremely successful in these

improvements. Items such as addressing the participant by name is a simple configuration of the system as it starts. It can be easily added to the initial questions as the system starts up. Scenery and toggle options align with the answer to the previous set of research questions that deal with the scope of this specific research project. Lastly the capability of the agent and the movements are improved with the next prototype. Each movement can be controlled and adjusted by the technician and the granularity of adjustments has been increased.

5.1.4 Error Handling

Unexpected inputs, responses and error handling was also identified during this survey and analysis. Although the 3-minute video doesn't provide a clear picture of this capability, the next prototype should be examined to answer the following questions:

- How does the platform handle errors?
- What is identified as a natural response for an unexpected input?
- How well can it identify sarcasm?

In coordination with the USC-ICT Simsensei team, there is a specific department that is dedicated to the dialogue of the EMPOWER system. The team that coordinated and supported this thesis is focused on improving the platform while the dialogue team is focused on the dialogue improvements. Currently as the system is today, it operates on linear input with very limited branching to support the requirement of an open-ended conversation. The agent does not currently generate new responses as the current state of machine learning cannot guarantee that they would be meaningful responses. As seen in the media, there are platforms such as the Microsoft AI that was trained by a Twitter user [11] to return racist responses. They removed the chatbot

from use to address the issue by of such learning and investigate their branching limitations that caused the system to respond with racist responses. With such a sensitive system and scope of the project, it is imperative that the agent always responds with meaningful responses. However, this is a huge area for future work and to improve the ML and response planning to optimize a system to be able to generate new and informative responses. Since the system is based on open-ended questions it doesn't really error on input. It adjusts to analyzing the mannerisms of the participants followed by the verbal responses. Additional consideration during the analysis was given to user behavioral signals related to depression, suicide ideation, and PTSD through the following 5 areas:

- Body tracking
- Facial expression analysis
- Agent Visualization
- Dialogue management
- Behavioral realization

These five areas will be examined within the Kevin prototype once the software and platform is made available to the researcher. Currently, with the Ellie platform and previous tests, these were areas that needed improvement while the system operated in a semi-autonomous mode. These proved to be improved during the 'Man behind the curtain' tests conducted which had two operators controlling the platform's responses. Ideally, we can examine the verbal content-language, visual-face/head behavior, and voice-speech prosody on a technical level to identify improvements for future integration.

5.2 Integration

Ideally, each of the above identified gaps and solutions has a method identified to be integrated with the 'Kevin' prototype. The specific methods of integration are dependent on the feedback received during the survey. Then each identified method will have a specific requirement and evaluation criteria to measure the success of its integration. This section will have each of these outlined. The setup, protocols, and evaluation of the specific experiments will be determined in coordination with USC-ICT and West Point, which were attempted over the summer of 2021, however, due to lack of funding and IRB approvals, the capstone integration was cancelled. To further evaluate the identified optimization, individual experiments were attempted. This researcher conducted individual interactions with the 'Kevin' prototype as-is, which in the current state is not fully capable of demonstrating the identified improvements.

5.3 Further Optimization

Further optimization will be based on the previous two sections - identified and developed following the system test with USC-ICT in late 2021 - early 2022. System test of 'Kevin' the second generation prototype that was developed from the original ELLIE [9]. Further development of virtual human multimodal capability to appropriately respond to individuals that are identified to be in need, to address the issue identified as a barrier of AI integration in the realm of Patient Safety. Possibly expand the coverage of such a program from Soldiers to be available to other high risk communities such as Correctional Officers, Nurses (during high stress COVID), or younger adults that are at high risk of suicide. Goal would be a fully integrated system that can identify all symptoms and risk factors as an all-inclusive open source application releasable to all applicable entities that can intervene.

Although the USC-ICT team has been aggressively working the virtual human platforms for a decade, all of the identified future work hinge on additional funding and approvals. Through this research and its results, significant gains are identified and the impact on the American Soldier can be extensive if the prototype is optimized before release. The hope of this work is to support the request for additional funding and scope expansion.

Chapter 6

CONCLUSION

An American Soldier's perspective and insights into barriers to care can provide the technical development team a significant advantage to optimizing the Virtual Human Agent. This survey returned valuable feedback that can solve integration issues with the Virtual Human platforms future use. Specifically, the ability to collect 51 survey responses while developing independent analysis and exploration of the VH Agent of Kevin for the project proved to be extremely valuable and optimizing the agent. Ideally an immediate improvement that can be actualized is providing the participant or end user the confidence to trust the system, this will prove to be the most valuable system improvement gleaned in this study. This confidence is increased by specifically addressing critical issues such as security of the system and the data, optimization and customization integration and error handling.

With additional funding the USC-ICT teams that are currently developing the EMPOWER platform and current prototype of Kevin could integrate all of the recommended changes identified through this survey. The customization of the platform such as various agents, environments, and ability to adjust the ability of the agent require extensive code writing and development. Additionally, the language team could continue to investigate various Machine Learning search methods in order to develop a training regime to build the reliability of unscripted responses for the VH agent to use through future interactions. Currently the system operates on linear, prerecorded responses due to the lack of reliability within a sensitive interaction space.

BIBLIOGRAPHY

[1] Mental health statistics. https://www.nimb.nih.gov, Mar 2021.

[2] T. V. Brook. Suicide rate among active-duty troops jumps to six-year high, covid-19 stress could make it even worse. *USA Today*, Oct 2020.

[3] G. B. Colombo, P. Burnap, A. Hodorog, and J. Scourfield. Analysing the connectivity and communication of suicidal users on twitter. *Computer Communications*, 73:291–300, 2016.

[4] J. Du, Y. Zhang, J. Luo, Y. Jia, Q. Wei, C. Tao, and H. Xu. Extracting psychiatric stressors for suicide from social media using deep learning. *BMC Medical Informatics and Decision Making*, 18(S2), 2018.

[5] Facebook. Under the hood: Suicide prevention tools powered by AI. *Facebook AI*.

[6] K. K. Fitzpatrick, A. Darcy, and M. Vierhile. Delivering cognitive behavior therapy to young adults with symptoms of depression and anxiety using a fully automated conversational agent (woebot): A randomized controlled trial. *JMIR Mental Health*, 4(2), 2017.

[7] T. M. Fonseka, V. Bhat, and S. H. Kennedy. The utility of artificial intelligence in suicide risk prediction and the management of suicidal behaviors. *Australian & New Zealand Journal of Psychiatry*, 53(10):954–964, 2019.

[8] A. S. Miner, A. Milstein, S. Schueller, R. Hegde, C. Mangurian, and E. Linos. Smartphone-based conversational agents and responses to questions about mental health, interpersonal violence, and physical health. *JAMA Internal Medicine*, 176(5):619, 2016.

[9] S. Rizzo. Simsensei & multisense: Virtual human and multimodal perception for healthcare support. https://www.youtube.com/watch?app=desktop&v=ejczMs6b1Q4&t=28s, Feb 2013.

[10] V. Venkataramani, J. Hoon, S. Kulkarni, N. Lawrence, M. Marchukov, D. Petrov, L. Puzar, Z. Amsden, N. Bronson, G. C. Iii, and et al. Tao: Facebook's distributed data store for the social graph. *Facebook INC. USENIX Annual Technical Conference (USENIX ATC '13)*, page 49–60, 2012.

[11] J. Vincent. Twitter taught microsoft's ai chatbot to be a racist asshole in less than a day. *The Verge*, Mar 2016.

[12] M. J. Vioules, B. Moulahi, J. Aze, and S. Bringay. Detection of suicide-related posts in twitter data streams. *IBM Journal of Research and Development*, 62(1), 2018.

[13] B. Zohuri and S. Zadeh. Global suicide rate among youngsters increasing significantly. *Online Journal of Neurology and Brain Disorders*, 3(5), 2020.

[14] B. Zohuri and S. Zadeh. The utility of artificial intelligence for mood analysis, depression detection, and suicide risk management. *J. of Health Science*, 8(2), 2020.

APPENDICES

Appendix A

THESIS FLYER

Appendix B

SURVEY QUESTIONS

Optimizing a Virtual Human Platform

Virtual Human Platform Optimization and Feedback

This 'survey' is requested to assist a researcher to develop an unbiased, multi-opinion perspective while analyzing the selected virtual human platform with regards to barriers to care, the 5 HCI principles and 6 Intelligence Attributes. Please provide your honest opinion by selecting one value to each area on a scale of 1 to 5. Each scale will be slightly different but will generally reflect 1 being the best, 3 being neutral and 5 being the worst. Followed by a field to include any additional comments for each principle and to recommend improvements.

Your personal or confidential information, such as your name and identifying information, will NOT be requested, collected or maintained.

Recommendations for improvements will be used to better understand access to care.

Optimizing a Virtual Human Platform

* Required

Consent

Please read and sign the consent form to validate if you choose to participate in this survey.

INFORMED CONSENT TO PARTICIPATE IN A RESEARCH PROJECT: "Optimizing a Virtual Human Platform to assist with Depression/Suicidal Ideation Identification for the American Soldier in coordination with USC-ICT"

INTRODUCTION - This form asks for your agreement to participate in a research project on optimizing a virtual human platform to assist individuals in need through a 'Understanding the Use of Care' survey. Your participation involves watching a 3 minute video illustration of the platform, and completing a survey to provide recommendations for improvements to the system to solicit open and honest feedback from the end user of the system.

PURPOSE OF THE STUDY AND PROPOSED BENEFITS

• The purpose of the study is to receive anonymized responses to the simple survey questions to be consolidated and quantified in support of integrated the recommended changes into a future system test conducted by USC-ICT and West Point.
• Survey data will be retained for a period of no more than 1 year to allow for examination and integration of the Virtual Human platform. The anonymized responses will be shared with other researchers and re-used for the EMPOWER project in support of USC-ICT and West Point system test.

YOUR PARTICIPATION

• If you are 18 or older and agree to participate, you will take a survey following the discussion and video demonstration. The survey should take about 5-10 minutes to complete.
• The survey link will be distributed via hard copy or digital link (whichever is optimal for you) as a web form hosted externally by Google. Your data will remain confidential as your name and identifying information will not be collected or maintained

PROTECTIONS AND POTENTIAL RISKS

• The possible risks or discomforts associated with participation in this study include feeling sad or frustrated as this is a sensitive topic for those that have experienced a significant life effect and received care.
• Your confidentiality will be protected as your name and identifiable information will not be collected with your survey.
• Survey responses will be stored on the restricted Google drive or hard copy maintained by researcher only until consolidated as one inclusive document of all survey results.
• Participants results will be consolidated to give quantitative results and recommendations for improvements.

RESOURCES AND CONTACT INFORMATION

• If you should experience any negative outcomes from this research, please be aware that you may contact Dr. Franz J. Kurfess (fkurfess@calpoly.edu or 805-756-7179) at Cal Poly.
• This research is being conducted under the guidance of Dr. Franz Kurfess, Professor in the Department of Computer Science and Software Engineering at Cal Poly, San Luis Obispo. If you have questions regarding this study or would like to be informed of the results when the study is completed, please contact the researcher(s) at cmmonaha@calpoly.edu.
• If you have concerns regarding the manner in which the study is conducted, you may contact Dr. Michael Black, Chair of the Cal Poly Institutional Review Board, at (805) 756-2894, mblack@calpoly.edu, or Ms. Trish Brock, Director of Research Compliance, at (805) 756-1450, pbrock@calpoly.edu.

AGREEMENT TO PARTICIPATE

If you are 18 or older and agree to participate in the survey of Understanding the Use of Care, please indicate your agreement by signing below. Please retain a copy of this form for your reference, and thank you for your participation in this research.

*

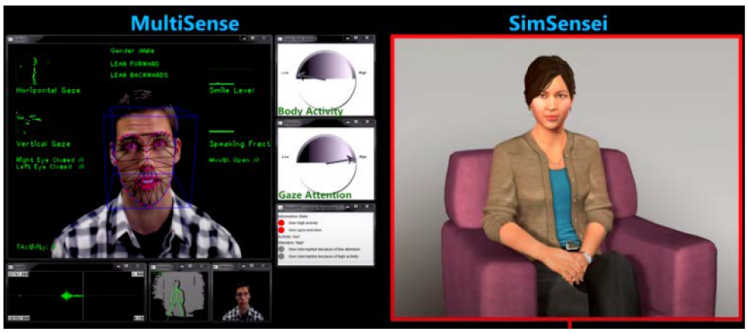
- ○ Yes, I am 18 or older and agree to participate by taking 'Understanding the Use of Care' survey.
- ○ No, I do not agree to participate by taking the 'Understanding the Use of Care' survey.

Optimizing a Virtual Human Platform

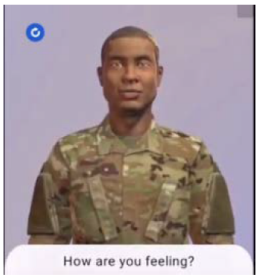
* Required

Access/Barriers to Care

This is a Virtual Human Agent named Ellie. Please watch the 3 minute demonstration and then complete the follow-on questions. **You can also view directly on YouTube to see it full screen**. https://www.youtube.com/watch?v=ejczMs6b1Q4&t=28s



This is a Virtual Human app called Battle Buddy. Please watch the 2 minute demonstration. **You can also view directly on YouTube to see it full screen**.
https://www.youtube.com/watch?app=desktop&v=aG-wWRNoIhA



1. If you had a life crisis or significant event would you seek assistance? *
   - ○ Yes
   - ○ No
   - ○ Maybe
   - ○ Not sure

2. If you were to seek trusted health care information, where would you go? Or ho would you prefer to interact? (select all you might use) *
   - ○ In person at a clinic
   - ○ Over the phone (Army OneSource, FMLC, etc)
   - ○ Through an 'app' or virtual buddy
   - ○ Online with a Virtual Human Agent - such as Ellie
   - ○ Social Media
   - ○ Other:
   - ○ [ ]

3. How secure do you believe the above methods ('app', virtual agent, etc.) is? *

   Very secure (my information will not be leaked)

- 1
- 2
- 3
- 4
- 5

Not secure (my information is vulnerable)

4. If you have concerns about security of interacting with your preferred healthcare method, what are they?

Your answer

5. Are you concerned about being honest with a virtual agent? *

Not concerned (I would be 100% honest)

- 1
- 2
- 3
- 4
- 5

Very concerned (I don't feel comfortable disclosing my info to a virtual agent)

6. If you feel that you can't be 100% open and honest, what are your hesitations?

Your answer

7. What might you want to use the agent for?
(Select as many options as you would like) *

- Resilience/Mindfulness techniques
- Stress-relief in high-stress scenarios
- Crisis mitigation
- Just someone to talk to
- Just 'help'
- Find other resources
- None

- Other:
- [ ]

8. Are there any specific characterizations that would make your interaction better? (Select all recommended preferences you would like to see integrated) *

- Male
- Female
- Non-Binary
- Other Gender
- Uniform
- Civilian Clothing
- Leadership Position
- Buddy rank (equal to your rank)
- In an office
- Outside atmosphere
- None recommended
- Other:
- [ ]

* Required

Evaluate the Agent

When thinking about the Virtual Agent (ELLIE) demonstration video - please try to imagine that you were the one interacting with ELLIE. In this imagined scenario, please attempt to answer the following questions in response to the 6 principles and 5 attributes to evaluate the agent.

Principle #1 – Cognitive Foundations, as the mental model of the Virtual Agent and how well it matches your mental model. Through your natural perception, memory and knowledge – the virtual agent should be designed to interact naturally and on the same level.

The next three questions pertain to Principle #1

How easy was the virtual agent designed? Was it natural and easy for you to interact with? *

Natural and easy to use
- 1
- 2
- 3
- 4
- 5

Complicated and hard to follow

Was your interaction familiar and expected OR was it unfamiliar/unexpected situations? *

Familiar and expected

- 1
- 2
- 3
- 4
- 5

Unfamiliar or unexpected situations

Were the limitations obvious? *

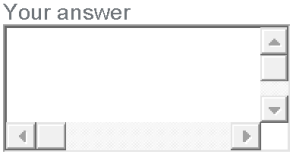
Yes - Completely obvious

- 1
- 2
- 3
- 4
- 5

No - Did not notice any limitations

Please provide any additional insight or recommendations for improving the agents' ease of use and natural interactions.

Your answer



Principle #2 – Interaction Spaces, the physical and cognitive space in which the user and the automation interact can contribute to the data collected. Through physical, visual, and textile, interaction the interaction can be inviting and solicit more interaction.

The next two questions pertain to Principle #2

Was your interaction pleasing? *

Yes (Extremely pleasing)

- 1
- 2
- 3
- 4
- 5

No (Not pleasing at all)

Did you want to continue the conversation? *

Yes (could talk all day)

- 1
- 2
- 3
- 4
- 5

No (couldn't end soon enough)

Do you think you would be less 'judged' by a virtual agent than an actual person? *

Judgment-free

- 1
- 2
- 3
- 4
- 5

Judged more

Please provide any additional insight or recommendations for improving the agents' interaction spaces: list any physical, visual, and textile changes that could be made to make the interaction more inviting.

Your answer

Principle #3 – Input-Output Devices, are the methods in which you interact with the agent. Audio, visual and textual devices contribute to the direct or indirect interactions.

Next two questions pertain to Principle #3

Was the input-output device used optimal for the type of interaction? *

Yes (optimized)

- 1
- 2
- 3
- 4
- 5

No (needs improvements)

Were there multiple choices for input-output device? *

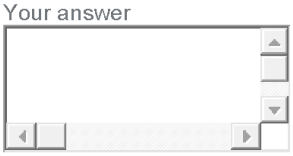
Yes (multiple options)

- 1
- 2
- 3
- 4
- 5

No (no options)

Please provide any additional insight or recommendations for improving the agents' input/output devices: list any other methods you think would make it easier for continued interaction.

Your answer



Principle #4 - Interaction Styles, are various methods a user can interact with the system through. These can vary from batch systems, command-line, full-screen interfaces, menus, forms, direct manipulation interfaces, graphical user interfaces, user assistance, speech and mobile devices. This can be further broken down into four parts: articulation, performance, presentation, and observation.

Next three questions pertain to Principle #4

How easy was it for you to articulate what you wanted to convey to the system? *

Easy

- 1
- 2
- 3
- 4
- 5

Difficult

How was the performance of the system to keep up with the required interactions? *

Fast and easy

- 1
- 2
- 3
- 4
- 5

Slow and difficult

How easy was it for you to understand what the system was presenting back to you? *

Easy
- 1
- 2
- 3
- 4
- 5

Difficult

Please provide any additional insight or recommendations for improving the agents' interaction style to make the interaction easier to you to articulate your intent to the agent or for the agent to be understandable.

Your answer

Principle #5 – Speech-based Interaction, elicit additional considerations such as ability to distinguish accents, mitigate noise, the range of auditory signal, transmitting and receiving channels, and encoding. For a human it may be seen as a speaker, through the air (noise) to the ears as the receiver. For a computer, it may be seen as the transmission from a speaker through the air, to a microphone.

Next three questions pertain to Principle #5

Did the speech interaction provide optimized channels and encoding? In other words, is the way it communicated easy for you to understand and do you feel it understood you? *

Yes (it was very easy/understood everything)
- 1
- 2
- 3
- 4
- 5

No (the encoding/channels are wrong - I couldn't understand anything)

Is the agent able to distinguish accents? *

Best
- 1
- 2
- 3
- 4
- 5

Worst

Was the range of signal and mitigation of noise appropriate? In other words, was the system sound levels ok? Was it free of interference or noise? *

Best
- 1
- 2
- 3
- 4
- 5

Worst

Additional insight on improvements for the HCI principles of the agent. To include recommendations for alternative devices, change in design, etc.

Your answer

Attribute #1 Context understanding as it relates to the intent of the input and incorporating inference to draw conclusions based on known facts. A level of intelligence is required to apply it to compute the intent of a statement - not merely pattern matching (noun and verb). Did the Virtual Agent understand your input? *

Most Intelligent (best)
- 1
- 2
- 3
- 4
- 5

Least Intelligent (worst)

Attribute #2 – Conversational design to handle the input in a conversational way rather than as a transactional platform. Conversational flow is a distinct characteristic of a human conversation; typically we do not jump from 1 concept to a completely

difference concept without a transition. Response planning is critical to mimicking human behavior. Did the Virtual Agent illustrate a conversational design within the responses? Can the Virtual Agent stay on topic? How many branches does the Virtual Agent have to a specific question? *

Most Intelligent (best)

- 1
- 2
- 3
- 4
- 5

Least Intelligent (worst)

Attribute #3 – The Depth of knowledge is critical. The simple facet of accepting multiple questions at once, offering multiple solutions to the same question or the ability to learn new responses based on context rather than static database via natural branching require further programming or complexities than a simple question-answering. Can the Virtual Agent accept more than one question at a time? Can the Virtual Agent learn new responses based on conversation correctly? Can it plan responses from future inferences? *

Most Intelligent (best)

- 1
- 2
- 3
- 4
- 5

Least Intelligent (worst)

Attribute #4 – Error detection and handling is a clear way to see how extensive the natural language processing of a Virtual Agent extends. The ability to recognize variations such as slang, idioms and colloquialisms is more complex than just a standard dictionary would provide. How the Virtual Agent reacts to out of range topics is important. Can the Virtual Agent detect slang? Can the Virtual Agent correctly identify misspellings? Can the Virtual Agent coach the user to clarify – to recover from out of scope questions? *

Most Intelligent (best)

- 1
- 2
- 3

- 4
- 5

Least Intelligent (worst)

Attribute #5 – Being able to determine that the pronoun is associated with a separate part of speech and relate it to that part is more intricate. Can the Virtual Agent correctly identify the correct meaning of the ambiguous terms (such as contractions or slang)? Can the Virtual Agent resolve the 'he' to reference the male that was mentioned three sentences before to provide the correct responses? *

Most Intelligent (best)

- 1
- 2
- 3
- 4
- 5

Least Intelligent (worst)

Attribute #6 – Emotional Intelligence (Empathy), illustrating empathy through 'human-like' behavior to help individuals virtually to interact based on the 'mood' and 'need' of the conversation. Did the Virtual Agent make you feel like you were talking to a real person? Did the Virtual Agent 'pay attention' to your feelings? Can the Virtual Agent detect your thought processes or feelings? *

Most Intelligent (best)

- 1
- 2
- 3
- 4
- 5

Least Intelligent (worst)

Additional insight on improvements for the 'intelligence' of the Virtual Agent.

Your answer

Optimizing a Virtual Human Platform

Voluntary Future involvement

If there are follow-on surveys, system tests or interaction opportunities with a Virtual Human Agent, the researcher is providing an opportunity to provide your military email address if you would like to be included Your email and information will NOT be collected here - as the questions will remain anonymized. Please email the researcher directly at christina.m.monahan.mil@mail.mil if you would like to participate in future work.

Appendix C

IRB APPROVAL

Human Subjects Approval - Conditions Cleared

Humans Subjects Research IRB <hs-irb@calpoly.edu> Sat, May 22, 2021 at 2:10 PM
To: Christina Monahan <cmmonaha@calpoly.edu>, Humans Subjects Research IRB <hs-irb@calpoly.edu>, Franz Kurfess <fkurfess@calpoly.edu>

Research team:

Your response to the conditions placed on the approval for the study referenced below has been reviewed and accepted by the Cal Poly IRB, and you may proceed with your project.

Study Title: Optimizing a Virtual Human Platform to assist with Depression/Suicidal Ideation Identification in coordination with USC-ICT
PI: Christina Monahan

Click here to access application: Protocol Approval Form

This approval will require a status check by 05/21/2022. Please be aware that it is your responsibility as the persons in charge of this research project to ensure that, with respect to human subjects, the work is carried out as described in the proposal and the rights of the subjects are fully protected.

You have listed the following Grants Development numbers as associated with this project: N/A

If the associated GDO numbers change, you will need to submit a modification request through IRB Manager. The IRB protocol number associated with this project is: 2021-142

We wish you success in your research efforts.

If you have any questions, please contact the Cal Poly IRB at hs-irb@calpoly.edu or (805 756-1450). Thank you.

Research Protocol:

Optimizing a Virtual Human Platform to assist with Depression/Suicidal Ideation Identification for the American Soldier

in coordination with USC-ICT

Survey – "Understanding the Use of Care"

Primary Investigator: Christina Monahan

Advisor: Franz J. Kurfess

Completion: 4 June 2021

Master's Thesis Research Protocol, Spring 2021

Cal Poly San Luis Obispo, Computer Science Department

Statement of Purpose, Benefits, and Hypotheses

Suicide surpassed homicide to be the second leading cause of death among people 10-24 years old in the United States [1]. To capitalize on the modern shift to electronic-based interactions [1], the use of Artificial Intelligence (AI) and Machine Learning (ML) methods to aid in identification have been implemented previously implemented in Virtual Human interviewing platforms. This survey will be used to solve the gap in early identification of depression and suicidal ideation using a virtual human interviewing platform but soliciting honest, open, and current feedback from Soldiers on how to optimize such system to encourage its use in the future. Specifically, the survey will provide an examination and identification of gaps in current processes and recommendations to be included in future development of the EMPOWER (Enhancing Mental Performance and Optimizing Warfighter Effectiveness and Resilience: From MultiSense to OmniSense) platform.

Benefits of such a survey will provide real-time, relevant, and honest insight as to where Soldiers are currently turning to for care, how can a platform be optimized to encourage interaction and provide the judgment-free interactive space for individual in crisis in the future.

Hypotheses for this specific survey are that the current systems that leadership believe are in place are not optimized for the newer generation of Soldiers to use. The electronic-based platforms are an interaction space that can be optimized to get Soldiers to seek care when in need.

Purpose of the study is to receive anonymized responses to the simple survey questions to be consolidated and quantified in support of integrated the recommended changes into a future system test conducted by USC-ICT and West Point.

Participants will be 18 or older Active Duty Soldiers and the data will be collected via a hard copy survey or digital link (whichever is optimal for the user) as a web form hosted externally by Google. User data will remain confidential as name and identifying information will not be collected or maintained. Risks will be minimized as there are inherent risks while discussing sensitive topics. The possible risks or discomforts associated with participation in this study include feeling sad or frustrated as this is a sensitive topic for those that have experienced a significant life effect and received care.

Methods

System Design

- The survey considers the research technologies currently employed within the USC-ICT applications, software, studies, and prototypes and identify AI/ML methods that could be applied, examined, or improved.
- The survey will consider the research technologies and identifying the following HCI principles for the platform:
    - Principle #1 – **Cognitive Foundations**, as the mental model of the platform and how well it matches the users mental model. Through the user's natural

perception, memory and knowledge, each platform should be designed to interact naturally and on the same level. Was the platform designed to be used easily? The interaction of the user should be familiar and expected. Limitations of the platform should not be obvious.

- Principle #2 – **Interaction Spaces**, the physical and cognitive space in which the user and the automation interact can contribute to the data collected. Through physical, visual, and textile, interaction the interaction can be inviting and solicit more interaction. Goal is the interaction is pleasing, and solicits the user to continue the interaction.
- Principle #3 – **Input-Output Devices**, are the methods in which the user interacts with the agent. Audio, visual and textual devices contribute to the direct or indirect interactions. The input-output device used should be optimal for the type of interaction used. A way to address this is to have multiple choices for input-output devices. This is also a consideration for Americans Disability Act, to ensure there are methods for those that may have reduced capabilities.
- Principle #4 - **Interaction Styles** are various methods a user can interact with the system through. These can vary from batch systems, command-line, full-screen interfaces, menus, forms, direct manipulation interfaces, graphical user interfaces, user assistance, speech and mobile devices. This can be further broken down into four parts: articulation, performance, presentation, and observation. These four parts are often viewed as a circular cycle of communication - How well a user is able to articulate the input to the system, the performance of the system to keep up with the required interactions, the presentation of the response from the system back to the user and lastly how well was the response observed by the user.
- Principle #5 – **Speech-based Interaction**, elicit additional considerations such as ability to distinguish accents, mitigate noise, the range of auditory signal, transmitting and receiving channels, and encoding. For a human it may be seen as a speaker, through the air (noise) to the ears as the receiver. For a computer, it may be seen as the transmission from a speaker through the air, to a microphone. A few questions are: Did the speech interaction (if available), provide appropriate channels and encoding? Was the platform able to distinguish accents? Was the range of signal and mitigation of noise appropriate?

- Identify and compare the machine learning 'Intelligence' of the platform based on the following attributes:
  - Criteria #1 **Context understanding** as it relates to the intent of the input and incorporating inference to draw conclusions based on known facts. A level of intelligence is required to apply it to compute the intent of a statement - not merely pattern matching (noun and verb).
  - Criteria #2 – **Conversational design** to handle the input in a conversational way rather than as a transactional platform. Conversational flow is a distinct characteristic of a human conversation; typically we do not jump from 1 concept to a completely different concept without a transition. Response planning is critical to mimicking human behavior.

- Criteria #3 – The **depth of knowledge** is critical. The simple facet of accepting multiple questions at once, offering multiple solutions to the same question or the ability to learn new responses based on context rather than static databases via natural branching require further programming or complexities than a simple question-answering.
- Criteria #4 – **Error detection and handling** is a clear way to see how extensive the natural language processing of a Platform extends. The ability to recognize variations such as slang, idioms and colloquialisms is more complex than just a standard dictionary would provide. How the platform reacts to out of range topics is important.
- Criteria #5 – **Coreference resolution**, word sense disambiguation and handling of contractions is another level of natural language processing that can be used to analyze the behavior of a platform. Although word sense and contractions should be able to be handled with simple code the coreference resolution is a little more complicated. Being able to determine that the pronoun is associated with a separate part of speech and relate it to that part is more intricate.
- Criteria #6 - **Emotional Intelligence (Empathy)**, illustrating empathy through 'human-like' behavior to help individuals virtually to interact based on the 'mood' and 'need' of the conversation. The platform's functionality to make a user feel like they were talking to a real person. The underlying implied ability to 'pay attention' to the users feelings, by keying in on simple identifiers. And the ability of the Platform to detect implied thought processes or feelings.

- Specifically for the Virtual Human Platform:
  - Male vs. Female.
  - Military roles
  - Customization

Subjects and Subject Characteristics

- The source subjects for this survey will be 18 or older Active Duty Soldiers serving at Fort Carson, Colorado within the 4th Infantry Division.
- Selection criteria is 100% voluntary as all Soldiers can provide insight but are not required.
- Goal for the survey is to solicit as many responses as possible but am aiming for 50 responses at a minimum.
- No prospective subjects are members of seven vulnerable groups specifically identified in the basic study online form. However, active duty Soldiers are likely a more vulnerable population than a typical person due to their training to obey orders and the stress of active duty life. Although I am an active duty Soldier, I will not use my rank, or position to pressure participation. I fully understand the stress of active duty life, and the intent of this survey is to get real unbiased feedback from Soldiers on how we can optimize this system to have it deploy successfully in the future.

Investigators

- CW4 Christina Monahan is a current California Polytechnic Masters student that is requesting to conduct a survey of other Active Duty Soldiers to get real feedback on how

the system can be optimized. Previously trained on the Human-Computer Interaction principles and Intelligence criteria for virtual agents.

- Risks to subjects will be minimized as only a written survey will be requested.
  - No physical harm is anticipated.
  - Psychological harm is low and will be minimized by disclosing the coverage of sensitive topic at the introduction.
  - Questions will be formed in order not to solicit emotional detailed event responses but remain focused on optimizing the interaction space.

Materials and Procedures

- Soldiers will volunteer to complete a 15 minute survey.
- Survey begins with a review and signature of the consent form (see Appendix 1).
- Next will be a review of the background and motivation of the overall research provided in Appendix 2 (below).
- Soldiers will watch a 3 minute video of the ELLIE agent from https://www.youtube.com/watch?v=ejczMs6b1Q4
- Soldiers will complete a 30 question survey on how they would like to access care and if it is a virtual agent – how can it be optimized. (See Appendix 3). This will be done via a digital link to be submitted in a Web Form and saved to a restricted Google Drive.

Survey Location

- Will be Fort Carson, Colorado as that is the military installation that the PI recently moved from and has established relationships with leadership, knowledge of processes and procedures, and access to facilities to conduct survey. Since this is conducted digitally, participants may be from other military installation outside of Fort Carson now.

Informed Consent Form

- Attached here in Appendix 1 and will be the first item reviewed prior to any discussion of the survey, background or illustration.
- This consent form is the first page of the electronic survey and will not be conducted separately.

Debriefing Requirement

- There will be no requirement for debriefing as this is a straight forward survey involving no deception or incomplete disclosure.

Recruitment Material

The following is a simple email I was going to distribute as recruitment material:

- My Master's Thesis is to help improve methods for Soldiers can seek care. I have an opportunity to create and conduct a survey to improve a virtual human agent. If you could possibly spare 15 minutes - I could really use your feedback.
- https://forms.gle/Tco4hgPZeN444Aow5

- This is the link to the Google form. I am not collecting any names or identifying information.
- First page is general information, second page is the consent form, and third page begins the survey with a 3-minute YouTube of ELLIE which I get to help modify the 2nd generation prototype that will undergo a system test at West Point in June.

References

References

[1] Zohuri, B., Zadeh, S. (2020). Global Suicide Rate Among Youngsters Increasing Significantly. Journal of Neurology and Brain Disorders, 3(5), 300-310. https://doi.org/10.32474/OJNBD.2020.03.000175

Appendix 1

INFORMED CONSENT TO PARTICIPATE IN A RESEARCH PROJECT:
"Optimizing a Virtual Human Platform to assist with Depression/Suicidal Ideation Identification for the American Soldier in coordination with USC-ICT"

INTRODUCTION

This form asks for your agreement to participate in a research project on optimizing a virtual human platform to assist individuals in need through a 'Understanding the Use of Care' survey. Your participation involves watching a 3 minute video illustration of the platform, and completing a survey to provide recommendations for improvements to the system to solicit open and honest feedback from the end user of the system.

PURPOSE OF THE STUDY AND PROPOSED BENEFITS

- The purpose of the study is to receive anonymized responses to the simple survey questions to be consolidated and quantified in support of integrated the recommended changes into a future system test conducted by USC-ICT and West Point.
- Survey data will be retained for a period of no more than 1 year to allow for examination and integration of the Virtual Human platform. The anonymized responses will be shared with other researchers and re-used for the EMPOWER project in support of USC-ICT and West Point system test.

YOUR PARTICIPATION

- If you are 18 or older and agree to participate, you will take a survey following the discussion and video demonstration. The survey should take about 5-10 minutes to complete.
- The survey link will be distributed via hard copy or digital link (whichever is optimal for you) as a web form hosted externally by Google. Your data will remain confidential as your name and identifying information will not be collected or maintained.

PROTECTIONS AND POTENTIAL RISKS

- The possible risks or discomforts associated with participation in this study include feeling sad or frustrated as this is a sensitive topic for those that have experienced a significant life effect and received care.

- Your confidentiality will be protected as your name and identifiable information will not be collected with your survey.
- Survey responses will be stored on the restricted Google drive or hard copy maintained by researcher only until consolidated as one inclusive document of all survey results.
- Participants results will be consolidated to give quantitative results and recommendations for improvements.

RESOURCES AND CONTACT INFORMATION

- If you should experience any negative outcomes from this research, please be aware that you may contact Dr. Franz J. Kurfess (fkurfess@calpoly.edu or 805-756-7179) at Cal Poly.
- This research is being conducted under the guidance of Dr. Franz Kurfess, Professor in the Department of Computer Science and Software Engineering at Cal Poly, San Luis Obispo. If you have questions regarding this study or would like to be informed of the results when the study is completed, please contact the researcher(s) at cmmonaha@calpoly.edu.
- If you have concerns regarding the manner in which the study is conducted, you may contact Dr. Michael Black, Chair of the Cal Poly Institutional Review Board, at(805) 756-2894, mblack@calpoly.edu, or Ms. Trish Brock, Director of Research Compliance, at (805) 756-1450, pbrock@calpoly.edu.

AGREEMENT TO PARTICIPATE

If you are 18 or older and agree to participate in the survey of Understanding the Use of Care, please indicate your agreement by signing below. Please retain a copy of this form for your reference, and thank you for your participation in this research.

___ Yes, I am 18 or older and agree to participate by taking 'Understanding the Use of Care' survey.

___ No, I do not agree to participate by taking the 'Understanding the Use of Care' survey.

_____ _____

Signature of Volunteer Date

_____ _____

Signature of Researcher Date

Appendix 2

Introduction

- Initial motivation is to impact the critical need of the American Soldier in crisis by improving identification to intervene early. Depression and suicidal ideation continue to increase across society today and is very personal to me. As an active duty Army Soldier (21 years), I intimately understand the sacrifices of service members in the military today and during combat. In 2019, two of my Soldiers decided to take their life and my closest co-worker attempted suicide. These three gentlemen will have an everlasting impact on my life and are the reasons why this topic is interesting to me. It is extremely important to recognize that the key age of the suicide statistic analyzed is both inclusive of most college students and of young military Soldiers as well.
- The focus of this systematic comparison of existing approaches of identification methods will begin with identifying currently applied ML and AI methods in recognizing and identifying symptoms and risk factors across millions of data points prevalent in advanced platforms such as ELLIE or EMPOWER. Followed by the analysis of adequate methods and identification of areas of improvement based on an independent experiment that will be conducted on the platform or in conjunction with upcoming studies conducted by USC-ICT. Furthermore, due to the isolation of the pandemic, this topic is extremely important and the results from the experiment can be a baseline for future study and development of automated systems to identification methods.
- Intended users of this systematic comparison are individuals in the following areas:
    - Computer Science
    - Psychology
    - Mental healthcare
    - Improving current AI methods

Background

- There are several risk factors identified to contribute to suicidal ideations. The first factor is typically a previous suicide attempt, followed by a diagnosis of a type of depression. [20] Although this is not always the case, this research is examining the link to "Depression". Further exploring this concept, there are several different types of Depression across a full spectrum of severities. Types such as Major Depression, Persistent Depression Disorder (Dysthymia), Perinatal Depression, Seasonal Affective Disorder (SAD), Psychotic Depression, Disruptive Mood Dysregulation Disorder, Premenstrual Dysphoric Disorder, and Bipolar Disorder.[20]
- Depression indicators considered in this research are not just a 'feeling' of sadness for a day or two, we are looking at a deeper level. Indicators such as [18]:
    - Persistent sad, anxious, or "empty" mood;
    - Feelings of hopelessness or pessimism;
    - Feelings of guilt, worthlessness, or helplessness;
    - Loss of interest or pleasure in hobbies or activities;
    - Decreased energy, fatigue, or being "slowed down";
    - Difficulty in concentrating, remembering, or making decisions;

    - Difficulty in sleeping, early-morning awakening, or oversleeping;
    - Appetite and/or weight changes;
    - Thoughts of death or suicide or suicide attempts;
    - Restlessness or irritability;
- These indicators for depression and suicidal ideation have many visual, audio and textual cues that can be identified through various AI and ML algorithms.
- Artificial Intelligence (AI) advancements provide capabilities to analyze the mood of a person by looking at his or her face via Facial Expression Recognition, CR (character recognition) or IP (image processing); examine a person's voice through Voice Recognition (VR) or analyze a person's text online in real-time. Each of these will be identified within the EMPOWER platform and related studies. An experiment to solicit user feedback will be developed and a study will be conducted to test a developed hypotheses based on feedback from Professor Rizzo to provide unbiased evaluation.

This analysis provides the opportunity of investigating methods that give professionals a technological advantage to identify people in need – possibly prevent suicide attempts and manage mood changes among the people that not only are isolated socially but show signs of depression on their face, their body movements and in their voice.